Hidden Changes: from CCTV to “Smart” video surveillance

Joseph Ferenbok and Andrew Clement

To appear in A. Doyle, R. Lippert & D. Lyon (Eds.), *Eyes Everywhere: The Global Growth of Camera Surveillance* (In Press). Devon Willan Publishing.

Introduction

Behind the familiar camera housings and lenses, and hidden from their surveilled subjects, video surveillance systems are becoming much “smarter”. Though many people still associate video surveillance with analog CCTV (closed-circuit television), these new digital systems differ significantly. Behind the scene, video analytics—computer-assisted video monitoring, analysis and indexing—is changing what can be ‘seen’ using surveillance camera networks. Video analytics (VA) is software that uses signal processing and pattern recognition techniques to automatically generate meaningful or semantic data from video images. Video analytics marks a paradigmatic shift in visual surveillance practices—in how information is purposed, and repurposed — and in the potential consequences for surveillance subjects.

How video surveillance systems are being networked and augmented remains largely out of public view. This obscured deployment contributes to “a lack of realism about what could be expected from CCTV systems” (Deisman et al., 2009, 15). The covert data flows and undisclosed algorithmic practices also marginalize public debate and hinder informed policy making. To help open these technological changes (and informational exchanges) to scrutiny, this chapter looks at how CCTV technologies are being transformed and integrated with information communication technologies (ICT) and signal processing algorithms to expand the scope and nature of video surveillance well beyond CCTV.

We begin by highlighting the central features of the conventional CCTV model and then show how “smart” or “intelligent” digital video surveillance aims to overcome the limitations of this model.

We review the enhanced functionalities of digital video techniques, especially video analytics, and how they are being applied in a growing range of surveillance settings. This chapter seeks to contribute to the broader discussion of what is at stake at a time of intensifying surveillance regimes both in North America and around the globe (see this volume). We are particularly interested to understand how camera surveillance augmented by video analytics is being deployed and what this deployment may generally mean for individual citizens and the public interest. The resulting better understanding of the implications for civil liberties and public policy of the new generation of video surveillance techniques may help redress the growing imbalance between the watched and the watchers, especially in 'open-street' public surveillance in retail contexts.

The CCTV Paradigm

CCTV represents the first generation of visual surveillance. It was first used by German scientists in 1942 for remote monitoring of rocket launches. (Reuter, 2000). At its most basic, a CCTV system is a camera linked to a monitor. Connected by a continuous or closed electronic circuit, the camera captures images that are then transmitted to the monitor. In this model the information flow is uninterrupted and unidirectional. The closed path from camera to monitor implies an indexical relationship between the surveyed space and the image—a visual truth that hides the mediation imposed by the technology of the medium, and the viewer from the viewed. This visual, photographic truth (Tagg, 1988) combined with the inherent one-way flow of information, supports institutional infrastructures of knowledge production that reinforce dominant social power.

The one-camera one-monitor closed-circuit model was readily adapted for surveillance of multiple spaces or for multiple camera viewpoints. The first 'open-street' deployment came in 1969, when the London Metropolitan Police used two temporary cameras in Trafalgar Square to monitor Guy Fawkes Day Activities (Norris and Armstrong, 1999). Over the next two decades CCTV systems evolved to include more than just cameras and monitors. To facilitate multiple camera inputs and viewing options additional devices, such as switchers, controllers and recorders were integrated. Figure 1 presents a typical CCTV system model including multiple cameras, a switcher/controller and a video recording device.

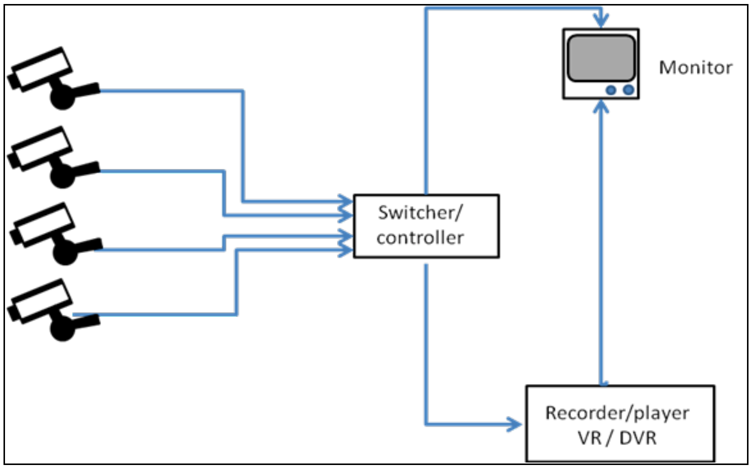


Figure 1: The Classic CCTV Model

This conventional CCTV model has one or more cameras that feed into a switcher or controller. Controllers may be used to direct PTZ (pan, tilt & zoom) cameras, allowing operators to change the field of view remotely. Information then flows from the switcher to a monitor for real-time remote viewing and a recorder for future viewing. The two key features of the system are that it is closed, so that no information is broadcast elsewhere, and that it represents images using analog signals. Analog signals or information means that image elements are coded by continuously varying properties of the physical medium, such as voltage in the transmission circuit and magnetic field strength in the case of recording tape. Analog signals also mean that images often have low resolution and quality, and are prone to various technological errors. Analog signals, compared to current technologies are also relatively hard to process, so that doing more than sequential viewing and storing involve complex customized hardware. So another limitation becomes the difficulty in handling the flood of visual information generated. Human operators have limited attention spans. The more cameras, the more information is produced, the more the switcher moves between channels, and the more information remains unseen and unrecorded.

Beyond Analog CCTV

Since the 1990's the adoption of digital techniques for handling visual data has facilitated the transition from analog cameras and closed-networks toward digital image capture, transmission, storage, analysis and encoding. Digital recording of multiple cameras on hard drives greatly reduces the costs associated with recording and storage. A contemporary video surveillance network model can receive signals from various camera types: analog or digital, with fixed views or pan, tilt and zoom capabilities. Surveillance networks may also incorporate multiple compression, transmission and encryption standards; incorporate local or networked video recording (NVR) and storage; incorporate digital signal processing (DSP); and employ computer-mediated real-time analytics. As internet protocol (IP) standards are increasingly adopted, individual components and entire networks can be readily linked together to aggregate, exchange or repurpose information over the internet. Figure 2 demonstrates some variations on information processing associated with more sophisticated digital video surveillance systems.

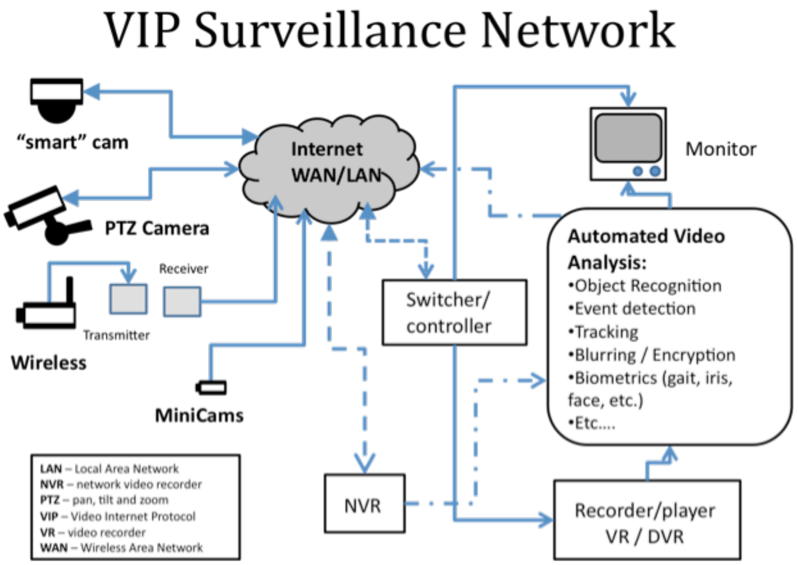


Figure 2: Video Internet Protocol (VIP) Surveillance Network

Although a wired CCTV camera is by no means obsolete, increasingly cameras are digital and information is at some point transmitted using wireless networks and/or over the internet. The switch to digital sensors provides the foundation for "smart" security—that is, security augmented by computer-mediated processing. Contemporary digital sensors are typically in the range of several megapixels, but surveillance cameras with 50 megapixel resolution and higher are commercially available. The imaging technology is improving rapidly: in 2010 the Department of Homeland Security announced development of a 100 megapixel camera with a specialized fish-eye lens for 360 degree surveillance (2010). The potential resolution or amount of information gathered using these digital sensors is significantly greater than analog cameras. To make this point more clear consider a face recognition system that uses images as small as "an average face size of 75 pixels between the centers of the eyes" (Phillips et al., 2007) for identification. For analog CCTV images, this represents roughly a fifth of a standard TV image. With digital images, where the resolution is measured in megapixels, a 75 pixel average distance between pupils can represent a much smaller fraction of the information captured in a single image. So a face in a digital frame with high-resolution imaging has more possibilities for cataloguing, analysis and identification than an analog image with the same field of view. The amount of data is not all that has changed: the quality of that data too has seen significant improvements.

Digital video signals also tend to have better and more consistent image quality than analog CCTV. Digital capture and recording eliminates many of the errors characteristic of analog video recording. Time-based errors occur because of the mechanical instability of video recording devices—for example the slight difference in tape speed between the beginning of a cassette and the end of one—making mixing or comparison of video difficult without correction. Dropout errors derived from the recording device produce blurred, skewed or blotchy images when using analog video recordings. Digital video recorders, although they present new challenges, virtually eliminate time-based and dropout errors. Finally, digital video eliminates generation loss (Watkinson, 1994). With magnetic tape-based analog video recorders, each copy—or each generation—introduces noise that degrades the original signal. Since digital recordings may be reproduced without generation loss, digital surveillance video does not fall prey to the same types of information degradation from being copied and distributed multiple times for multiple viewings at multiple sites.

The transition to digital technologies represents a significant change in the development of video surveillance—a kind of renaissance in quality and a movement away from strictly closed-circuit systems. The move to digital means information can exist in multiple archives and in multiple formats over multiple networks—allowing for unprecedented distribution, manipulation, aggregation and sharing. This movement towards more open, networked, heterogenous infrastructures of surveillance marks a notable shift in how visual information is exchanged, stored, indexed and analyzed. At the extreme end, PTZ cameras are being controlled and monitored by geographically distributed people, either with access permissions, or via interactive broadcasts entirely open to all visitors. An example of this is the Texas Sheriff’s strategy for border control (Reay, 2008). Roaming video surveillance cameras were deployed so that the public could voluntarily monitor the Texas border for illegal migration from an online site. Not all open video surveillance feeds are advertized, but many are openly broadcasting, some unwittingly—a search for ‘"ViewerFrame?Mode="’ produces a horde of IP addresses openly streaming surveillance images that often include offices and commercial spaces. These examples of video surveillance encoded into packets and accessible either by password protected sites or as open video, can no longer be understood as CCTV. With commercially available cameras increasingly ‘internet ready’, this decoupling of camera from monitor and from storage or recording device means that visual information can travel across the globe for analysis or retention. This shift in information flow is typically understated when CCTV cameras are discussed and has significant implications for privacy, security and democratic rights of all surveilled subjects.

As challenging to privacy and public policy as these changes in the flow of visual information represent, they only pave the way for more and different challenges that algorithmic processing of that information will likely bring. The ability to integrate video analysis software into multiple nodes of a surveillance network represents a shift toward what its promoters refer to as, “smart” visual surveillance. If analog CCTV represents the first generation, then the so called ‘smart’ surveillance ‘solutions’, that integrate video analytics with digital storage and networking, are clearly a new generation of video surveillance.

Hidden Processing

Increasingly, surveillance systems integrate one or more layers of analytics that mediate the flow of information between the operator(s) and the camera(s). This intermediate video processing is most commonly referred to as Video Analytics (VA), though vendors also use terms such as smart video surveillance, Intelligent Video Analytics (IVA), Intelligent Analytics (IA), Video Content Analysis (VCA) to describe a range of video image processing techniques. Video Analytics incorporate a range of functionality “from systems that classify and store simple data, through more complex systems that compare the captured data to other data and provide matches, to systems that attempt to predict events based on the captured data” (Norris and McCahill, 2006). This digital processing enables a form of algorithmic surveillance (Introna and Wood, 2004) that is hidden from even the most observant surveillance subject.

Programming computers to detect, classify, code and semantically label temporal and spatial patterns of pixels in a video stream is a complex problem. Video analytic architectures tend to be modular -- that is a series of software modules strung together -- where outputs of one module are used as inputs for the next. Modules are typically independent but use common protocols throughout a processing chain. There are distinct algorithms for detecting, classifying and identifying objects and their trajectories as well as for initiating responses. Generally an early processing stage includes a module that attempts to distinguish and track objects across successive video images. Differences between frames can help demarcate moving objects in the foreground from the background. Once detected, the background is subtracted, or discarded, so subsequent object and event detection modules can process the pixels that remain, for instance by comparing them with patterns in earlier frames or pre-defined reference models.

The image information can then pass on to object and event classification modules where patterns are classified and alerts triggered automatically. The criteria for alerts are based on predefined or user-specified models for various types of objects and events, such as the presence of a gun, a person running or an unexpected package. In the classification stages video analytic engines produce meta-data that both guides the visual processing and provides semantic descriptions of the video inputs. This classification of objects and events can also produce an index of objects and recognized events, which in turn can be used as meta-data for future search and retrieval and as data for further processing. For example, objects recognized as humans can be characterized, measured, compared and slotted into generalized categories (male/female, white/asian/… etc.), and profiled to assess risk in both security and retail domains, and opportunities in forms of targeted marketing and other forms of intervention. Cisco Systems Inc., encourages their retail store customers to monitor shoppers based on a range of demographic criteria, potentially influencing how customers are sorted and treated. In a pamphlet aimed at retailers, Cisco advertises that “cameras can also identify gender and age, and send out real-time alerts [and]…. when combined with an IP network, video… can be easily retrieved, shared, integrated, and analyzed within the store and across the enterprise.” (2009b).

Based on the class of object recognized—e.g. a vehicle, body, or face—other algorithms may be employed to identify objects individually and track them between frames and across different video streams. If an object is classified as a license plate, for example, automatic number plate recognition (ANPR) (also known as automatic license plate recognition (ALPR)) software can then be used to correct any distortion and apply optical character recognition (OCR) to identify the registration.

An object determined to be a body can be analyzed for individual identity or behavioural pattern (e.g. ‘loitering’). Though perhaps further off than ALPR in terms of routine adoption, but at least as challenging with respect to policy and individual rights, are the analytic techniques for behavioural analysis and biometrics such as gait or face recognition. Tracking and identifying people in video is an active area of research, development and commercialization (Cai et al., 2009, Bojkovic and Samcovic, 2006, Gorodnichy, 2006, Shaokang et al., 2008, Suman, 2008, Goffredo et al., 2008). It is driven by the hopes that video analytics will enable video surveillance networks to follow individuals, detect anomalies, anticipate potential danger and take appropriate action, all in real-time. This convergence of biometric identification infrastructures, notably those based on facial images, and VA represents a new horizontal integration of contemporary video surveillance systems. See Figure 3. High resolution surveillance cameras supported by digital processing linked to private, public or government biometric databases may significantly extend our already extensive data shadows into public spaces and subject us to profiling, risk assessment, and proactive interventions.

<INSERT FIGURE 3 AROUND HERE>

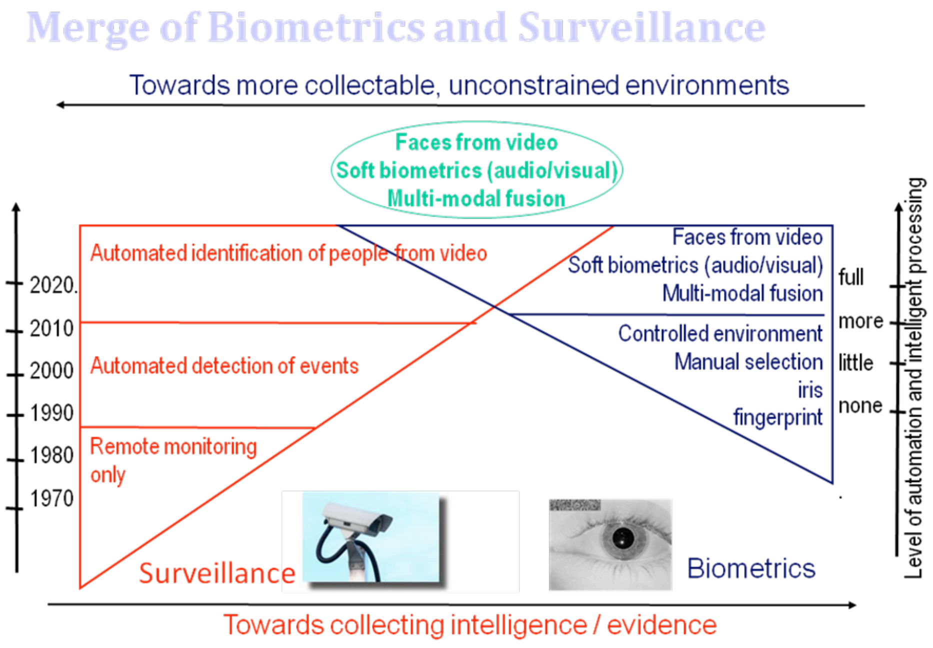
Video Analytics (VA) addresses at least two major limitations of the conventional analog CCTV model: live monitoring and retrospective searching. Watching video surveillance can be tedious and boring. Often there can be hours or days of video from multiple sources where very little of interest actually happens. The volume of information produced by multiple cameras running 24 hours seven days a week means that much of the information captured by analog CCTV cameras is not viewed in real-time or retained, and if recorded, remains effectively not viewed. Bosch Security Systems presents the problem this way:

> No matter how few or how many cameras your system uses, monitoring everything effectively presents a serious challenge. Even observing just a single screen for long periods pushes concentration to the limit -- after only 20 minutes, an operator can miss as much as 90% of the activity in a scene (2009a).

Video Analytics is presented as a promising solution. Processing the video input can help prioritize information in real-time and bring significant events to the foreground for human agents that require intervention or further monitoring. Bosch Security Systems offers Intelligent Video Analysis (IVA) to help "operators focus by introducing a new level of automation to CCTV monitoring" (2009a). They bill their system as an extra set of unblinking eyes that help operators to stay focused and on top of the action. Video analytics introduces "...a new level of automation to CCTV monitoring. Edge-based, real-time processing identifies alert conditions, giving your security team the information it needs to react swiftly and take action." (2009a).

Video Analytics also potentially addresses another significant limitation with conventional CCTV surveillance: sequential, time based information access, requiring linear searches through recordings.

With standard recording equipment, video information could not generally be accessed in other than a linear fashion, say by event type. This makes reviewing CCTV footage time consuming and labour intensive. Video analytics tools can recognize and catalogue objects and events to produce a semantic index of visual data, which facilitates more advanced processing and analysis – e.g. enabling searches by object, gender, age, habit, predefined event, etc,. This more sophisticated video search technique has multiple applications across many domains.


Merge of Biometrics and Surveillance
Towards more collectable, unconstrained environments
Faces from video
Soft biometrics (audio/visual)
Multi-modal fusion
2020.
2010
2000
1990
1980
1970
Automated identification of people from video
Automated detection of events
Remote monitoring only
Surveillance
Faces from video
Soft biometrics (audio/visual)
Multi-modal fusion
Controlled environment
Manual selection
iris
fingerprint
full
more
little
none
Level of automation and intelligent processing
Biometrics
Towards collecting intelligence / evidence

D. Gorodnichy © 2009

Figure 3: Towards "smart" surveillance; Merging of biometrics and surveillance techniques (Gorodnichy, 2009)

This shift to "smart" video surveillance is being made almost entirely out of public view. The transition is hidden in the sense that existing infrastructures can be upgraded with video image processing software without any overt or visible change to existing cameras. In fact, this is a selling point. IBM's Smart Surveillance Solution is designed to be interoperable and "integrate with existing video and capture systems to provide video/sensor analytic capabilities" (Sullivan, 2008). So once the visual images are captured they can be remotely stored, indexed and analyzed for what IBM refers to as "valuable business information."

Applying video analytics

VA is being applied in a variety of surveillance contexts, including such broadly defined domains as: public safety, security, border control, finance and retail data-mining. For public safety and security VA may be used to identify and report people who need emergency medical attention: a person who collapses in a parking lot for instance. For border control, video analytics are being employed at airports to 'detect' abandoned or suspicious items.

The business opportunities that come from applying computer vision techniques in everyday settings have attracted the interest of hundreds of companies. There are large integrators like IBM, Bosch and Cisco who offer comprehensive 'solutions' to surveillance, security, and visual data mining; and smaller companies like IQeye, Smart Camera Systems, VideoIQ Inc., and 3VR that specialize in specific areas of video analytics. The variety of vendors means that there are a broad range of customizable systems designed to trigger automatic and often, user-defined, alerts. As a result, the application of VA in video surveillance can vary widely, including counting the number of cars that use a road way, as part of traffic safety (Coleshill et al., 2007, Pflugfelder et al., 2005), investigating banking fraud and conducting market research.

In financial institutions "smart" cameras are used to assist forensics investigations. Automated analysis of ATM video images can provide legal evidence while reducing investigation costs (Russo, 2008). In retail spaces, cameras pointed at employees may be looking for transaction fraud; while the cameras pointed at customers may help analyze shopping habits through digital signage networks (DSN) (Dixon, 2010). DSN are a growing phenomenon where stores employ cameras embedded in digital signage to record potentially useful behavioural information. Video analytics can report how long a customer interacts with the sign and demographic information about the customer to establish customer profiles. These cameras can be hidden behind video screens, and retailers generally provide little notice about whether they employ this form of video capture and analysis (Dixon, 2010). With digital video surveillance infrastructures in place, and as the utility of mining visual information increases, the temptation to gain valuable insights into customer patterns will likely lead to this becoming increasingly commonplace.

As a vendor of integrated systems, IBM claims benefits to their customers can include: cross-agency sharing; remote access; reduced security overhead; reduced need for guards; and nonlinear search capabilities. Facilitated by video analytics, nonlinear search capabilities allow for “queries based on color, object, size and other attributes” (Sullivan, 2008). According to IBM, this type of fine grained analytics will help its clients benefit by “getting business information out of security-related data”(Sullivan, 2008). Repurposing security surveillance video data by means of video analytics allows retailers to potentially mine and capitalize on the behaviours and patterns of both clients and staff. It also threatens to erode the principles of purpose specification, use limitation and data minimization that underpin all privacy regulation.

The events of 9/11 provided a brief but important impetus to the development of video analytics. Promoters of the latest generation of video surveillance were quick to promote its virtues in stopping future terrorist attacks. One of the most prominent of these was Dr. Joseph Atick, co-founder of Visionics Corporation (CNN.com, 2001), but despite his confident claims, his Face-it technology proved much less than ready for deployment during tests at Logan International Airport a year later (Murphy and Bray, 2003). Failures at other sites like Tampa Florida (Stanley and Steinhardt, 2002) contributed to the reading of automated systems as expensive, prone to false alarms, hard to configure, unreliable and easy to fool. For example, researchers demonstrated multiple ways to circumvent face recognition technologies (Alexander and Smith, 2003). However, although the initial roll out of video analytic systems following 9/11 fell prey to over-stated claims and poor system performance (BSIA, 2010), video analytics has developed apace. With modest (or at least less easily refutable) claims and expanded capabilities, video analytics is finding wider markets.

As research continues to improve real-time video signal processing and analysis, images captured by video surveillance may increasingly serve as gateways to our identities. At a time when visual information is becoming increasingly personal, linkable, mobile, searchable, and thereby valuable, images caught on surveillance cameras pose an increasing threat to privacy and identity integrity. Collectors and custodians of visual information should adhere to fair information practices. They should understand the capacities and limitations of their systems, keep to justifiable purposes, and provide information on their information practice policies involving any form of automated processing, storage or transmission of personal information. This ideal remains elusive.

Hidden Alternatives

The digital techniques for video image capture, storage, networking, and algorithmic analysis discussed so far are aimed at enhancing the discriminatory capabilities of surveillant authorities, but with no corresponding reciprocal visibility or accountability. In this way such techniques contribute to further imbalance established relationships of power away from surveillance subjects and toward those deploying the video surveillance. But while the asymmetric "big brother", panoptic and one-way metaphors that dominate the discussion of the social implications of video surveillance have a valid basis and help inform resistance to the more egregious forms of surveillance, they also tend to obscure the possibilities that digitalization can offer for opening up alternative forms of surveillance that might help mitigate some systemic imbalance. Can the techniques of digital signal processing, object coding, and video analytics that surveillance operators deploy for their advantage also be used on behalf of surveillance subjects? Can alternate forms of surveillance be used to demonstrate privacy enhancing surveillance strategies? At least technically, and under favourable conditions, the answer appears to be 'yes' – they can help protect personal privacy and enable greater public accountability and oversight of legitimate surveillance operations.

An obvious approach to privacy protection offered by video analytics is through the anonymization of personal image data. Most straightforward technically is encryption, or the systematic mathematical scrambling of the digital data. Only with the appropriate key can the encrypted images be viewed. This is currently routine in many uses of wireless cameras to prevent surreptitious eavesdropping of the transmission from the camera to its base station. However, once the video signals reach the more protected wired circuits and storage devices, typically they are de-encrypted and from then on appear in plain form. Where video surveillance is used exclusively for forensic purposes, the stored images could remain encrypted until law enforcement authorities obtain an authorized key to decrypt the specific time period when a serious incident is strongly suspected to have occurred. The keys could be held by the judiciary, and only provided upon legitimate formal request, such as a search warrant submitted to a judge. The computer system hosting the video image database could also automatically log and regularly report on such key release authorizations and decryptions, including such details as the date, time, rationale and authorizing body. Even when there is a need for viewing the live video stream, all subsequent storage and viewing could be regulated in this fashion.

Another approach to anonymization is through the blurring of specified video objects. The same techniques for detecting objects, such as faces, bodies and licence plates, in the video stream for subsequent analysis and recognition can also be used to blur them. Once spotted, these objects can be permanently smudged by randomizing the visual information. The software for accomplishing this is simpler than that required for the enhancement and analysis techniques mentioned above. The best known example is Google Blur, the software used by the search engine giant to smudge the faces and licence plates in Google Street View images. Of course the success of this technique depends on the accuracy of the underlying object recognition algorithms—and mistakes have been well documented in the media. Also, if the blurring is not performed in the camera, but later in the process, it would be possible to retrieve the un-blurred images.

Encryption can be combined with object detection to produce a broader range of anonymization options. Researchers at the University of Toronto have developed a technique they refer to as secure visual object coding. The technique selectively encrypts particular types of objects, such as faces, and treats them separately from the rest of the video stream (Martin and Plataniotis, 2008). In normal live viewing the background would appear as usual, but with the protected objects appearing in outline form with a randomized texture. See figure 4.

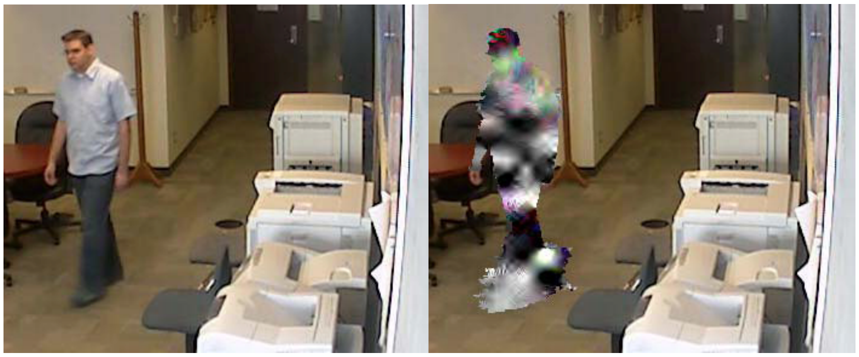


Figure 4. Demonstration of secure visual object coding technique, showing original surveillance frame and with encrypted object texture (Martin and Plataniotis, 2008).

Once the images are encrypted in this way, it is very difficult for anyone without proper authorization to view the full video image. This could enable live monitoring, say of subway platforms or public places, but without revealing individual identities. Access to personal information can then be controlled through a rigorous administrative and technical regime. Secure visual object coding is still in experimental development and testing, but it already illustrates the potential for approaches that exploit digital characteristics for privacy protection. In particular, it can help reinforce the data minimization principle by setting the standard for the amount of personal data used for security and other purposes at a much lower level. However, it is unlikely that this or other similar techniques will be adopted and implemented without strong policy, regulatory or other forms of external pressure.

Furthermore, even if "improved" privacy-protective surveillance techniques become prevalent, they alone cannot address all risks posed by the shift to digital video surveillance. The following section explores the implications of this asymmetric and differential visibility in terms of everyday identification and surveillance. This analysis highlights the need to mitigate this asymmetry and the challenge of making the surveillance infrastructure and the propagation of faces and other personal visual information collected through it more visible and hence more accountable.

Social Implications for Surveillance Subjects

To this point we have concentrated on how capabilities of video surveillance technologies have grown since the 1990s, mainly in response to the demands for enhancing security and the increased availability of flexible, inexpensive, interoperable digital hardware and software. We turn now to address issues of what these changes mean for surveilled subjects' well-being and civil rights.

It hardly can be claimed the central challenges posed by video surveillance, even in its simpler CCTV form, have been adequately addressed. Lingering questions remain: Is video surveillance effective, and under what circumstances? For what purposes is it, or should it be put? Who is looking at my live and recorded images? Who is doing what, and why, with my personal information and how will this affect me? All remain largely unanswered. The on-going transformation of video surveillance technologies and practices generally make these questions even harder to answer. This further suggests that the well-founded concerns for civil liberties associated with conventional CCTV are likely to be exacerbated rather than mitigated by digital growth. Concerns over reasonable expectations of privacy in public spaces, already an issue identified by the results of the Urbaneye project (Hempel & Töpfer, 2004), will likely be exacerbated by digital repurposing of security information for other forms of visual data mining.

Technological changes alone never produce specific social outcomes. Organizational, cultural and other factors also play important roles. However, it is useful to consider how the adoption of digital video surveillance techniques creates new, socially significant affordances and constraints. Digital technologies typically greatly reduce the component costs for the capture, transmission, storage and display of video surveillance images. Even if the conventional CCTV model remained un-changed, this would encourage the wider adoption of video surveillance for its original security purposes and make it more attractive for other purposes, such as observing consumer behaviour. At the very least these lower costs mean more people would be subject to its potentially intrusive, chilling and privacy-invasive effects.

Conventional CCTV cameras, with their recognizable housings and prominent public placement, are becoming almost invisible through their ubiquity in contemporary urban landscapes. The much-reduced size of digital cameras makes video surveillance easier to hide and further obscures it from public view. But even when the cameras appear in their familiar bulky form, they are only one part of a much larger surveillance assemblage, in which the growing number and variety of actors, human and otherwise, remain unseen. We have noted that digital technologies now make it possible to connect large arrays of cameras together in wide area networks, to disseminate the signals widely, and to archive unprecedented volumes of video material in multiple remote locations for later analysis and viewing. This enlarges the potential number of viewers and distributes them across space and time. With conventional CCTV systems it was hard enough for a determined person to find out who was watching her, where her personal images were stored and for how long. This basic openness requirement for privacy protection and democratic accountability more generally is now decidedly more difficult to achieve with networked systems - unless surveillance operators become more forthcoming than previously. So far there is no sign of this.

But the most significant social challenges that digital, analytic surveillance pose comes from the enhanced capabilities that digitalization offers for automated image analysis and subsequent decision-making. The step of transforming the visual image stream into data that can be processed algorithmically opens a vast new array of surveillance possibilities. Some are relatively simple and benign, such as detecting litter so cleaners can be dispatched to remove it. When people become the subject of algorithmic gaze, the risks magnify. The more rudimentary forms of detection rely on comparison to "normal" appearance and behaviours, so deviations from the norm invite scrutiny. Walking too slowly may be interpreted as loitering, walking too quickly as fleeing. Carrying or wearing something unusual can be treated as suspicious. Some techniques are aimed at classifying surveillance subjects – particularly ethno-racially and socio-economically - based on visual data. Introna and Wood (2004) in particular highlight the ethnic biases of algorithmic surveillance and face recognition technologies. They further note that face recognition systems are a "particularly good example of a silent technology" (Introna and Wood, 2004) -- a technology whose significance and application are often hidden within existing infrastructures, such as ID card schemes and CCTV networks.

It is this potential to connect live video surveillance with existing digital records about individuals which raise the greatest long term societal concerns because of the potential for continuous, ubiquitous, surreptitious, fine-grained surveillance of whole populations. The hitherto separate surveillance infrastructures of video and administrative data are now on the verge of being brought together via the recent growth of facial image databases running facial recognition software for managing ID documents such as drivers licences, health cards and passports. As these databases go on-line, it will be increasingly tempting for law enforcement and other agencies to search them using visual images captured in the field. There are already signs of this in the call by Ontario's police chiefs "to give officers access to a database of driver's licence photographs to help instantly verify the identity of suspects and traffic accident victims." (Globe and Mail, 2010). In this case the images would be stills, but if they are successful with this request, can the next step of pulling images from live video feeds be far behind?

The failures of the face recognition field trials following 9/11 illustrate the immaturity of the video analytics field. While progress has been made in improving the performance of video analytic software, there are some inherent technical limitations, especially in the area of facial recognition that are likely to persist. The complexity of the algorithms and the overall inscrutability of the increasingly complex networked configurations make it difficult even for technical experts, familiar with the systems, to assess whether they are functioning reliably and effectively. In the face of the obvious short comings of conventional CCTV, and promise of video analytics to overcome them, the incentive to install sophisticated albeit faulty systems, before they are field proven, could be irresistible. As long as the failures remain out of public scrutiny, the allure of "smart" new techniques will remain bright and their deployment will multiply with unknown consequences.

The deviations, classifications and identifications produced by correctly functioning automated surveillance schemes are typically of little consequence, but sometimes can fit threat profiles that trigger discriminatory, and sometimes forceful, responses. In forensic applications, automated techniques can be a boon for investigators searching through large volumes of recordings searching for particular patterns. In such situations there will be usually the opportunity for competent investigators to probe more deeply, and discard the "false positives" before any harm is done. However, in real-time monitoring settings, especially when there is strong pressure to intervene proactively to preserve public safety, there will be little such opportunity for reflection, possibly resulting in injurious over-reaction. While the widely reported cases of police killing innocent people based on mis-interpreting visual and other cues (e.g. the Robert Dziekański Taser incident (2010) in the Vancouver airport and the London shooting of Jean Charles de Menezes (2010) following the July 7 2005 bombing) are extreme examples and so far appear not to involve video analytics, they illustrate the possible severity of outcomes when relying on automated threat profiling in security operations.

Conclusion

In the four decades since video surveillance techniques have been deployed in public settings, the key technologies have undergone a revolution—a paradigm shift from analog remote viewing to computer-mediated content analysis. As part of the wider digital transformation of information and communications technologies, video surveillance capabilities have expanded enormously in terms of accuracy, speed, storage capacity, image analysis, database linkage and identity recognition. While these developments are largely invisible to surveillance subjects, the behind the scenes changes are so significant that they call into question the assumptions based on the 'classic' CCTV model - about who or what is watching, for what purposes and with what consequences. Driven by an apparently insatiable demand for more detailed personal information in the pursuit of "security" and economic growth, the scope and intensity of video surveillance activities are raising qualitatively new challenges, far outstripping our understanding of its effectiveness, limitations and implications for civil liberties. Lagging further still is the institutional policy response, which is inadequate for regulating even the old style CCTV deployments.

The growing scope and depth of digital video analytics needs to be considered thoroughly as the technologies develop and their applications ramify. We need new understandings and attitudes towards how visual data are or should be stored, transmitted, analyzed and acted upon. Under most privacy law, images captured by organizations of identifiable individuals, whether actually identified or not, constitutes personal information and should therefore come under some level of legal protection. So far, there is little sign that current measures are effective in this regard. When individuals become routinely identified through biometric techniques such as iris, gait or face recognition and without their knowledge, the privacy stakes are much higher. But how will the organizations responsible be brought to account? Who is in a position to comprehend and intervene appropriately is overseeing this emerging phenomenon?

If the current lax to non-existent oversight of video surveillance prevails during the current phase of rapid but largely hidden digitalization, then the civil liberties risks will continue to grow while becoming harder to rectify.

Acknowledgement(s): We appreciate the support of the Identity, Privacy and Security Institute, University of Toronto

Works Cited:

Ontario police forces seek access to driver’s-licence photos - The Globe and Mail. Available at: http://www.theglobeandmail.com/news/national/ontario-police-forces-seek-access-to-drivers-licence-photos/article1564150/ [Accessed May 10, 2010].

(2009a) CCTV - Bosch Security Systems The Americas. Bosch Security Systems

(2009b) Video Analytics: Enabling Retailers and Casinos To Build Brand Loyalty, Increase Competitive Advantage, and Improve Customer Service IN CISCO SYSTEMS, I. (Ed.). Cisco IBSG.

(2010) New video camera offers very high resolution from afar -- and up close. Homeland Security News Wire.

ALEXANDER, J. & SMITH, J. (2003) Engineering privacy in public: Confounding face recognition. *Privacy Enhancing Technologies.*

BOJKOVIC, Z. & SAMCOVIC, A. (2006) Face Detection Approach in Neural Network Based Method for Video Surveillance. *Neural Network Applications in Electrical Engineering, 2006. NEUREL 2006. 8th Seminar on.*

BSIA (2010) BSIA reports an increase in the use of Video Content Analysis.

CAI, Y., KAUFER, D., HART, E. & SOLOMON, E. (2009) Semantic Visual Abstraction for Face Recognition. IN ALLEN, G., NABRZYSKI, J., SEIDEL, E., DONGARRA, J., VANALBADA, G. D. & SLOOT, P. M. A. (Eds.) *Computational Science - Iccs 2009, Part I.*

CNN.COM (2001) How the facial recognition security system works

COLESHILL, E., FERWORN, A. & STACEY, D. (2007) Traffic Safety using Frame Extraction Through Time. *System of Systems Engineering, 2007. SoSE '07. IEEE International Conference on.*

DEISMAN, W., DERBY, P., DOYLE, A., LEMAN-LANGLOIS, S., LIPPERT, R., LYON, D., PRIDMORE, J., SMITH, E., WALBY, K. & WHITSON, J. (2009) A Report on Camera Surveillance in Canada Part One. Surveillance Camera Awareness Network (SCAN)

DIXON, P. (2010) The One-Way-Mirror Society: Privacy Implications of the new Digital Signage Networks. World Privacy Forum.

GOFFREDO, M., CARTER, J. N., NIXON, M. S. & IEEE (2008) *Front-view Gait Recognition*.

GORODNICHY, D. (2006) Editorial: Seeing Faces in Video by Computers. *Image and Vision Computing* 551-556.

GORODNICHY, D. (2009) Merge of Biometrics and Surveillance. *PowerPoint.* Toronto, IPSI.

HEMPEL, L. & TÖPFER, E. (2004) Working Paper No. 15: CCTV in Europe Final Report. Berlin, Centre for Technology and Society, Technical University Berlin.

INTRONA, L. D. & WOOD, D. (2004) Picturing Algorithmic Surveillance: the politics of face recognition systems. *Surveillance & Society,* 2**,** 177 -198.

JEAN CHARLES DE MENEZES (2010, July 7) In *Wikipedia, the free encyclopedia*. Retrieved July 7, 2010, from http://en.wikipedia.org/wiki/Jean_Charles_de_Menezes

MARTIN, K. & PLATANIOTIS, K. N. (2008) Privacy Protected Surveillance Using Secure Visual Object Coding. Toronto, Multimedia Laboratory.

MURPHY, S. & BRAY, H. (2003) Face recognition devices failed in test at Logan. *Boston Globe.* Boston.

NORRIS, C. & ARMSTORNG, G. (1999) *The Maximum Surveillance Society: the Rise of CCTV,* Oxford, Berg.

NORRIS, C. & MCCAHILL, M. (2006) CCTV: Beyond penal modernism? *British Journal of Criminology,* 46**,** 97-118.

PFLUGFELDER, R., BISCHOF, H., DOMINGUEZ, G. F., NOLLE, M. & SCHWABACH, H. (2005) Influence of camera properties on image analysis in visual tunnel surveillance. *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE.*

PHILLIPS, J., SCRUGGS, T., O'TOOLE, A. J., FLYNN, P. J., BOWYER, K. W., SCHOTT, C. L. & SHARPE, M. (2007) FRVT 2006 and ICE 2006 Large-Scale Results. Gaithersburg, National Institute of Standards and Technology.

REAY, D. L. (2008) Texas Border Sheriff's Coalition institute border surveillance cameras., Sheriffs News.

REUTER, C. (2000) *The V2 and the German, Russian and American Rocket Program,* New York, S. R. Research and Publishing.

ROBERT DZIEKAŃSKI TASER INCIDENT (2010, June 30) In *Wikipedia, the free encyclopedia.* Retrieved July 7, 2010, from http://en.wikipedia.org/wiki/Robert_Dzieka%C5%84ski_Taser_incident

RUSSO, S. (2008) Digital Video Surveillance: enhancing physical security with analytic capabilities. IBM Corporation.

SHAOKANG, C., BERGLUND, E., BIGDELI, A., SANDERSON, C. & LOVELL, B. C. (2008) Experimental Analysis of Face Recognition on Still and CCTV Images. *Advanced Video and Signal Based Surveillance, 2008. AVSS '08. IEEE Fifth International Conference on.*

STANLEY, J. & STEINHARDT, B. (2002) *Drawing a Blank: The failure of facial recognition technology in Tampa, Florida* AN ACLU SPECIAL REPORT. American Civil Liberties Union.

SULLIVAN, F. (2008) Leveraging security data for business intelligence: IBM Smart Surveillance Solution. Frost & Sullivan.

SUMAN, A. (2008) Using 3D pose alignment tools in forensic applications of Face Recognition. *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems.* Washington DC, IEEE Xplore.

TAGG, J. (1988) *The Burden of Representation: essays on Photographies and histories,* Massachusetts, University of Massachusetts.

WATKINSON, J. (1994) The future of digital video recorders. *International Conference on Storage and Recording Systems, 1994.*