RC22886 (W0308-109) August 28, 2003
Computer Science

IBM Research Report

Blinkering Surveillance: Enabling Video Privacy through Computer Vision

Andrew Senior, Sharath Pankanti, Arun Hampapur, Lisa Brown, Ying-Li Tian, Ahmet Ekin
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Blinkering Surveillance: Enabling Video Privacy Through Computer Vision

Andrew Senior, Sharath Pankanti, Arun Hampapur, Lisa Brown, Ying-Li Tian and Ahmet Ekin
{aws,sharat,arunh,lisabr,yltian,aekin}@us.ibm.com
IBM T. J. Watson Research Center,
PO Box 704, Yorktown Heights, NY 10598

Abstract

In this paper we describe a technology for protecting privacy in video systems. The paper presents a review of privacy in video surveillance and describes how a computer vision approach to understanding the video can be used to represent "just enough" of the information contained in a video stream to allow video-based tasks (including both surveillance and other "person aware" applications) to be accomplished, while hiding superfluous details, particularly identity, that can contain privacy-intrusive information. The technology has been implemented in the form of a privacy console that manages operator access to different versions of the video-derived data according to access control lists. We have also built PrivacyCam*—a smart camera that produces a video stream with the privacy-intrusive information already removed.*

1 Introduction

In recent years we have seen a world-wide rise in the use of Closed-Circuit Television (CCTV) cameras, and are now beginning to see a corresponding rise in video processing systems that can interpret the video, mining information using computer vision algorithms to extract usable data such as movements, identities and event times from the raw video. While CCTV systems have typically been used for surveillance, low-cost cameras are enabling a wide range of other applications that will involve cameras being located on devices, in buildings and public spaces. Already, with CCTV systems monitored by human operators, unobtrusive or deliberately hidden cameras are used to spy on people and for voyeurism. Video surveillance can readily be used as a tool for state control and oppression. When the algorithms currently under development mature and achieve human-like efficiency, their sheer scale will immeasurably increase the power of CCTV systems in benign and malign applications.

The Pandora's box of automated video surveillance is already open, but among the technologies being developed we find also the hope for controls on the negative uses of video surveillance. Algorithms similar to those used to extract data from the raw video can be used to filter that same video, altering it and restricting the amount of privacy-intrusive data contained in the video, while preserving enough information to be useful for the original task. We hope that, in combination with whatever social and legal controls may be applied to prevent oppressive surveillance, these technological methods can be applied to "blinker big brother" and restrict the capabilities of CCTV to intrude on privacy.

This paper begins (Section 2) by describing the rise of video surveillance, and depicting public concerns over its widespread deployment. Section 3 describes what we mean by video privacy, and gives some examples of how it is treated the same as, or differently from, data privacy. Section 4 outlines a model for video privacy, followed by a description, in Section 5, of the system we have built for the preservation of video privacy and the PrivacyCam that extends this model to a standalone device. Section 6 discusses how effectively such a system can be expected to operate in practical terms and in terms of trustworthiness of implementation and public confidence.

2 The rise of video surveillance

Video surveillance is becoming ubiquitous in urban life. Video cameras are being installed in urban areas throughout the developed world, intended principally as a deterrent to crime. The argument is that crimes will not be committed (or will be committed elsewhere) because of the likelihood of being caught in the act by active surveillance, or identified later from video recordings. Armitage et al. [6] list ten ways in which (reported) crime might be reduced by the deployment of CCTV, though the actual effects seem to be limited. Welsh and Farrington [45] in a meta-study of 22 CCTV

studies found an average 4% reduction in crime, and Armitage [5] concludes "Unless publicity is maintained, any initial reductions in crime can fade." A study by Loveday and Gill [22] found that offenders show little concern for CCTV. There is less scepticism when it comes to CCTV's ability to solve crimes as found in a study of police officers opinions by Levesley [21].

Video surveillance to deter shoplifting has long been found in larger department stores but the deployment of surveillance cameras in public spaces has often been a reaction to a particular event: the IRA city of London bomb in 1993; the September 11th attacks in New York; or a single murder in a small town. Pressure to deploy CCTV is increased by the "balloon effect" that it causes — where some crime is simply displaced from areas with CCTV to areas that are perceived as being less likely to detect crime which, in turn, install CCTV to deter and displace crime.

Surveillance is spreading as the hardware becomes more affordable. Prices of video cameras have tumbled in recent years, as technology has improved and production quantities have increased. Similarly, video storage costs have fallen as video recorders have become a commodity item and now digital storage is becoming even less expensive, with higher quality, than analog video. Finally, installation costs have fallen and look set to fall further as wireless networks obviate the need for cables. Today a colour video camera with wireless transmitter and receiver retails for as little as $70. The adoption of 3G camera phones also raises the possibility of an instantly-deployable wireless, multi-viewpoint camera network.

2.1 Public concerns

In general the rise of video surveillance has been tolerated or welcomed by those being watched, because of the perceived benefits in terms of public safety and crime-fighting, but there have always been dissenting voices pointing out the potential abuses of video surveillance for invading individuals' privacy. Recent technological developments and the threat of blanket video surveillance have heightened these concerns and led to growing public concern about the less benign effects of mass surveillance.

Many writers [16, 23, 46] have likened the effects of video surveillance to the "Panopticon" of Jeremy Bentham [7]. This was a design for a prison wherein a guard could see every act of the prisoners, and would make the prisoners aware of the fact, leading them to believe that they were being constantly watched. The idea behind the Panopticon was that, faced with this omniscience, the prisoners would be subdued into good behaviour. "The Panopticon functions as a kind of laboratory of power. Thanks to its mechanisms of observation, it gains in efficiency and in the ability to penetrate into men's behaviour; knowledge follows the advances of power, discovering new objects of knowledge over all the surfaces on which power is exercised." [16] George Orwell, in "1984", satirized the Soviet Union, imagining a future society with powerful surveillance abilities reaching into peoples' homes via the "telescreen". In this way "Big Brother" could know their every act and inspire the self-censorship intended by the Panopticon:

> So long as he remained within the field of vision which the metal plaque commanded, he could be seen as well as heard. There was of course no way of knowing whether you were being watched at any given moment.[30]

The Orwellian dystopia inspires dread in us, but the rise of surveillance is bringing a number of undesirable effects, even without the advent of a totalitarian society in which a "Big Brother" sees and hears all. In technological terms, the apparatus of the Ministry of Truth is commonplace today. As early as the 1970s, Martin and Norman [24] noted that "a surprising amount of what George Orwell imagined now looks plausible, " and Barry Steinhardt, Director of the American Civil Liberties Union (ACLU) Technology and Liberty Program [43] says "Many people still do not grasp that Big Brother surveillance is no longer the stuff of books and movies."

The ACLU has outlined [4] a number of concerns about video surveillance, describing five abuses of CCTV:

- Criminal abuse,
- Institutional abuse,
- Abuse for personal purposes,
- Discriminatory targeting, and
- Voyeurism.

A study of the use of video surveillance in Britain [28, 29] found that "The young, the male and the black were systematically and disproportionately targeted, not because of their involvement in crime or disorder, but for 'no obvious reason' and on the basis of categorical suspicion alone." Norris has also found [33], not surprisingly, that video surveillance is used for voyeurism, and that surveillance was never used to watch over those at risk: "Operators simply do not look out for those they think may be vulnerable to ensure they do not become the victim of mishap or predators, but focus on stereotypical categories of those they think may be likely to offend. Women were also far more likely to be the object of

voyeuristic rather than specifically protectional surveillance...."[29] The ACLU is not alone in concluding there is "a lack of proportion between benefits and risks."

2.2 Total Information Awareness

In a 1995 report, Privacy International [2] listed "Advanced CCTV equipment" among the technologies being exported to developing, particularly non-democratic countries and "used to track the activities of dissidents, human rights activists, journalists, student leaders, minorities, trade union leaders, and political opponents. It is also useful for monitoring larger sectors of the population."

In the U.S.A., retired Admiral John Poindexter recently conceived the "Total Information Awareness" (TIA) project which aims to gather and mine large quantities of data, of all kinds, and use these to detect and track criminals and terrorists. The Orwellian potential for such a project raised an outcry that resulted in the project being renamed the *Terrorist Information Awareness* project, an epithet calculated to stifle objection in post-September 11th America.

The Association for Computer Machinery sent a letter to the Senate Committee on Armed Services Chairman expressing their concerns about the TIA program.

> As computer scientists and engineers we have significant doubts that the computer-based TIA Program will achieve its stated goal of "countering terrorism through prevention". Further, we believe that the vast amount of information and misinformation collected by any system resulting from this program is likely to be misused to the detriment of many innocent American citizens.

They go on to say:

> Privacy is a fundamental American value. Fair Information Practices were developed because policymakers recognized that there are critical issues of privacy when aggregating data that was collected for other purposes. First formulated by a Department of Health, Education and Welfare committee in 1973, the Code of Fair Information Practices is the foundation for the federal Privacy Act of 1974 and the privacy laws of the country. It prohibits secret databases and mandates fairness, accountability, and due process for individuals about whom information is gathered. The need for oversight and control is especially great when aggregation and analysis of personal information is done without the knowledge or consent of the people being monitored.

2.3 Automated surveillance

CCTV deployment is undoubtedly expanding rapidly. McCahill and Norris [25] estimate that there are more than 4 million CCTV cameras in operation. These are often little monitored and of poor quality, installed as a deterrent without much regard for practical use. Automatic processing of surveillance video, however, will bring in a new era of CCTV with constant monitoring, recording and indexing of all video signals. Some CCTV systems have already publicly deployed face recognition software which has the potential for identifying, and thus tracking, people as effectively as cars are recognized today (see below). Currently face recognition technology is limited to operate on small databases or under good conditions with compliant subjects [31].

Many groups around the world [9, 11, 17, 20, 26, 40] are developing software tools to automate and facilitate the task of "watching" and understanding surveillance videos. These systems also have the potential for gathering much richer information about the people being observed, as well as beginning to make judgments about their actions and behaviours, as well as aggregating this data across days, or even lifetimes. It is these systems that magnify the potential for video surveillance, taking it from an expensive, labour-intensive operation with patchy coverage and poor recall, to an efficient, automated system that observes everything in front of any of its cameras, and allows all that data to be reviewed instantly, and mined in new ways: tracking a particular person throughout the day; showing what happens at a particular time of day over a long period; looking for people or vehicles who return to a location, or reappear at related locations.

Algorithms exist for tracking people, understanding their interactions, determining which way they are looking and so on. Compression algorithms have reduced the storage needs, as digital (networked, off-site) storage has tumbled in price. Further algorithms bring the potential to automatically track individuals across multiple cameras, with tireless uninterrupted monitoring, across visible and non-visible wavelengths. Such computer systems may in future be able to process many thousands of video streams—whether from cameras installed for this purpose by a single body, or preinstalled private CCTV systems, access to which is subpoenaed or coerced—resulting in blanket, *omnivident* surveillance networks.

While the technologies to achieve all of this have not yet matured to adequate reliability, the London conges-

tion charging scheme [3] is an efficient, wide-area tracking system that heralds what might be done in the future to track people. The initial congestion charging system uses up to seven cameras at each of 230 locations to read licence plates of cars driving into or within the Congestion Charging Zone of central London. In addition, mobile cameras attached to laptop computers can be set up in other locations. The system has been created to levy a charge on anyone driving in the zone during peak hours, but has resulted in a system with the potential for much more. The system can reliably track vehicles passing in front of the cameras and could be used to know the movements of vehicles in the zone, determine if they were speeding, and the data captured could even be built up into a database of the regular habits of individual motorists. Similar cameras are already in use in London for spotting stolen vehicles.

2.4 Non-surveillance applications

While surveillance has driven the widespread deployment of cameras, low cost cameras and more sophisticated algorithms are enabling many other applications that involve the installation of cameras that will see people. These range from today's traffic cameras and cameras that anticipate drownings in swimming pools [1] to "human aware" buildings that adjust heating, lighting, elevators and telephones according to the locations and activities of people, as well as controlling physical access and assisting with speech recognition by lip-reading [32]. Many future devices and systems will have cameras installed because they are a low-cost sensor that "sees the world as humans see it". In an increasingly networked world, what guarantees do we have that this video is not being recorded or used for purposes besides the original intent?

3 What is video privacy

Faced with the current explosion in video camera deployment, by governments, corporations and individuals, together with the new technologies for exploiting the video, it is important to ask what protections are, or could be put, in place to protect individuals' privacy.

The problem of protecting privacy is ill-posed in the sense that privacy means different things to different people, and attitudes to its protection vary from the belief that this is a right and obligation, to an assumption that anyone demanding privacy must have something to hide [13]. Brin [12] argues that at some level privacy cannot be preserved and suggests how society can deal with that. Danielson [15] views the ethics of video surveillance as "a continuously modifiable practice of social practice and agreement". What is considered acceptable or intrusive in video privacy is a result of cultural attitudes (Danielson contrasts attitudes in the UK and Canada) but also technological capability. A report of the US General Accounting Office [38] quotes the 10th Circuit Court of Appeals decision to uphold the use of surveillance cameras on a public street without a warrant on grounds that "activity a person knowingly exposes to the public is not a subject of Fourth Amendment protection, and thus, is not constitutionally protected from observation." However technology (with capabilities such as high zooms, automatic control, relentless monitoring, night vision and long term analysis) enables surveillance systems to record and analyze much more than we might believe we are "exposing to the public". It has been argued that the "chilling" effect of video surveillance is an infringement of US first amendment rights.

3.1 The transparent society

Faced with an inevitable spread of surveillance technology to include blanket coverage of urban areas by fixed cameras and the possibility of future tiny aerial vehicles that could go anywhere carrying cameras, Brin [12] suggests that we are faced with a choice. The surveillance infrastructure is inevitable, he opines, but our choice is whether to entrust the access of the cameras to authorities, as today and as in Orwell's Oceania, or whether to democratize access to the surveillance mechanisms and use these same tools to "watch the watchers" and so protect the populace against abuses of the tremendous power that the surveillance apparatus affords.

3.2 Video privacy vs. general data privacy

In many legal systems, video privacy falls under the legislation dealing with general data privacy and thence data protection. In the European Union, for instance, this is covered by EU directive 95/46/EC which is enacted by member states in their own legislation and came into force in March 2000. In the United Kingdom, with perhaps the densest video surveillance, the relevant legislation is the 1998 Data Protection Act (DPA) which outlines the principles of data protection, saying that data must be:

- fairly and lawfully processed;
- processed for limited purposes;
- adequate, relevant and not excessive;
- accurate;
- not kept longer than necessary;
- processed in accordance with the data subject's rights;

- secure;
- not transferred to countries without adequate protection.

The act requires all CCTV systems to be registered with the Information Commissioner [42], extending the 1984 Data Protection act that only required registration of CCTV systems that involved "Automatic Processing" of the data. It further gives specific requirements on proper procedure in a CCTV system in order to protect privacy:

> Users of CCTV systems must prevent unauthorized access to CCTV control rooms/areas; all visitors must be authorized and recorded in the visitors log and have signed the confidentiality proforma. Operators/staff must be trained in equipment use and tape management. They should also be fully aware of the Codes of Practice and Procedures for the system. The observation of the data by a third party is to be prevented e.g. no unauthorized staff must see the CCTV monitors.

It has been estimated [25] that 80% of CCTV systems in London's business district are not compliant with the DPA.

The act also guarantees the individual's right of access to information held about them, which extends to access to CCTV recordings of the individual, with protections on the privacy of other individuals who may have been recorded at the same time. [1]

The European Convention on Human Rights guarantees the individual's right to privacy (see `http://www.crimereduction.gov.uk/cctv13.htm`) and further constrains the use of video surveillance, most explicitly constraining its use by public authorities.

The Swiss Federal Data Protection Commissioner has published these guidelines: [14]

> When private individuals use video cameras, for example to protect individuals or prevent material damage, this is subject to the federal law of 19th June 1992 on data protection (DPL; SR 235.1) when the images filmed show identified or identifiable individuals. This applies irrespective of whether the images are stored or not. The processing of the images—such as acquisition, release, immediate or subsequent viewing or archiving—must comply with the general principles of data protection.

3.2.1 Why video is different

A big difference between ordinary data privacy and video data privacy is the amorphous nature of the latter, and the difficulty in processing it automatically to extract useful information. A video clip can convey negligible amounts of information (e.g. there is nobody in the street at 4 a.m.) or may contain very detailed and specific information (about times, a person's appearance, actions). Privacy is hard to define, even for explicit textual information such as name, address and social security number fields in a database, knowledge of which can be used for identity theft, fraud and the mining of copious information about the individual from other databases. It becomes much harder to assess the privacy-invasion that might result from the unstructured but potentially very rich information that could be harvested from surveillance video. A simple video of a person passing in front of a surveillance camera by itself affords little power over the individual, except in a few rare circumstances (proving or invalidating an alibi for instance).

There are already strong restrictions on the use of microphones for surveillance because of the presumption of privacy of conversations, but video has been less restricted because there is an expectation of being observed when entering a public space. The UK DPA exempts from controls data where, "The information contained in the personal data has been made public as a result of steps deliberately taken by the data subject." While the act of walking along the street could be construed as deliberate steps to make ones visual appearance public, we have seen that the DPA does provide privacy safeguards for CCTV.

Hitherto the unmanageabilty of surveillance video has limited its potential for abuse. It takes time to review video to find "interesting" excerpts, and the storage requirements have added to privacy reasons to ensure that recordings are retained for only short periods of time. Long term storage, and detailed analysis have been reserved for situations with strong economic or forensic motivation.

However, the advent of the sophisticated computer al-

[1] "The DPA supports the right of the individual to a copy of any personal data held about them. Therefore data controllers are obliged to provide a copy of the tape if the individual can prove that they are identifiable on the tape , and they provide enough detail to locate the image (e.g. 1 hour before/after the time they believe they were captured by CCTV, their location and what identifiable features to look for). They must submit an appropriate application to the Data Controller and pay a £10 fee. However, the request can be refused if there are additional data/images on the tape relating to a third party. These additional images must be blurred or pixelated out, if shown to a third party. A good example would be a car accident where one party is attempting to claim against another. The data controller is obliged to say no to a civil request to view the tape, as consideration must be given to the other party. A request by the police is a different matter though."

gorithms of section 2.3 to automate the extraction of data from video, mean that video is becoming as easy to mine as a queryable, machine-readable database. The data mined from an omnivident surveillance network will have a potential power that can only be guessed at today. Even in a liberal democracy and with many checks and balances, the potential for abuse is large. One possible threat is increasingly arbitrary justice—laws which are rarely enforced (such as speeding, or drug possession) end up being applied selectively and unfairly. The potential for this expands as the state becomes more able to monitor every individual's every action, though automation can (as with speed cameras) take some of the arbitrariness out of the system.

3.3 Technological video privacy

Little work has been done on the protection of video privacy beyond the creation of legislation, principally in Europe, that describes how a CCTV system can be run and what can be done with the data. In recent years a few attempts have been made to use technology to protect the privacy of people observed by CCTV systems.

A Sony patent [8] describes a system that detects skin tone and replaces it with another colour. This invention has the purpose of hiding surveillance subjects' race to avoid discriminatory surveillance. Matsushita [44] have patented a system for obscuring a "privacy region" being observed by a pan-tilt-zoom camera. Newton *et al.* [27] recently described a system for "de-identifying" faces by transforming faces in shared surveillance video to prevent them from being recognized by a face recognition system.

4 A model for video privacy

From the previous sections we have seen that the following aspects are crucial to privacy in video surveillance systems:

- **What data is present:** The fundamental determinant of video privacy is what information is captured and conveyed by the surveillance.
- **Consent:** Clearly if the subject willingly consents to be observed, privacy is less of an issue, but consent can vary from consciously choosing to walk in front of a surveillance camera; to walking in front of one because by I have no option; walking in front of a hidden camera or being spied upon in secret, in an area where I can reasonably expect privacy (a hospital, my home). When consent is given it is usually on the understanding of certain privacy protections. The Institute for Applied Autonomy has developed a tool on a handheld device "iSee"[34] to suggest routes through cities avoiding surveillance cameras. On the other hand many people install wireless "nanny cams" and unwittingly broadcast to their neighbourhood unencrypted video of the insides of their own homes.
- **Who sees the data:** Is the data restricted to the police? To security professionals? To the management of my employer? What procedures are in place to enforce the policy? How well is the data protected against hackers, burglars or subpoena?
- **How long is the data kept:** Not only does this limit the period over which the data can be used, but it also limits the number of people who can get to see it. A great difference exists between systems that present video feeds for synchronous review by guards and those which store the video or other information derived from it.
- **How raw is the data:** Raw video on a tape is unwieldy and difficult to "use" without significant resources, addition of a time-stamp or other meta-data begins to make the video more accessible and consequently more likely to be intrusive of privacy. Meta-data may be stored even if the original video is discarded, and can be searched with or without the video.
- **What form is the data in:** This applies not only to the physical or electronic medium (Is it on tapes which might be removed? Is it transmitted over a network that might be tapped? Is it encrypted? Is the information accessible to a single person, or does it require the keys of multiple people to be read?)

Consideration of these aspects leads us to a model for protecting privacy of individuals observed by video surveillance systems.

- **What data is present:** At a very basic level, the design of the camera system should be designed to limit the data capture to areas where surveillance is needed and not intrusive. Blinkers, blinds or physical stops on the motion of steerable cameras, as well as lens caps and indicators of when the camera is in operation can both restrict the field of view of cameras and reassure the public that surveillance is bounded. A low resolution camera or deliberately defocussed lens are further ways in which privacy might be defended simply while preserving the systems usefulness.
- **Consent:** Willing consent is hard to achieve in public places. Signs are often used to inform the public (often to intentionally increase the "Panopticon"

effect) but generally consent is not sought by those deploying CCTV.

A future system might be derived that uses face information from the video to permit access to portions of the data representing a person presenting a request. In this way access to the video guaranteed through freedom of information provisions (as in the UK DPA) might be automated, and privacy clashes in the requested video might also be detected.

- **What form is the data in:** Data should be stored digitally and encrypted. In this way stealing the tapes or eavesdropping on transmissions no longer permits access to the video. Indeed, encryption should be carried out at the camera to prevent eavesdropping at any stage.

- **Who sees the data:** In addition to the physical and procedural controls already required by data protection laws, we propose a series of controls on access through a secure console. Video is encrypted and only accessible through a decoding console (see Figure 1) with the approved key. Further, a user key is required to access the data, with a system of access control rules detailing who can view what data under what circumstances, and additional restrictions such as key sharing to require multiple authorizations. Operations such as playback, searching, freeze frame etc may require different levels of authorization.

- **How long is the data kept:** With viewing managed through the privacy console, data lifetime can be managed with keys independent of the lifetime of (perhaps illicit) copies of the encrypted data.

- **How raw is the data:** This is the crucial aspect and one where video privacy can be most effectively enhanced. We propose a system of video processing to mask out privacy-invasive features. The methods for doing this are detailed in the following sections. Raw video that has not been masked in this way can be processed after-the-fact to extract meta-data.

4.1 Absolute vs relative ID

A major distinction among video surveillance systems, that significantly correlates with how likely they are to intrude on privacy, is the level of anonymity they afford. We distinguish three types of system: *Anonymous*, *Relative ID* and *Absolute ID*:

- **Anonymous** A typical CCTV system without computer augmentation is anonymous—it knows nothing about the individuals that are recorded onto the tape or presented on the monitors. While open to abuse by individuals watching the video, it does not facilitate that abuse.

- **Absolute ID** These systems have some method of identifying the individuals observed (such as face recognition or a badge swipe correlated with the video) and associating them with a personal record in a database. Such systems require some kind of enrollment process [10] to register the person in the database and link the personal information (such as name, social security number) with the identifying characteristic (face image or badge number).

- **Relative ID** These systems can recognize people they have seen before, but have no enrollment step. Such systems can be used to collect statistics about people's comings and goings, but do not know any individual information. A relative ID system may use weaker methods of identification (such as clothing colours) to collect short term statistics as people pass from one camera to another, but be unable to recognize people over periods of time longer than a day.

Clearly, anonymity protects the individual's privacy. An absolute ID system might, for instance be made to "Give a report on the movements of Joe Bloggs at the end of each day". A relative ID system with a "strong identifier" might be converted retrospectively into an Absolute ID with a manual enrollment.

5 Privacy preserving video console

In accordance with the factors in our model of video privacy, we have built a prototype system to record and redistribute surveillance video in a way designed to minimize the intrusion into individuals' privacy. Our prototype of the privacy-preserving surveillance video console concentrates on the **What data is present** and **How raw is the data** issues, and uses conventional technologies (encryption, access control lists) to deal with the other issues. The system works by re-rendering the video stream to hide the privacy-intrusive details while preserving the information necessary for the system to be useful. We have also embodied the principles of privacy protection in a "PrivacyCam"—a camera with onboard processing that produces a video stream with the privacy-intrusive information already removed, that can be used for a variety of automated, video-dependent applications as well as surveillance.

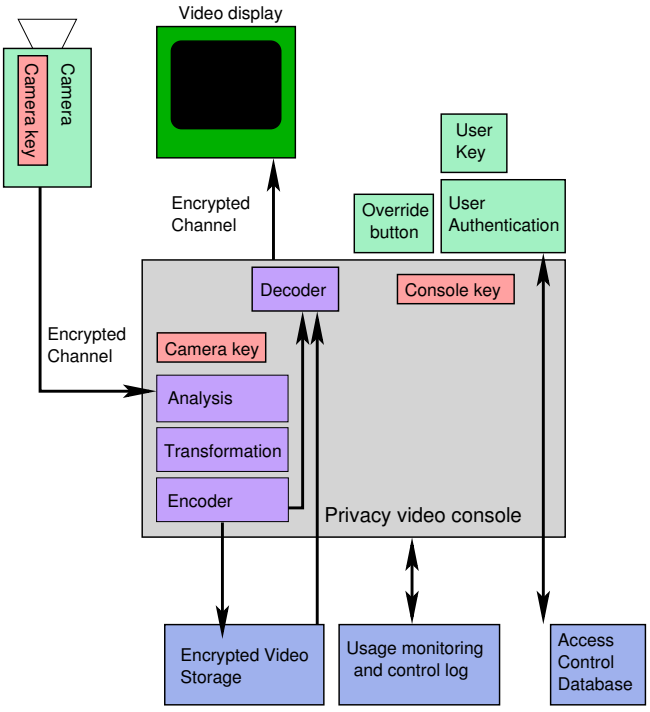


Figure 1: A privacy-preserving console for managing surveillance video.

5.1 System architecture

The basic premise of our privacy proposal is that various information components of the video content can be automatically extracted; these components can be made accessible to different system users based on their authorization levels.

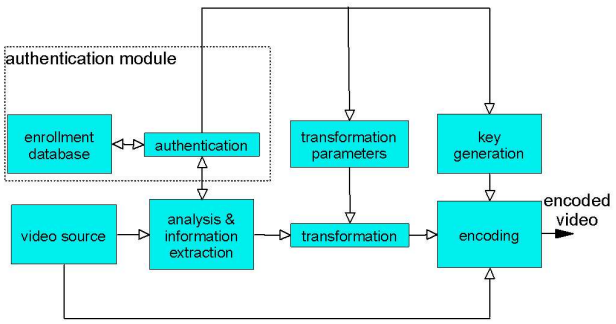
Figure 1 shows a block diagram of the complete system. At the highest level, the architecture consists of a selective video encoding system transmitting an encoded video signal to a selective video decoding system, either live or with intermediate storage. The decoding and encoding systems operate under the control of a user authentication system.

The encoding system (Figure 2a) consists of video analysis, transformation, and encryption subsystems. The video analysis subsystem (Figure 2b) takes a stream of one or more live or recorded videos and analyzes the video at successively more sophisticated levels to extract separate streams of information – for instance about the appearance of the background, and various attributes of different moving objects. The transformation subsystem selectively transforms the information extracted from the video based on the system policy. Finally the transformed, extracted information is encrypted using the encryption subsystem, using different keys for different information streams. The video thus encoded may contain multiple copies of essentially the same information, although each of the copies may be encoded with a different key. The encoded information stream may include an encrypted version of the original video in its unprocessed form. The encoded video is modular in nature and each module (or channel) represents one or more components of extracted or raw video information.

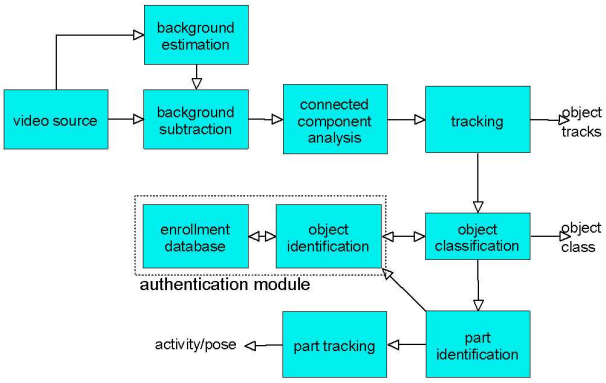
The decoding system (Figure 2c) primarily consists of user interface and video synthesis subsystems. The former establishes the identity of the system operator through the user authentication module (which authenticates the operator through token-, knowledge- or biometrics-based authentication [10]) and enables the operator to apply selective operations (e.g., query/view/freeze frame/export to analogue tape) to the synthesized video or unencrypted video information. The video synthesis module decrypts the encoded video information received from the encoding system (or storage) using the keys and authorization from the user authentication system. The decrypted video information may be used for both the reconstruction of the (transformed) video and for answering operator queries. All the operator-accessed information and operator actions are securely logged. Similarly an audit trail is maintained of all the data processing operations to enable guarantees of data integrity, for legal admissibility.

5.2 Video analysis

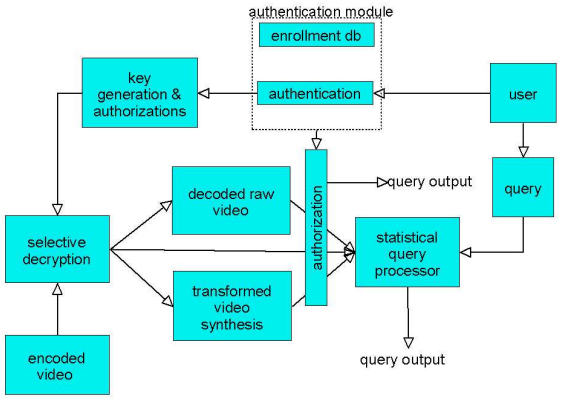
We propose a layered approach to granting access to the different kinds of data extracted by the system. Depending on the user authorization, the user interface may grant access to the original raw video or even video enhanced with additional information, or it may present only reconstructed video with much detail deliberately obscured, or simply present statistical information derived from the video, such as a count of the number of people in a space. The determination of what information can be allowed to which users is very much dependent on the situation and the types of users, but we provide a set of tools and basic algorithms that handle the most common cases. Figure 2d shows one possible layering of access, with law-enforcement officials being able to subpoena the original video. Security guards are able to see only (identity-obscured) re-rendered video, except when they use an override button whose usage is carefully logged (with time and the video at that time). Other registered users may be permitted to access other information, and anonymous users can make simple inquiries about statistics. Devices may be registered as users — for instance an elevator control computer may be given access to the number of people standing in front of the doors when an elevator is summoned.



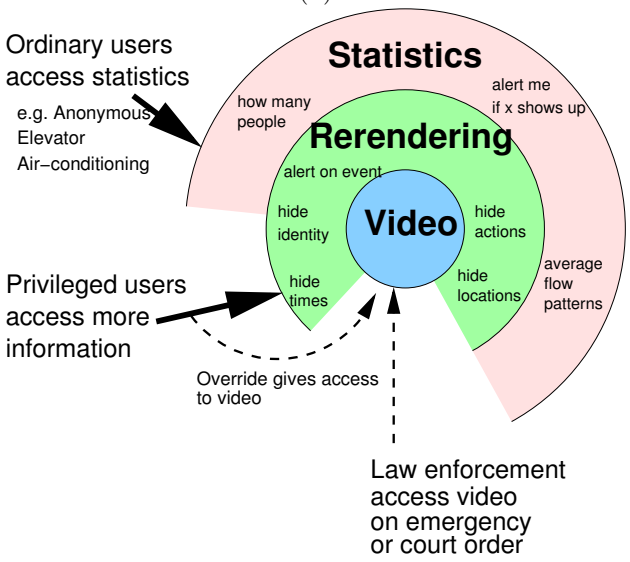
(a)



(b)



(c)



(d)

Figure 2: Privacy system architecture (a) Encoding (b) Analysis (c) decoding (d) A layered approach to the presentation of surveillance video.

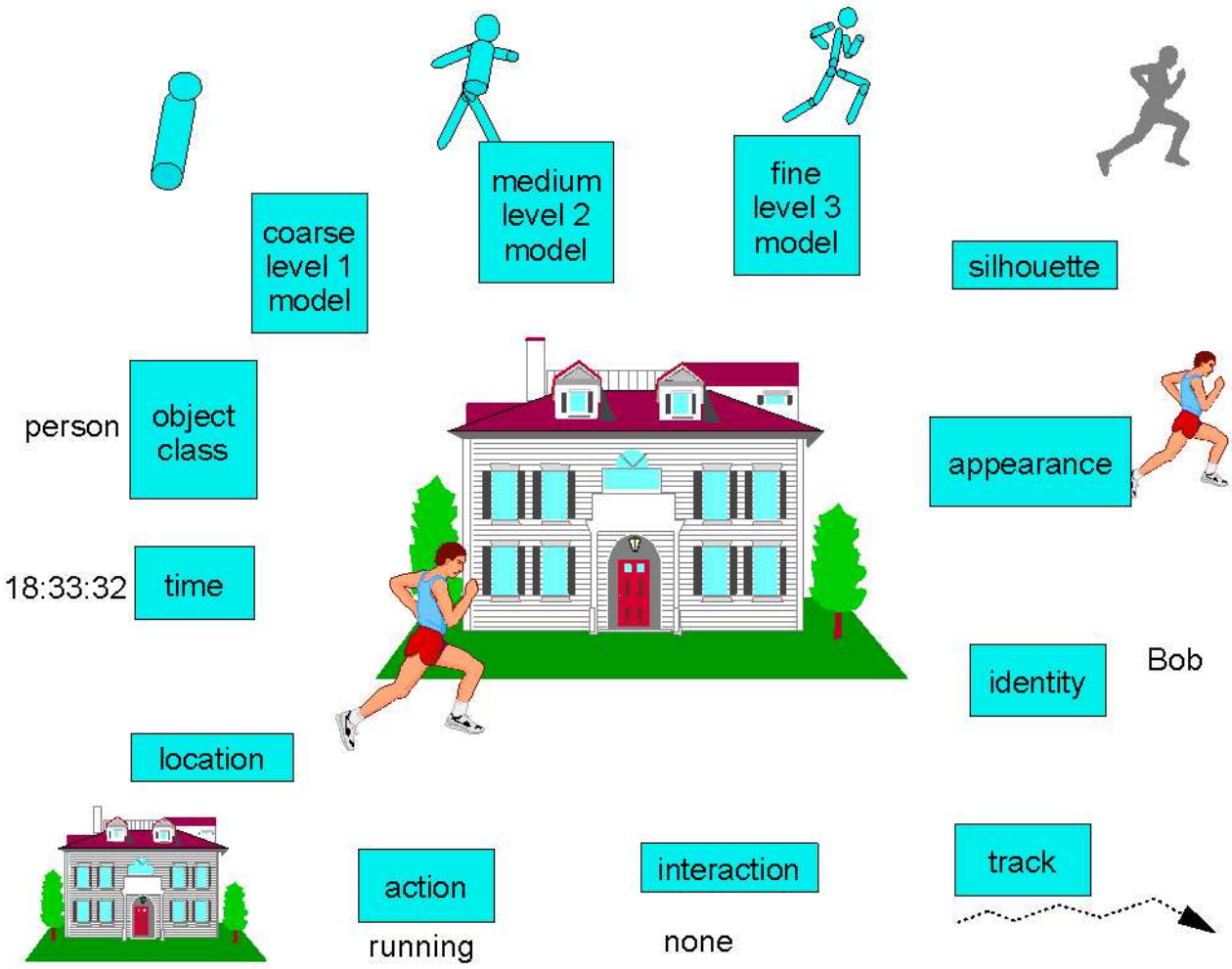


Figure 3: Video re-rendering options. The central image shows a video frame. Surrounding it are different methods of re-rendering the video which can be selected or composed together if the operator's privileges allow.

The algorithms in the analysis subsystem (Figure 2b) use computer vision techniques to "understand" the video: extracting *objects of interest*; distinguishing between "background" and "foreground"; separating people from vehicles; distinguishing people who walk in groups; and even distinguishing between different limbs within a person.

Here we only summarize some of the basic principles of the image analysis. Readers are referred to papers about the PeopleVision system for more details of our implementation of video analysis algorithms [18, 35, 36]. Detection of objects of interest can be accomplished by one of two strategies. In a generic object detection approach, *all* objects of interest are defined in terms of one or more attributes (features) of video image sequence. For example, we could observe that we are mostly interested in moving objects and detect the objects based on their motion, or difference from some archetypal *background model*, deviations from which are assumed to be "interesting" [11, 19, 39, 41]. Differentiating a detected generic object into specific object categories (or a false alarm) is then solved using a more specific model for each object.

In a model-specific approach, each object of interest is modelled and explicitly detected using model-based techniques. For example, in a surveillance application one may be predominantly interested in humans and vehicles. The detection of these objects may proceed from models of vehicles and the human body or face [37]. A hybrid approach which combines model-specific and generic strategies for object detection is also possible.

Once an object is detected in the video, its category (e.g., person), identity (e.g., John Doe, vehicle G104F, or object ID #20351), location (e.g., in Room 22-101), pose (e.g., sitting), etc. can be inferred by a further processing of video and the context. The locus of the object in successive frames determines the object *track*. Changes in the object over time can be used to infer activity; relating the activity of multiple objects defines a group interaction. Thus, the content of the video sequence can be richly represented in terms of the features extracted through video analysis.

The transformation subsystem transforms the video content by first selecting a component information of the video and then obscuring that piece of information, or its complement. For example, a particular system policy may dictate that location information in the video be completely erased. Another system policy may require that all faces in the video be masqueraded so that only gender (say, but not identity, age or expression) information be available from the transformed video. Similarly, the system policies may choose to partially/fully obscure or statistically perturb one or more components of extracted information such as location, pose, activity, track, and so on. Some simple global operations (such as noise, jitter, colour desaturation, blurring, time/space downsampling) may be prevent effective machine processing of the video stream. The transformed information components constitute an encoded video *channel*. Figure 3 and Table 1 enumerate some examples of the selection and obscuration methods.[2]

Transformation	Description
Null	No change.
Annihilation	Removal of information.
Decimation	Reduction in resolution.
Iconization	Replacement with appropriate limited prototypes.
Distortion-I	Replacement with fixed spatial or temporal warping.
Distortion-II	Replacement with dynamic spatial or temporal warping.
Caricature	Replacement with deliberate exaggerated characterization
Masquerading-I	Deliberate fixed overt misrepresentation.
Masquerading-II	Deliberate dynamic covert misrepresentation.
Substitution-I	Deliberate fixed covert misrepresentation.
Substitution-II	Deliberate dynamic covert misrepresentation.
Randomization-I	Random perturbation while preserving ensemble statistical properties.
Randomization-II	Random perturbation without preserving ensemble statistical properties.

Table 1: Example transformations. Different *selected* components of the information may undergo different transformations. For instance, one could completely obliterate the face of a person (*annihilation*) but choose to keep the rest of the person's appearance intact (*null*).

The protection and restriction of the information in our system follows a three pronged approach, using transformation, summarization and encryption to deliver only authorized information to any user. The transformation subsystem implements an obscuration policy for a channel, and users having access to a channel will not be able to infer the information obscured in that channel because that information is irrevocably lost and cannot be recovered from that channel. A user may, however, have access to some information (albeit at different levels of detail) through multiple channels, and the system must be designed such that information from multiple streams (perhaps from different colluding users) cannot be used to reconstruct information not in any of the streams. The statistical query processor delivers information of even less sensitivity. The encryption and decryption processes protect all information channels from tampering and interception. User authentication can be combined with identification of faces in the

[2]Some examples of re-rendered videos can be found at `http://www.research.ibm.com/peoplevision/videoprivacy.html`

video and transformation to make video of an individual available to him/her without obscuration except to protect the privacy of others present at the same time.

5.3 The PrivacyCam

The PrivacyCam is a standalone implementation of some of the concepts that we have described above. In the version that we have built, the camera's output is in the form of a re-rendered NTSC video stream so the PrivacyCam can simply replace a standard CCTV camera, but with privacy-preserving features built in. In this case the camera is designed with on-board processing power, so the video encoding, transformation and encryption take place on the camera before transmission. The on-board processor implements any of the processing algorithms available in the privacy console. In addition to the re-rendered output video stream, encrypted information streams can also be transmitted via other output ports, such as over a wireless network. Such a privacy camera may be limited to only ever produce a single type of video stream as output, or could be integrated with a privacy console to allow authenticated requests to show the original video or some other information stream.

6 Guaranteeing video privacy

Video information processing systems, including the system outlined here, are error prone. Perfect performance can not be guaranteed, even under fairly benign operating conditions, and the system makes two types of errors when separating video into streams: missed detection (of an event or object) and false alarm. We can trade these errors off against one another, choosing an *operating point* with high sensitivity that has few missed detections, but many false alarms, or one with low sensitivity that has few false alarms, but fails to detect events when they occur.

The problems of imperfect video processing capability can be minimized by selecting the appropriate system operating point. The *costs* of missed detection and false alarm are significantly different, and differ in privacy protection from those for a surveillance system. Given the sensitive nature of the information, it is likely that a single missed detection may reveal personal information over extended periods of time. For example, failing to detect, and thus obscure, a face in a single frame of video could allow identity information to be displayed and thus compromise the anonymity of days of aggregated track information associated with the supposedly anonymous individual. On the other hand, an occasional false alarm (e.g. obscuring something that isn't a face) may have a limited impact on the effectiveness of the installation. The operating point can be part of the access-control structure—higher authority allows the reduction of the false alarm rate at a higher risk of compromising privacy. Additional measures such as limiting access to freeze-frame or data export functions can also overcome the risks associated with occasional failures in the system.

Even with perfect detection, anonymity cannot be guaranteed. Contextual information may be enough to uniquely identify a person even when all identifying characteristics are obscured in the video. Obscuring biometrics (face, gait) and weak identifiers (height, pace length, clothing colour) will nevertheless reduce the potential for privacy intrusion. In general, these privacy-protection algorithms, even when operating imperfectly, will serve the purpose of making it harder, if not impossible, to run automatic algorithms to extract privacy-intrusive information, and making abuses by human operators more difficult or costly.

6.1 Increasing public acceptance

Naturally, the techniques described in this paper can be considered as an optional layer on a CCTV system, and one that will cost more and risk impinging on the effectiveness of the surveillance offered. We must then ask the question of why anybody would accept this extra burden. The main reason is likely to be through legislation. In the future, it may be required by law that CCTV systems impose privacy protection of the form that we describe. Indeed it may even be argued that existing legislation would require the deployment of these techniques as they become commercially available.

Without legislation, it may still be that companies and institutions deploying CCTV choose, or are pressured (by the public, shareholders or customers), to "do the right thing" and include privacy-protecting technology in their surveillance systems. Liability for infringement of privacy may encourage such a movement.

We must also ask, however, what guarantee a citizen has that a claimed privacy protection is actually in force. McCahill and Norris [25] estimate that nearly 80% of CCTV systems in London's business space do not comply with current data protection legislation. Legislating public access to surveillance systems as proposed by Brin [12] is one solution, but that still begs the question—is there some data that has not been opened to the public? A potential solution is certification and registration of systems, perhaps along the lines of the system that has evolved for internet privacy (e.g. `www.TRUSTe.org`). Vendors of video systems might invite certification of their privacy-protection system by some independent body. For purpose-built devices with a dedicated camera sensor (like PrivacyCam) this would

suffice. Individual surveillance installations could also be certified for compliance with installation and operating procedures, with a certification of the privacy protection offered by the surveillance site prominently displayed on the equipment and CCTV advisory notices. Such notices might include a site (or even camera) identification number and the URL of the surveillance privacy registrar where the site can be looked up to confirm the certification of the surveillance system. Consumer complaints would invoke investigations by the registrar, and conscientious companies could invite voluntary inspections.

7 Conclusions

Video surveillance and person-aware video systems are here to stay, and will grow ever more powerful. Thus far, controls on the intrusions of privacy that these technologies bring have been very limited and primarily legislative. We have presented a model for future systems that take a *technological* approach to defending video privacy, and have described two systems we have implemented that use computer vision techniques to re-render video information in a useful but privacy-preserving manner. We have also begun to address issues of performance and public acceptance, and hope to encourage more work in this little-researched field.

References

[1] Poseideon. `http://www.poseidon-tech.com/`.

[2] Big brother incorporated: A report on the international trade in surveillance technology and its links to the arms industry. Technical report, Privacy International, London, November 1995. `http://www.privacyinternational.org/reports/big_bro/intro.html`.

[3] Congestion charging: Enforcement technology. *BBC LDN*, 2003. `http://www.bbc.co.uk/ london/ congestion/ technology.shtml`.

[4] ACLU. What's wrong with public video surveillance? 2002. `http://archive.aclu.org/issues/ privacy/ CCTV_Feature.html`.

[5] Rachel Armitage. To CCTV or not to CCTV: A review of current research into the effectiveness of CCTV systems in reducing crime. Technical report, NACRO, London, May 2002. `http://www.nacro.org.uk/data/ briefings/ nacro-2002062800-csps.pdf`.

[6] Rachel Armitage, Graham Smyth, and Ken Pease. Burnley CCTV evaluation. In Kate Painter and Nick Tilley, editors, *Surveillance of Public Space: CCTV, Street Lighting and Crime Prevention*, Crime Prevention Studies. Criminal Justice Press, 1999.

[7] Jeremy Bentham. *Panopticon Letters*. London, 1787. `http://cartome.org/ panopticon2.htm`.

[8] A.M. Berger. Privacy mode for acquisition cameras and camcorders. U.S. Patent 6,067,399, Sony corporation, 23 May 2000.

[9] J. Black and T. Ellis. Multi camera image tracking. In *International Workshop on Performance Evaluation of Tracking and Surveillance*, 2001.

[10] R.M. Bolle, J.H. Connell, S. Pankanti, N.K. Ratha, and A.W. Senior. *Guide to Biometrics: Selection and Use*. Springer-Verlag, New York, 2003.

[11] Terry Boult, Ross J. Micheals, Xiang Gao, and Michael Eckmann. Into the woods: Visual surveillance of noncooperative and camouflaged targets in compex outdoor settings. *Proceedings of the IEEE*, 89(10):1382–1402, October 2001.

[12] David Brin. *The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom*. Perseus Publishing, 1999.

[13] Michael Caloyannides. Society cannot function without privacy. *IEEE Security and Privacy magazine*, May/June 2003.

[14] Swiss Federal Data Protection Commissioner. *Leaflet on video surveillance by private individuals*. 3003 Bern, January 2003. `http://www.edsb.ch/e/doku/ merkblaetter/ video.htm`.

[15] P. Danielson. Video surveillance for the rest of us: Proliferation, privacy, and ethics education. In *International Symposium on Technology and Society*, pages 162–167, 6–8 June 2002.

[16] Michel Foucault. *Discipline and Punish: The Birth of the Prison*, chapter III Discipline. 3 Panopticism. Vintage books, 1994. Surveiller et punir. 1975. Tr. Alan Sheridan.

[17] A. Hampapur, L. Brown, J. Connell, A. Ekin, M. Lu, H. Merkl, S. Pankanti, A. Senior, and Y.L. Tian. Multi-scale tracking for smart video surveillance. *IEEE Transactions on Signal Processing*, 2004. to appear.

[18] A. Hampapur, S. Pankanti, A.W. Senior, Y.-L. Tian, L. Brown, and R. Bolle. Face cataloger: Multi-scale imaging for relating identity to location. In *IEEE conference on Advanced Video and Signal Based Surveillance*, pages 13–20, Miami, July 2003. IEEE Computer Society.

[19] T. Horprasert, D. Harwood, and L. S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *ICCV'99 Frame-Rate Workshop*, 1999.

[20] S. Khan and M. Shah. Tracking people in presence of occlusion. In *Asian Conference on Computer Vision*, 2000.

[21] Levesley. Police use of CCTV. Technical report, Home Office, 2003. Cited in CCTV Image www.cctvusergroup.com October 2003 p. 16.

[22] Karryn Loveday and Martin Gill. *CCTV*, chapter What do Offenders Think about CCTV? Perpetuity Press, 2003. Cited in CCTV Image www.cctvusergroup.com October 2003 p. 17.

[23] David Lyon. *The Electronic Eye: The Rise of Surveillance Society*. Minneapolis: University of Minnesota Press, 1994. http://www.rochester.edu/College/FS/Publications/ Lyon.html.

[24] James Martin and Adrian Norman. *The Computerized Society*. Harmondsworth, Penguin/New York, Random House, 1973.

[25] Mike McCahill and Clive Norris. *CCTV*. Perpetuity Press, 2003.

[26] S. McKenna, J.S. Jabri, Z. Duran, and H. Wechsler. Tracking interacting people. In *International Conference on Face and Gesture Recognition*, pages 348–53, March 2000.

[27] Elaine Newton, Latanya Sweeney, and Bradley Malin. Preserving privacy by de-identifying facial images. Technical Report CMU-CS-03-119, Carnegie Mellon University, School of Computer Science, Pittsburgh, 2003. http://privacy.cs.cmu.edu/people/ sweeney/video.html.

[28] Clive Norris. Surveillance, order and social control. Technical report, Department of Social Policy, University of Hull, Hull, UK, 1997. End of Award Report to the Economic and Social Research Council in respect of grant L210252023, http://archive.aclu.org/issues/privacy/ CCTV_Norris.pdf.

[29] Clive Norris and Gary Armstrong. *The Maximum Surveillance Society*. Berg, Oxford, 1999.

[30] George Orwell. *1984*. 1948. http://www.online-literature.com/ orwell/ 1984/.

[31] P.J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and M. Bone. FRVT 2002: Facial recognition vendor test. Technical Report NISTIR 6965, Defence Advance Research Project Agency, DoD Counterdrug Technology Development Office, National Institute of Justice, Arlington, VA, March 2003.

[32] G. Potamianos, C. Neti, G. Gravier, A. Garg, and A.W. Senior. Recent advances in the automatic recognition of audiovisual speech. *Proceedings of the IEEE*, 2003.

[33] Jeffrey Rosen. A watchful state. *New York Times*, page 38, October 2001. http://www.nytimes.com/2001/10/07/ magazine/ 07SURVEILLANCE.html.

[34] Erich W. Schienke and IAA. On the outside looking out: an interview with the institute for applied autonomy (IAA). *Surveillance & Society*, 1(1), September 2002. http://www.surveillance-and-society.org/journalv1i1.htm http://www.appliedautonomy.com/isee/.

[35] A Senior. Tracking with probabilistic appearance models. In *Third International workshop on Performance Evaluation of Tracking and Surveillance systems*, 2002.

[36] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance models for occlusion handling. In *International Workshop on Performance Evaluation of Tracking and Surveillance*, 2001.

[37] Andrew W. Senior. Recognizing faces in broadcast video. In *IEEE International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pages 105–110, September 1999.

[38] Richard Stana. Video surveillance. Technical Report GAO–03–748, United States General Accounting Office, June 2003.

[39] C. Stauffer. Automatic hierarchical classification using time-based co-occurrences. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, June 23-25*, pages 333–339, 1999.

[40] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):747–757, August 2000.

[41] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. IEEE International Conference on Computer Vision*, volume 1, pages 255–261, 1999.

[42] UK Government. Implications of the data protection act, 2000. http://www.crimereduction.gov.uk/cctv9.htm.

[43] American Civil Liberties Union. Is the U.S. turning into a surveillance society? January 2003. http://www.aclu.org/Privacy/Privacylist.cfm?c=39.

[44] Jyoji Wada, Koji Kaiyama, Ken Ikoma, and Haruo Kogane. Monitor camera system and method of displaying picture from monitor camera thereof. European Patent EP 1 081 955 A2, Matsushita Electric Industrial Co. Ltd., April 2001.

[45] Brandon C. Welsh and David P. Farrington. Crime prevention effect of closed circuit television: A systematic review. Research Study 252, UK Home Office, London, August 2002. http://www.homeoffice.gov.uk/rds/pdfs2/ hors252.pdf.

[46] Reginald Whitaker. *The End of Privacy: How Total Surveillance Is Becoming a Reality*. The New Press, Dec. 1998.