A Progress Review of Intelligent CCTV Surveillance Systems

Anthony C Davies[1)] and Sergio A Velastin[2)]

[1] Visiting Professor, School of Computing and Information Systems, Kingston University, Penrhyn Road, Kingston, Surrey, KT1 2EE, England (and Emeritus Professor, King's College London), e-mail: tonydavies@ieee.org
[2] Reader, School of Computing and Information Systems, Kingston University. e-mail: sergio.velastin@kingston.ac.uk

Keywords: Closed-circuit television, surveillance, image processing, security, vehicle tracking, crowd-monitoring.

1. INTRODUCTION

The installation of closed circuit television (CCTV) cameras in urban environments is now commonplace and well-known. Public attitudes to these systems are in two opposing categories:

(a) concerns over invasion of privacy and fears of authoritarian control of the population

(b) welcoming the increased safety in public spaces and reductions in antisocial behaviour.

Fears in category (a) are not surprising, since within living memory, and in some cases until recently, there have been countries in Europe governed by regimes with a strong commitment to the oppressive monitoring and control of their citizens, and the continuous tracking of actual or supposed 'dissidents'.

Support for category (b) arises from public concerns over both real and imagined risks of urban crime and terrorism. Controlling anti-social behaviour and protection against terrorist threats is generally perceived to have a high priority, making intrusive monitoring relatively acceptable and encouraging the installation of advanced surveillance systems. Recent uses of the phrase 'homeland security', and plans for technologically-advanced personal ID cards has made this a topical issue.

Forms of visual monitoring which were once the exclusive domain of well-funded secretive government security agencies are becoming readily available at an affordable cost to the public, and very sophisticated systems are being developed and installed for general police uses.

2. HISTORICAL STAGES OF DEVELOPMENT

CCTV based surveillance has developed from simple systems comprising a camera connected directly to a viewing screen with an observer in a control room, watching for incidents of crime or vandalism or searching for targeted individuals, to complex multi-camera systems with many computers. The computers carry out image processing, object recognition and scene analysis, prior to presenting data to observers. Sophisticated recording and playback techniques can be provided, with searching capabilities, and suitable for audit and presenting observed results as evidence in legal proceedings.

Early steps in this development were to add manual camera control (pan, tilt and zoom) in order to track events or objects of particular types. As the number of cameras in each system increased, they first exceeded the number of monitoring screens (so requiring sequential switching) and then exceeded the capability of the observing teams to watch events effectively. Their attention span is inevitably limited [1]. Adding computational intelligence to alert the observers to the infrequent image sequences which contained events of possible importance was thus a natural development as computing resources became both cheaper and more powerful.

The need for video recording arose to relieve the observers from overload, to provide an audit capability and for detailed studies of images leading up to an incident Colour, including the use of infrared imaging, has in many cases replaced monochrome, and digital encoding with compression, for storage and/or transmission, has become economical.

The availability of many powerful computers within the overall surveillance system has enabled the required increased automation, with the use of 'computer intelligence' to detect and analyse significant events and alert human observers when appropriate. Detection thresholds usually need to be biased in favour of false positives, since these can usually be quickly recognised and disregarded by a human observer, whereas missing real incidents could be a serious deficiency.

Decreasing cost and increasing processing capability has made distributed multi-computer systems feasible and affordable, and able to include a significant amount of pre-processing of images 'per camera' (as opposed to sending the raw video data to a centralised processor). Image compression methods have advanced considerably, so that the data rates and storage required can be substantially reduced. Adoption of very efficient but lossy image compression methods is not always acceptable because of fears that the images may then be challenged and not be accepted as evidence in legal proceedings.

When video recordings from CCTV systems are required for legal purposes, it is necessary to ensure the feasibility of an independent and authenticated audit, without which such recordings cannot be used as evidence in court proceedings. Recorded image-sequences associated with criminal events can often be

subjected to advanced off-line image enhancement methods, making it easier to identify the particular individuals involved in the crime. Searching large amounts of recorded video data may be required in order to try to find other scenes containing the same individuals.

CCTV surveillance systems are required to have high availability and run continuously, and yet be extendible. For example, adding cameras and computing resources is often desired, but may need to be done without taking the existing system out of operation.

A British newspaper reported in 2004 that over four million CCTV cameras were deployed in the UK [2]

3. CROWD AND PERSON MONITORING

Early uses of computers included the estimation of global properties of urban crowds such as density and flow [3,4]. Applications were to learn about pedestrian behaviour in densely-populated urban spaces such as city rail-stations during 'rush-hours', shopping malls and airport departure lounges. Additionally this data was expected to be useful to assist architects in the planning of urban environments. The techniques did not attempt to identify individual pedestrians in a crowd – rather, the crowd was monitored as a generalised entity, and 'average' properties sought.

Ideal gas theory provides a basis for predicting the behaviour of gases although it takes no account of the trajectories of individual molecules. In an analogous way, some aspects of crowd behaviour can be estimated without considering the individuals making up the crowd. Some analogies with fluid behaviour and with the behaviour of charged particles in an electric field may be observed.

3.1 Crowd Density

If the area of an image occupied by pedestrians can be identified, then the ratio of 'crowd area' to 'background area' provides a rough estimate of crowd density. In low density crowds, the 'edges' of all pedestrians may be extracted from the image, and the total number of edge pixels has been found to be a measure to estimate the number of people in the crowd (Fig.1).
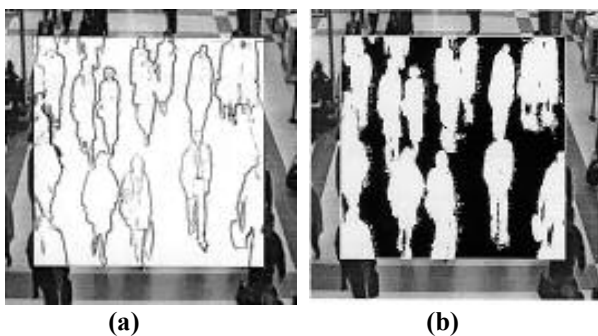

| (a) | (b) |

Fig. 1 Pedestrian extraction from crowd-scene (from [5])
(a) 'edges' of pedestrians (b) area of pedestrians

Compensation is needed for the different apparent size of objects at various distances. Calibration for camera angle is also essential. Variation of lighting levels and directions presents many problems to be overcome. In many monitoring situations in public spaces, the cameras have to be mounted in places which give very oblique images, making image interpretation more difficult.

By monitoring density, it is possible to set thresholds above which safety might be endangered (Fig. 2). For example, entry to city-centre rail stations may have to be periodically closed during rush-hours in order to keep crowd density on platforms at a safe level.

**Fig. 2   Example of automatic identification of a congested area, denoted by array of white dots**

3.2 Crowd Flow

Crowd flow investigations can provide real time data of security value. Optical flow (the change in image brightness from one image to the next, expressed as a vector field) can provide basic motion data. This uses a gradient method to follow changes in brightness in the image [6]

**Fig. 3(a)   Extracted motion-vector field –
arrows show direction of motion**

Block matching is an alternative effective technique. Typically, a rectangular block of pixels around a selected pixel in the first image is compared with a larger search block around the same pixel in the second image.
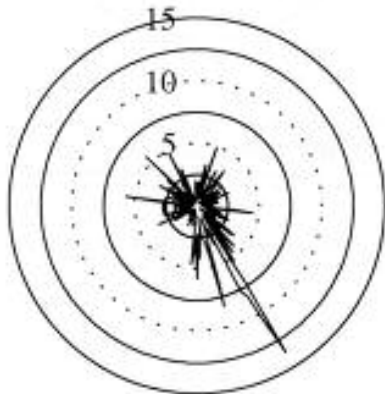
Fig 3(b) Polar velocity plot derived from a crowd flow estimate [5]

All pixel blocks in the search area are compared with the one in the first image, and the one with greatest similarity is used as an indicator of the estimated movement direction and speed (Fig 3(a)). An overview of the movement may be obtained by a polar velocity plot (Fig.3(b)), which also enables unusual flows to be detected automatically (for example, a group of pedestrians moving very quickly in a non-standard direction).

3.3 Detecting Unusual Behaviour

Detection of individuals who remain in one place for long periods while surrounding crowds are moving, or of individuals moving in a different direction from the majority, may be a potential indication of planned or actual criminal behaviour. Automatic detection of such events in the image sequences from a CCTV system may be used to alert human observers, who can then decide if the events are significant. Tracking of specific individuals is also possible. Distinguishing approaching and receding pedestrians (for example by automatic marking with different coloured tags) may be operationally useful.

4. VEHICLE MONITORING

Speed monitoring is often carried out by simple cameras triggered by a vehicle exceeding the speed limit. The photograph captures the vehicle registration number. Instead, this may now be automated within a CCTV surveillance system. This requires not only identifying the moving vehicle within the scene, but also locating the position of the registration plate, and then automatically reading the number. To achieve this reliably under all conditions of lighting and weather is a challenging image-processing task, but operationally successful systems have been developed.

Reliable methods of automatically reading vehicle registration numbers provide alternative possibilities for excess-speed detection. For example, detecting a vehicle at position x_1 at time t_1 and then detecting the same vehicle at position x_2 at time t_2 immediately provides a lower bound for the velocity of travel between these two points of $distance(x_1,x_2)/(t_2-t_1)$. If that is greater than the speed limit, then it can be concluded that the vehicle has exceeded the limit. Thus with a complex network of cameras capable of identifying particular vehicles (e.g. by registration numbers) it is possible to detect extreme speed violations without measuring speed, even when the vehicles are moving within the junctions of a complex highway network. Installing such systems has implications for civil liberties, since the same data provides information on the movements of particular vehicles, and therefore has the potential to be used (misused) to track individuals.

There is now considerable government interest in the enhancement of such capabilities, for such applications as 'congestion charging' (automatically billing the owners of vehicles which are observed travelling within city centres during chargeable periods of time) and 'automated road pricing' (toll-collection with no toll-booths).

5. PUBLIC ACCEPTANCE AND INVOLVEMENT

In most countries, cameras in public areas are not now considered an invasion of privacy. They are a common sight at highway junctions, pedestrian crossings and traffic-signal locations (for traffic monitoring) and in city centres, shopping malls, airports and rail stations (for person monitoring). The trends are towards smaller cameras, which may not be readily visible (sometimes augmented by highly-visible dummy cameras to mislead criminals).

There is now public awareness and acceptance of 'webcams' observing places of interest and supplying images to the internet on a continuous basis. It is becoming affordable to install internet-connected cameras in and around the home, so that remote viewing is possible while away on vacation, etc. to provide reassurance that the house has not burnt down, been flooded, or broken into by thieves. Of course making this data available over the internet does open a risk of interception.

6. TRACKING MOVING OBJECTS: PEOPLE AND VEHICLES [7]

A key technique used to identify moving objects in video data streams is to subtract the current frame from an estimate of the background scene, based on the idea that anything 'new' in the current frame must be mobile object(s). Since the background alters as a result of lighting changes and camera movement, in practice the background estimation needs to be adaptive. Otherwise, effects of sunshine and clouds which would hardly be noticed by a human observer can produce major errors in identifying the moving parts [8].

The moving parts have to be segmented into the distinct moving objects, typically referred to as 'blobs'.

A standard method of tracking objects in a video sequence involves identifying each blob of interest, marking it with an easily visible bounding rectangle,

and in many cases, adding an identifier (Fig. 4). In the case of vehicles, an identifier may be extracted automatically by locating and automatically reading the registration plate. In the case of persons, of course this is not possible.

Fig.4 Pedestrians marked by bounding rectangles (approaches marked by ■ and departures by ⊢)

The object is then 'followed' from image to image. Difficulties naturally arise from occlusion: if the object disappears behind an obstacle, it is possible to use its velocity to estimate the place and time of its emergence assuming no change in its velocity (more sophisticated systems might use acceleration data too). Commonly, a Kalman filter is used to provide better estimates of future trajectories of objects which move into an occluded zone. [9,10]. However, it is easy to imagine many situations where this is unreliable – for example two people being tracked may disappear behind an obstacle, and while there may meet, hold a conversation,, then split up and depart in opposite directions. To automatically determine which one is which from the image sequences following their re-emergence is obviously not at all easy [11].

Two situations need to be handled: *dynamic occlusion*, which is the occlusion of one moving object by another, and *static occlusion* which is the result of moving objects interacting with static objects (pillars, access gates, etc. or the boundary of the image). In the former case, appearance templates of the moving object(s) can be used to complete partially-obscured observations [12]. In the latter case, observations over a long time may be used to create three-dimensional (depth) information about the image, and this may then be used to make interpretations and predictions about the occlusions [13]

Since surveillance is typically achieved by multiple cameras, the 'hand-over' of identified objects from the field of view of one camera is needed. Sometimes the fields overlap, in other cases, there may be a part of the scene not covered by any camera.

Mobile surveillance cameras (for example, mounted on vehicles) present some additional problems because the observed background moves from frame to frame, and a constant background has to be estimated from this sequence of differing views. Only then is it possible to identify those objects which are really moving with respect to this background.

7. SURVEILLANCE FOR PUBLIC SAFETY

Because of the large number of cameras often deployed in public areas, the automatic detection of events of importance for safety and security has become important – the events are required to trigger an alarm, to alert humans able to make decisions about the need for action [14,15]. For example, one might wish to identify unattended baggage in a rail station. Of course, the sudden appearance of a suitably-shaped stationary object may be detected by conventional image processing methods. Alternatively, if a person being tracked 'splits' into a moving person-like object and a smaller stationary object (e.g. depositing an item and walking away from it), this could be used to trigger an alarm.

It is relatively easy to automatically detect individuals in forbidden areas (Fig. 5) or individuals who loiter for excessive times in one place (which can be a sign of criminal intent). Falling pedestrians may be identified as person-like objects which take up a horizontal position with little or no motion. This may be of particular importance needing rapid response if observed on rail tracks [16].

Fig. 5 Pedestrian automatically identified as too close to platform edge (marked by rectangle and bar)

8. SOME SPECIFIC ACTIVITIES OF THE RESEARCH TEAMS AT KINGSTON UNIVERSITY

The Digital Imaging Research Centre (DIRC) at Kingston University has a number of teams working on Video Surveillance projects. Some of this research was started at King's College London and some at City University London, and these teams moved to Kingston, taking the activities with them. Example projects include:

• **'PRISMATICA':** Participation in a European Commission project [17, 18] to make public transport systems more attractive to passengers, safer for passengers and staff and operationally cost-effective. An innovative aspect was the integration of operational, legal, social and technical aspects of CCTV surveillance systems. Partners included RATP-Paris, ATM-Milan,

STIB-Brussels, PPT-Prague, ML-Lisbon, Kings College London, University College London, INRETS-France, CEA-France, TIS-Portugal, SODIT-France, FIT-Italy, ILA-Germany, and Thales-France

PRISMATICA aims included development of concepts for pro-active surveillance systems to provide decision-support tools for human operators in complex and large environments. The tools should automatically guide the attention and actions of the managers of a transport network, while keeping the technology itself as transparent as possible.

This project incorporated a two-stage assessment. First, the architecture was tested in the Paris Métro (Gare de Lyon) and successfully demonstrated communications mechanisms and protocols. Next there was a major deployment of the system in Liverpool Street Underground station, London. This station takes commuters to/from one of the biggest financial centres in Europe and connects with main railways and buses. There are more than seventy cameras in this station covering approximately 80% of its total area.

• **Monitoring Public Spaces:** developing intelligent surveillance tools for integration within existing urban CCTV infrastructures to improve incident detection and assist control room operations. Incidents of interest in public transport systems include overcrowding, loitering, busking, begging, jumping over access barriers, drug-dealing. Fears of terrorism lead to continuous monitoring for abandoned luggage or suspect packages. For all these situations, automation and the tracking of objects provide an opportunity to deploy staff only where they are really needed. This enables security personnel to be freed from having to acquire and manually track targets [19].

• **Plug and Play Surveillance:** devising designs which enable newly installed cameras and associated computational intelligence to easily integrate into an existing camera network.

• **Learning Camera Topology:** An intelligent surveillance system must capture and track objects to establish a history of their behaviour, classify the objects (as people or vehicles, etc. of particular types), and establish their trajectories in a 3D space. In many cases, a number of cameras are used with partially overlapping areas of view, and there may also be areas which are not covered by any camera. Synthesising all the images into a single real-time description of the scene is a complex data-fusion task. Solving this problem also means that camera configuration may be changed without the need for a full re-calibration.

• **Learning Semantic Scene Models:** The aim is to label regions in a scene according to activity (e.g. entry zones, exit zones, stopping zones, junctions, etc.) from video data streams from the scene. The activity is often time-dependent – for example commuter-flows are usually in an opposite direction for travel in morning and evening, because of travel to and from the workplace respectively. The purpose is to assist subsequent interpretation of moving-object behaviour in the scene [11]

• **Automated Extraction of Evidence from CCTV footage:** The aim is to reduce the time and effort of reviewing lengthy CCTV video sequences to locate specific sequences showing incidents of interest by automating the recovery and information management of video evidence.

In many investigations of major crimes in urban environments, CCTV evidence plays an increasingly crucial role in establishing the identity of vehicles and individuals. Gathering this data is extremely time consuming, and involves manual annotation of CCTV archives by police and other experts. A joint project of Kingston University, University of Surrey and Sira,Ltd. is exploring the linkage of the meta-data structure of the video interpretation process with the linguistic structure of police descriptions of evidence. An aim is to validate the effectiveness of this data-fusion process by an automatic generator of galleries of vehicle registration plate images and person images (Fig. 6). Viewing of such galleries by local police and the general public is known to be an effective method of acquiring knowledge.

Fig. 6 Examples of vehicle registration number and person galleries associated with car-park usage

9. FUTURE DEVELOPMENTS

The progress in integrated circuit technology, the increased capabilities of digital cameras and the new communications methods (wireless LANs, mobile phones) all contribute to the continued deployment of more complex and advanced CCTV surveillance systems, which will become increasingly unobtrusive as cameras decrease in size, and likely to be integrated with other sensors (audio, thermal, etc.) and data-bases.

The software will inevitably increase in capability, so that tracking and recognition of people and vehicles will become much more effective. Automatic recognition of behaviour patterns will improve, so that it will become easier to detect and predict both legitimate and illegal activity. Automatic analysis of gestures and the ability of computers to support lip-reading at a distance

[20] is already indicating a direction in which surveillance might develop.

There is no assurance that these systems will always be used responsibly and only by those with the public interest and safety in mind. Misuse by official agencies and adoption by criminal elements in Society may happen if there are insufficient safeguards.

Just as the invention of the word processor did not result in the paperless office, the development of improved CCTV surveillance systems is not likely to lead to the crime-free city centre.

10. ACKNOWLEDGEMENTS

References:

[1] E. Wallace et al, "Good Practice for the Management and Operation of Town Centre CCTV", *European Conf. on Security and Detection*, 28-30 Apr. 1997, pp36-41 (IEE Conf. Pub. 437)

[2] M. Frith, "Big Brother Britain, 2004" *The Independent*, 12 Jan 2004

[3] A.C. Davies, J.H. Yin, S.A. Velastin, "Crowd monitoring using image processing", *IEE Electronics and Communication Engineering J.*, 7, Feb 1995, pp 37-47

[4] S. A. Velastin, J. H. Yin, A. C. Davies, M. A. Vicencio-Silva, R. E. Allsop A. Penn, "Automated measurement of crowd density and motion using image processing," *7th Int. Conf. on Road Traffic Monitoring and Control, London,* UK, 1994, pp. 127-132.

[5] J. H. Yin: "Automation of crowd data-acquisition and monitoring in confined areas using image processing," Ph.D. Thesis, King's College London, University of London, September 1996.

[6] B.K.P. Horn, B.G. Schunck, "Determining Optical Flow", *Artificial Intelligence*, 17, 1981, pp. 185-203

[7] G. Foresti, C. Micheloni, L. Snidaro, P. Ramagnino,T Ellis, "Active Video-based Surveillance System", *IEEE Signal Processing Magazine*, March 2005, pp25-37

[8] C. Stauffer, W.E.L. Grimson, "Learning Patterns of Activity using Real-Time Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, August 2000, pp747-757

[9] I. Haritaoglu, D. Harwood, L.S. Davis, "Real-time Surveillance of people and their Activities", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, August 2000, 22, 8, pp809-830

[10] H. Tao, H.S. Sawhney, R. Kumar, "Dynamic Layer Representation and its Applications to Tracking", *IEEE Conference on Computer Vision and Pattern Recognition*, 2, June 2000, pp134-141

[11] D. Makris, T.J. Ellis, "Learning Semantic Scene Models from Observing Activity in Visual Surveillance" *IEEE Transactions on Systems Man and Cybernetics - Part B*, (accepted for publication in 2005)

[12] A. Senior, "Tracking People with Probabilistic Appearance Models", *3rd IEEE International Workshop on Performance Evaluation of Tracking and Surveillance,* Copenhagen,1st June 2002, pp48-55

[13] J. Renno, D. Greenhill, J. Orwell and G.A. Jones "Learning the Semantic Landscape: Embedding scene knowledge in object tracking", *Realtime Imaging,, Special Issue on Video Object Processing,* (accepted for publication in 2005)

[14] B.A. Boghossian, "Motion-based Image processing Algorithms applied to Crowd Monitoring Systems", Ph.D. Thesis, King's College, University of London, Oct. 2000

[15] L.M. Fuentes, S.A. Velastin,"From tracking to advanced surveillance", *Proc. Int. Conf. on Image Processing (ICIP 2003)*, Barcelona, Spain, Sept 2003, III-121-4

[16] S. A. Velastin, M. A. Vicencio-Silva, B. Lo, L. Khoudour, "A Distributed Surveillance System For Improving Security In Public Transport Networks," *Measurement and Control*, vol. 35, No. 8, pp. 209-13, 2002

[17] S.A. Velastin, B. Boghossian, B. Lo, J. Sun, M.A. Vicencio-Silva, "PRISMATICA: Toward Ambient Intelligence In Public Transport Environments", *IEEE Transactions on Systems Man and Cybernetics*, 35, Jan 2005, pp164-182

[18] http://www.prismatica.com

[19] M. Valera, S.A. Velastin "A review of the State-of-the-Art in distributed surveillance systems", in *Intelligent Distributed Systems* (Eds. S.A. Velastin, P. Remagnino), IEE Publications 2005

[20] Intel Developer Forum, Berlin, 28 Apr. 2003 and "Lip-Reading Computers Are Born", *Internet Magazine,* 29 Apr 2003