Intelligent distributed surveillance systems: a review

M. Valera and S.A. Velastin

Abstract: This survey describes the current state-of-the-art in the development of automated visual surveillance systems so as to provide researchers in the field with a summary of progress achieved to date and to identify areas where further research is needed. The ability to recognise objects and humans, to describe their actions and interactions from information acquired by sensors is essential for automated visual surveillance. The increasing need for intelligent visual surveillance in commercial, law enforcement and military applications makes automated visual surveillance systems one of the main current application domains in computer vision. The emphasis of this review is on discussion of the creation of intelligent distributed automated surveillance systems. The survey concludes with a discussion of possible future directions.

1 Introduction

Intelligent visual surveillance systems deal with the real-time monitoring of persistent and transient objects within a specific environment. The primary aims of these systems are to provide an automatic interpretation of scenes and to understand and predict the actions and interactions of the observed objects based on the information acquired by sensors. The main stages of processing in an intelligent visual surveillance system are: moving object detection and recognition, tracking, behavioural analysis and retrieval. These stages involve the topics of machine vision, pattern analysis, artificial intelligence and data management.

The recent interest in surveillance in public, military and commercial scenarios is increasing the need to create and deploy intelligent or automated visual surveillance systems. In scenarios such as public transport, these systems can help monitor and store situations of interest involving the public, viewed both as individuals and as crowds. Current research in these automated visual surveillance systems tends to combine multiple disciplines such as those mentioned earlier with signal processing, telecommunications, management and socio-ethical studies. Nevertheless there tends to be a lack of contribution from the field of system engineering to the research.

The growing research interest in this field is exemplified by the IEEE and IEE workshops and conferences on visual surveillance [1–6] and special journal issues that focus solely on visual surveillance [7–9] or in human motion analysis [10]. This paper surveys the work on automated surveillance system from the aspects of:

- image processing/computer vision algorithms which are currently used for visual surveillance;
- surveillance systems: different approaches to the integration of the different vision algorithms to build a completed surveillance system;
- distribution, communication and system design: discussion of how such methods need to be integrated into large systems to mirror the needs of practical CCTV installations in the future.

Even though the main goal of this paper is to present a review of the work that has been done in surveillance systems, an outline of different image processing techniques, which constitute the low-level part of these systems, is included to provide a better context. One criterion of classification of surveillance systems at the sensor level (signal processing) is related to sensor modality (e.g. infrared, audio and video), sensor multiplicity (stereo or monocular) and sensor placement (centralised or distributed). This review focuses on automated video surveillance systems based on one or more stereo or monocular cameras because there is not much work reported on the integration of different types of sensors such as video and audio. However some systems [11, 12] process the information that comes from different kinds of sensors as audio and video.

1.1 Evolution of intelligent surveillance systems

The technological evolution of video-based surveillance systems started with analogue CCTV systems. These systems consist of a number of cameras located in a multiple remote location and connected to a set of monitors, usually placed in a single control room, via switches (a video matrix). In [13], for example, integration of different CCTV systems to monitor transport systems is discussed. Currently, the majority of CCTV systems use analogue techniques for image distribution and storage. Conventional CCTV cameras generally use a digital charge coupled device (CCD) to capture images. The digital image is then converted into an analogue composite video signal, which is connected to the CCTV matrix, monitors and recording equipment, generally via coaxial cables. The digital to analogue conversion does cause some picture degradation and the analogue signal is susceptible to noise. It is possible to have CCTV digital systems by taking advantage of the initial digital format of the

IEE Proceedings online no. 20041147

doi: 10.1049/ip-vis:20041147

Paper first received 16th April and in revised form 6th September 2004

The authors are with Digital Imaging Research Centre, School of Computing & Information Systems, Kingston University, UK

E-mail: m.valera@kingston.ac.uk

captured images and by using high performance computers. The technological improvement provided by these systems has led to the development of semi-automatic systems, known as second generation surveillance systems. Most of the research in second generation surveillance systems is based on the creation of algorithms for automatic real-time detection events aiding the user to recognise the events. Table 1 summarises the technological evolution of intelligent surveillance systems (1st, 2nd and 3rd generation), outlining the main problems and current research in each of them.

Table 1: Summary of technical evolution of intelligent surveillance systems

1st generation	
Techniques	Analogue CCTV systems
Advantages	– They give good performance in some situations – Mature technology
Problems	Use analogue techniques for image distribution and storage
Current research	– Digital versus analogue – Digital video recording – CCTV video compression
2nd generation	
Techniques	Automated visual surveillance by combining computer vision technology with CCTV systems
Advantages	Increase the surveillance efficiency of CCTV systems
Problems	Robust detection and tracking algorithms required for behavioural analysis
Current research	– Real-time robust computer vision algorithms – Automatic learning of scene variability and patterns of behaviours – Bridging the gap between the statistical analysis of a scene and producing natural language interpretations
3rd generation	
Techniques	Automated wide-area surveillance system
Advantages	– More accurate information as a result of combining different kind of sensors – Distribution
Problems	– Distribution of information (integration and communication) – Design methodology – Moving platforms, multi-sensor platforms
Current research	– Distributed versus centralised intelligence – Data fusion – Probabilistic reasoning framework – Multi-camera surveillance techniques

2 Applications

The increasing demand for security by society leads to a growing need for surveillance activities in many environments. Lately, the demand for remote monitoring for safety and security purposes has received particular attention, especially in the following areas:

- Transport applications such as airports [14, 15] maritime environments [16, 17], railways, underground [12, 13, 19–21], and motorways to survey traffic [22–26].
- Public places such as banks, supermarkets, homes, department stores [27–31] and parking lots [32–34].
- Remote surveillance of human activities such as attendance at football matches [35] or other activities [36–38].
- Surveillance to obtain certain quality control in many industrial processes, surveillance in forensic applications [39] and remote surveillance in military applications.

Recent events, including major terrorist attacks, have led to an increased demand for security in society. This in turn has forced governments to make personal and asset security a priority in their policies. This has resulted in the deployment of large CCTV systems. For example, London Underground and Heathrow Airport have more than 5000 cameras each. To handle this large amount of information, issues such as scalability and usability (how information needs to be given to the right people at the right time) become very important. To cope with this growing demand, research and development has been continuously carried out in commercial and academic environments to find improvements or new solutions in signal processing, communications, system engineering and computer vision. Surveillance systems created for commercial purposes [27, 28] differ from surveillance systems created in the academic world [12, 19, 40, 41], where commercial systems tend to use specific-purpose hardware and an increasing use of networks of digital intelligent cameras. The common processing tasks that these systems perform are intrusion and motion detection [11, 42–46] and detection of packages [42, 45, 46]. A technical review of commercial surveillance systems for railway applications can be found in [47].

Research in academia tends to improve image processing tasks by generating more accurate and robust algorithms in object detection and recognition [34, 48–52], tracking [34, 38, 48, 53–56], human activity recognition [57–59], database [60–62] and tracking performance evaluation tools [63]. In [64] a review of human body and movement detection, tracking and also human activity recognition is presented. Other research currently carried out is based on the study of new solutions for video communication in distributed surveillance systems. Examples of these systems are video compression techniques [66, 67], network and protocol techniques [68–70], distribution of processing tasks [71] and possible standards for data format to be sent across the network [12, 19, 62]. The creation of a distributed automatic surveillance system by developing multi-camera or multi-sensor surveillance systems, and fusion of information obtained across cameras [12, 36, 41, 72–76], or by creating an integrated system [12, 20, 53] is also an active area of research.

3 Techniques used in surveillance systems

This Section summarises research that addresses the main image processing tasks that were identified in Section 2. A typical configuration of processing modules is illustrated in Fig. 1. These modules constitute the low-level building blocks necessary for any distributed surveillance system.

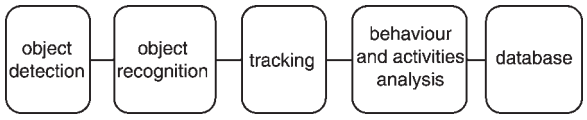

Fig. 1 *Traditional flow of processing in visual surveillance system*

Therefore, each of the following Sections outlines the most popular image processing techniques used in each of these modules. The interested reader can consult the references provided in this paper for more details on these techniques.

3.1 Object detection

There are two main conventional approaches to object detection: 'temporal difference' and 'background subtraction' The first approach consists in the subtraction of two consecutive frames followed by thresholding. The second technique is based on the subtraction of a background or reference model and the current image followed by a labelling process. After applying one of these approaches, morphological operations are typically applied to reduce the noise of the image difference. The temporal difference technique has good performance in dynamic environments because it is very adaptive, but it has a poor performance on extracting all the relevant object pixels. On the other hand, background subtraction has better performance extracting object information but it is sensitive to dynamic changes in the environment. See Figs. 2 and 3.

An adaptive background subtraction technique involves creating a background model and continuously upgrading it to avoid poor detection when there are changes in the environment. There are different techniques to model the background, which are directly related to the application. For example, in indoor environments with good lighting conditions and stationary cameras, it is possible to create a simple background model by temporally smoothing the sequence of acquired images over a short time as described in [38, 73, 74].

Outdoor environments usually have high variability in scene conditions, thus it is necessary to have robust adaptive background models, even though these robust models are computationally more expensive. A typical example is the use of a GM (Gaussian model) that models the intensity of each pixel with a single Gaussian distribution [77] or with more than one Gaussian distribution (Gaussian mixture models). In [34], due to the particular characteristics of the environment (a forest), they use a combination of two Gaussian mixture models to cope with a bimodal background (e.g. movement of trees in the wind). The authors in [59] use a mixture of Gaussians to model each pixel. The method they adopted handles slow lighting changes by slowly adapting the values of the Gaussians. A similar method is used in [78]. In [54] the background model is based on estimating the noise of each pixel in a sequence of background images. From the estimated noise the pixels that

Fig. 3 *Example of background subtraction technique used in motion detection. In this example a bounding box is drawn to fit the object detected*

represent moving regions are detected. Other techniques use groups of pixels as the basic units for tracking, and the pixels are grouped by clustering techniques combining colour information (R, G, B) and spatial dimension (x, y) to make the clustering more robust. Algorithms as such EM (expectation minimisation) are applied to track moving objects as clusters of pixels significantly different from the corresponding image reference. For example, in [79] the authors use EM to simultaneously cluster trajectories belonging to one motion behaviour and then to learn the characteristic motions of this behaviour.

In [80] the reported object detection technique is based on wavelet coefficients to detect frontal and rear views of pedestrians. By using a variant of Haar wavelet coefficients to low-level process the intensity of the images, it is possible to extract high-level information of the object (pedestrian) to detect, for example, shape information. In a training stage, the coefficients that most accurately represent the object to be detected are selected using large training sets. Once the best coefficients have been selected, they use a SVM (support vector machine) to classify the training set. During the detection stage, the selected features are extracted from the image and then the SVM is applied to verify detection of the object. The advantage of using wavelet techniques is in not having to rely on explicit colour information or textures. Therefore they can be useful in applications where there is a lack of colour information (a usual occurrence in indoor surveillance). Moreover, using wavelets implies a significant reduction of data in the

Fig. 2 *Example of temporal difference technique used in motion detection*

learning stage. However, the authors only model front and rear views of pedestrians. In the case of groups of people that stop, talk or walk perpendicular to the view of the camera, the algorithm is not able to detect people. Furthermore, an object, with similar intensity characteristics to the front or rear of a human, is likely to generate a false positive. Another line of research is based on the detection of contours of persons by using principal component analysis (PCA). Finally, as far as motion segmentation is concerned, techniques based on optic flow may be useful when a system uses moving cameras as in [26], although there are known problems when the image size of the objects to be tracked is small.

3.2 Object recognition, tracking and performance evaluation

Tracking techniques can be split into two main approaches: 2-D models with or without explicit shape models and 3-D models For example, in [26] the 3-D geometrical models of a car, a van and a lorry are used to track vehicles on a highway. The model-based approach uses explicit *a priori* geometrical knowledge of the objects to follow, which in surveillance applications are usually people, vehicles or both. In [24] the author uses two 2-D models to track cars: a rectangular model for a passing car that is close to the camera and a U-shape model for the rear of a car in the distance or just in front of the camera. The system consists of an image acquisition module, a lane and car detector, a process co-ordinator and a multiple car tracker. In some multi-camera systems like [74], the focus is on extracting trajectories, which are used to build a geometric and probabilistic model for long-term prediction, and not the object itself. The *a priori* knowledge can be obtained by computing the object's appearance as a function of its position relative to the camera. The scene geometry is obtained in the same way. In order to build shape models, the use of camera calibration techniques becomes important. A survey of different techniques for camera calibration can be found in [81]. Once *a priori* knowledge is available, it may be utilised in a robust tracking algorithm dealing with varying conditions such as changing illumination, offering a better performance in solving (self) occlusions or (self) collisions. It is relatively simple to create constraints in the objects' appearance model by using model-based approaches; e.g. the constraint that people appear upright and in contact with the ground is commonly used in indoor and outdoor applications.

The object recognition task then becomes a process of utilising model-based techniques in an attempt to exploit such knowledge. A number of approaches can be applied to classify the new detected objects. The integrated system presented in [53] and [26] can recognise and track vehicles using a defined 3-D model of a vehicle, giving its position in the ground plane and its orientation. It can also recognise and track pedestrians using a prior 2-D model silhouette shape, based on B-spline contours. A common tracking method is to use a filtering mechanism to predict each movement of the recognised object. The filter most commonly used in surveillance systems is the Kalman filter [53, 73]. Fitting bounding boxes or ellipses, which are commonly called 'blobs', to image regions of maximum probability is another tracking approach based on statistical models. In [77] the author models and tracks different parts of a human body using blobs, which are described in statistical terms by a spatial and colour Gaussian distribution. In some situations of interest the assumptions made to apply linear or Gaussian filters do not hold, and then nonlinear Bayesian filters, such as extended Kalman filters (EKF) or particle filters have been proposed. Work described in [82] illustrates that in highly non-linear environments particle filters give better performance than EKF. A particle filter is a numerical method, which weights (or 'particle') a representation of posterior probability densities by resampling a set of random samples associated with a weight and computing the estimate probabilities based on these weights. Then, the critical design decision using particle filters relies on the choice of importance (the initial weight) of the density function.

Another tracking approach consists in using connected-components [34] to segment the changes in the scene into different objects without any prior knowledge. The approach gives good performance when the object is small, with a low-resolution approximation, and the camera placement is chosen carefully. HMMs (hidden Markov models) have also been used for tracking purposes as presented in [40], where the authors use an extension of HMM to predict and track objects trajectories. Although HMM filters are suitable for dynamic environments (because there is no assumption in the model or in the characterisation of the type of the noise, as is required when using Kalman filters), offline training data are required. Recent research has been carried out on the creation of semi-automatic tools that can help create the large set of ground truth data that is necessary for evaluating the performance of the tracking algorithms [63].

3.3 Behavioural analysis

The next stage of a surveillance system recognises and understands activities and behaviours of the tracked objects This stage broadly corresponds to a classification problem of the time-varying feature data that are provided by the preceding stages. Therefore, it consists in matching a measured sequence to a pre-compiled library of labelled sequences that represent prototypical actions that need to be learnt by the system via training sequences. There are several approaches for matching time-varying data. Dynamic time warping (DTW) is a time-varying technique widely used in speech recognition, image patterns as in [83] and recently in human movement patterns [84]. It consists of matching a test pattern with a reference pattern. Although it is a robust technique, it is now less favoured than dynamic probabilistic network models like HMM (hidden Markov models) and Bayesian networks [85, 86]. The last time-varying technique that is not as widespread as HMM, because it is less investigated for activity recognition, is neural networks (NN). In [57] the recognition of behaviours and activities is done using a declarative model to represent scenarios, and a logic-based approach to recognise pre-defined scenario models.

3.4 Database

One of the final stages in a surveillance system is storage and retrieval (the important aspects of user interfaces and alarm management are not considered here due to lack of space) Relatively little research has been done in how to store and retrieve all the obtained surveillance information in an efficient manner, especially when it is possible to have different data formats and types of information to retrieve. In [62] the authors investigate the definition and creation of data models to support the storage of different levels of abstraction of tracking data into a surveillance database. The database module is part of a multi-camera system that is presented in Fig. 4.

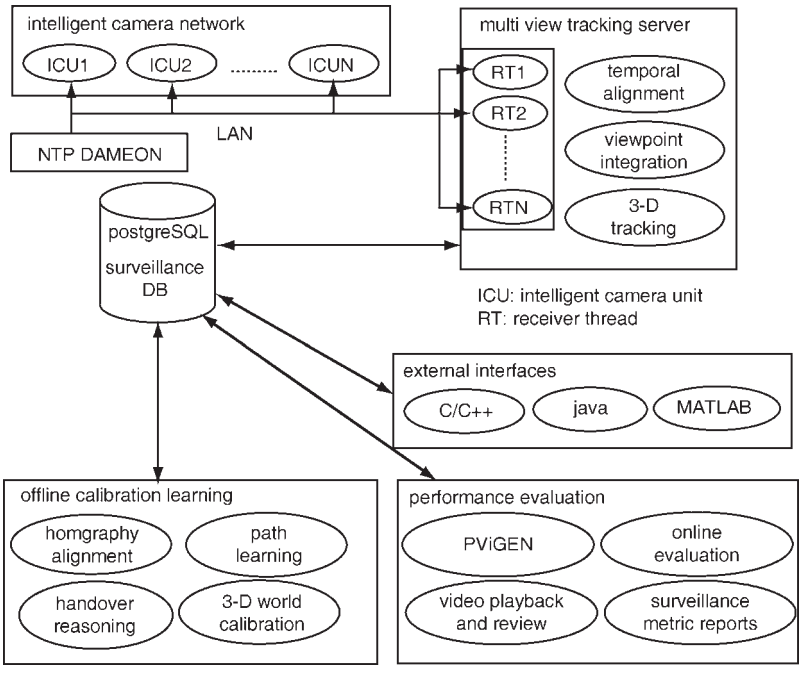

Fig. 4 *Architecture of multi-camera surveillance system (from Makris et al. [62])*

In [61] the authors develop a data model and a rule-based query language for video content based indexing and retrieval. Their data model allows facts as well as objects and constraint. Retrieval is based on a rule-based query language that has declarative and operational semantics, which can be used to gather relations between information represented in the model. A video sequence is split into a set of fragments and each fragment can be analysed to extract the information (symbolic descriptions) of interest to store into the database. In [60] retrieval is performed on the basis of object classification. A stored video sequence consists of 24 frames; the last frame is the key frame that contains information about the whole sequence. Retrieval is performed using a feature vector where each component contains information obtained from the event detection module.

4 Review of surveillance systems

The previous Section reviewed some core computer vision techniques that are necessary for the detection and understanding of activity in the context of surveillance. It is important to highlight that the availability of a given technique or set of techniques is necessary but not sufficient to deploy a potentially large surveillance system, which implies networks of cameras and distribution of processing capacities to deal with the signals from these cameras. Therefore in this section we review what has been done to propose surveillance systems that address these requirements. The majority of the surveillance systems reviewed in this paper are based on transport or parking lot applications. This is because most reported distributed systems tend to originate from academic research which has tended to focus on these domains (e.g. by using university campuses for experimentation or the increasing research funding to investigate solutions in public transport).

4.1 Third generation surveillance systems

Third generation surveillance systems is the term sometimes used in the literature to refer to systems conceived to deal with a large number of cameras, a geographical spread of resources, many monitoring points, and to mirror the hierarchical and distributed nature of the human process of surveillance Those are important prerequisites, if such systems are going to be integrated as part of a management tool. From an image processing point of view, they are based on the distribution of processing capacities over the network and the use of embedded signal processing devices to give the advantages of scalability and robustness potential of distributed systems. The main goals that are expected of a generic third generation vision surveillance application, based on end-user requirements, are to provide good scene understanding, oriented to attract the attention of the human operator in real time, possibly in a multi-sensor environment, surveillance information and using low cost standard components.

4.2 General requirements of third generation of surveillance systems

Spatially distributed multi-sensor environments present interesting opportunities and challenges for surveillance. Recently, there has been some investigation of data fusion techniques to cope with the sharing of information obtained from different types of sensors [41]. The communication aspects within different parts of the system play an important role, with particular challenges either due to bandwidth constraints or the asymmetric nature of the communication [87].

Another relevant aspect is the security of communications between modules. For some vision surveillance systems, data might need to be sent over open networks and there are critical issues in maintaining privacy and authentication [87]. Trends in the requirements of these systems include the desirability of adding automatic learning capability to provide the capability of characterising models of scenes to be recognised as potentially dangerous events [57, 85, 86, 88]. A state-of-the-art survey on approaches to learn, recognise and understand scenarios may be found in [89].

4.3 Examples of surveillance systems

The distinction between surveillance for indoor and outdoor applications occurs because of the differences in the design

at the architectural and algorithmic implementation levels. The topology of indoor environments is also different from that of outdoor environments.

Typical examples of commercial surveillance systems are DETEC [27], Gotcha [28] or [29]. They are usually based on what is commonly called motion detectors, with the option of digital storage of the events detected (input images and time-stamped metadata). These events are usually triggered by objects appearing in the scene. DETEC is based on specialised hardware that allows one to connect up to 12 cameras to a single workstation. The workstation can be connected to a network and all the surveillance data can be stored in a central database available to all workstations on the network. Visualisation of the input images from the camera across internet links is described in [29].

Another example of a commercial system intended for outdoor applications, is DETER [18, 78] (detection of events for threat evaluation and recognition). The architecture of the DETER system is illustrated in Fig. 5. It is aimed at reporting unusual moving patterns of pedestrians and vehicles in outdoor environments such as car parks. The system consists of two parts: the computer vision module and the threat assessment or alarms management module. The computer vision part deals with the detection, recognition and tracking of objects across cameras. In order to do this, the system fuses the views of multiple cameras into one view and then performs tracking of the objects. The threat assessment part consists of feature assembly or high-level semantic recognition, the off-line training and the on-line threat classifier. The system has been evaluated in a real environment by end-users, and it had good performance in object detection and recognition. However, as is pointed out in [78], DETER employs a relatively small number of cameras because it is a cost-sensitive application. It is not clear whether the system has the functionality for retrieval and even though the threat assessment performance is good, there is no feedback loop in this part that could help improve performance.

Another integrated visual surveillance system for vehicles and pedestrians in parking lots is presented in [53]. This system has a novel approach to deal with interactions between objects (vehicles and pedestrians) in a hybrid tracking system. The system consists of two visual modules capable of identifying and tracking vehicles and pedestrians in a complex dynamic scene. However, this is an

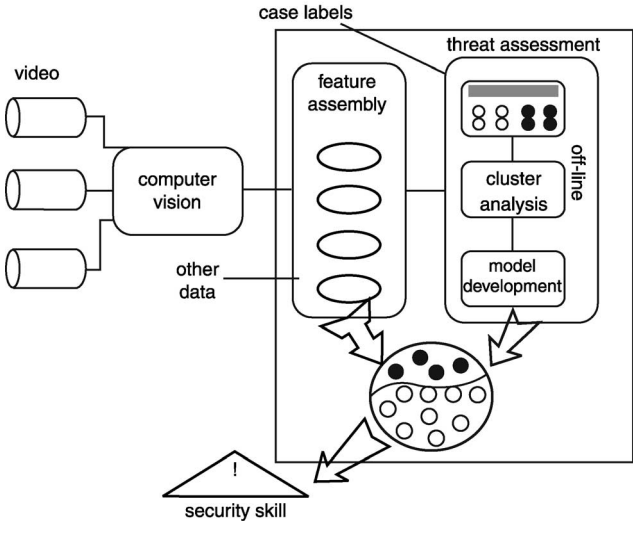

Fig. 5 *Architecture of DETER system (from I. Pavlidis et al. [78])*

example of a system that considers tracking as the only surveillance task, even though the authors pointed out in [53] the need for a semantic interpretation of the tracking results for scene recognition. Furthermore, a 'handover' tracking algorithm across cameras has not been established.

It is important to have a semantic interpretation of the behaviours of the recognised objects in order to build an automated surveillance system that is able to recognise and learn from the events and interactions that occur in a monitored environment. For example in [90], the authors illustrated a video-based surveillance system to monitor activities in a parking lot that performs a semantic interpretation of recognised events and interactions. The system consists of three parts: the tracker which tracks the objects and collects their movements into partial tracks; the event generator, which generates discrete events from the partial tracks according to a simple environment model and finally, a parser that analyses the events according to a stochastic context-free grammar (SCFG) model which structurally describes possible activities. This system, as the one in [53], is aimed at proving the algorithms more than at creating a surveillance system for monitoring a wide area (the system uses a single stationary camera). Furthermore, it is not clear how the system distinguishes between cars and pedestrians because the authors do not use any shape model.

In [25] visual traffic surveillance for automatic identification and description of the behaviour of vehicles within parking lots scenes is presented. The system consists of a motion module, model visualisation and pose refinement, tracking and trajectory-based semantic interpretation of vehicle behaviour. The system uses a combination of colour cues and brightness information to construct the background model and applies connectivity information for pixel classification. Using camera calibration information they project the 3-D model of a car onto the image plane and they use the 3-D shape model-based method for pose evaluation. The tracking module is performed using EKF (extended Kalman filters). The semantic interpretation module is realised by three steps: trajectory classification, then an on-line classification step using Bayesian classifiers, and finally natural language descriptions are applied to the trajectory patterns of the cars that have been recognised. Although this system introduces a semantic interpretation for car behaviours, it is not clear how this system handles the interactions of several objects in the same scene, and consequently the occlusions between objects. Another possible limitation is the lack of different models to represent different types of vehicles (c.f. [53], which includes separate 3-D models for a car, van and lorry).

Other surveillance systems, which have been applied to different applications (e.g. road traffic, ports, and railways), can be found in [13, 16, 21–23]. These automatic or semi-automatic surveillance systems apply more or less intelligent and robust algorithms to assist the end-user. The importance to this review of some of these systems is the illustration of how the requirements of wide geographical distribution impinge on system architecture aspects.

The author in [13] expresses the need to integrate video-based surveillance systems with existing traffic control systems to develop the next generation of advanced traffic control and management systems. Most of the technologies in traffic control are based on CCTV technology linked to a control unit and in most cases for reactive manual traffic monitoring. However, there are an increasing number of CCTV systems using image processing techniques in urban road networks and highways. Therefore, the author in [13] proposes to combine these systems with other existing surveillance traffic systems like surveillance systems based

on networks of smart cameras. The term 'smart camera' (or 'intelligent camera') is normally used to refer to a camera that has processing capabilities (either in the same casing or nearby), so that event detection and storage of event video can be done autonomously by the camera. Thus, normally, it is only necessary to communicate with a central point when significant events occur.

Usually integrated surveillance systems consist of a control unit system, which manages the outputs from the different surveillance systems, a surveillance signal processing unit and a central processing unit which encapsulates a vehicle ownership database. The suggestion in [13] of having a control unit, which is separated from the rest of the modules, is an important aspect in the design of a third generation surveillance system. However, to survey a wide area implies geographical distribution of equipment and a hierarchical structure of the personnel who deal with security. Therefore for better scalability, usability, and robustness of the system, it is desirable to have more than one control unit. Their design is likely to follow a hierarchical structure (from low-level to high-level control) that mirrors what is done in image processing where there is a differentiation between low-level and high-level processing tasks.

Continuing with traffic monitoring applications, in [22] a wide-area traffic monitoring system for highway roads in Italy is presented. The system consists of two main control rooms, which are situated in two different geographical places, and nine peripheral control rooms, which are in direct charge of road operation: toll collection, maintenance and traffic control. Most of the sensors used to control traffic are CCTVs. Images are centrally collected and displayed in each peripheral control room. They have installed PTZ (pan, tilt and zoom) colour cameras in places where the efficiency of CCTV is limited, e.g. by weather conditions. The system is able to detect automatically particular conditions and therefore to attract human attention. Each peripheral control room receives and manages, in a multi-session environment, the MPEG-1 compressed video for full motion traffic images at transmission rates up to 2 Mbps, from each peripheral site. There is integration of image acquisition, coding and transmission subsystems in each peripheral site. In some peripheral sites that control tunnels, they have a commercial subsystem that detects stopped vehicles or queues. Even though this highway traffic monitoring system is not fully automatic, it shows the importance of having a hierarchical structure of control and image processing units. Moreover, it shows the importance of coding and transmission bandwidth requirements for wide-area surveillance systems.

The authors in [23] present a video-based surveillance system for measuring traffic parameters. The aim of the system is to capture video from cameras that are placed on poles or other structures looking down at traffic. Once the video is captured, digitised and processed by onsite embedded hardware, it is transmitted in summary form to a transportation management centre (TMC) for computing multi-site statistics like travel times. Instead of using 3-D models of vehicles as in [25] or [26], the authors use feature-based models like corners, which are tracked from entry to exit zones defined off-line by the user. Once these corner features have been tracked, they are grouped into single candidate vehicles by the sub-features grouping module. This grouping module constructs a graph over time where vertices are sub-feature tracks, edges are grouping relationships between tracks, and connected components correspond to the candidate vehicle. When the last track of a connected component enters the exit region, a new candidate vehicle is generated and the component is removed from the grouping graph. The system consists of a host PC connected to a network of 13 DSPs (digital signal processors). Six of these DSPs perform the tracking, four the corner detection, and one acts as the tracker controller. The tracker controller is connected to a DSP that is an image frame-grabber and to another DSP which acts as a display. The tracker update is sent to the host PC, which runs the grouper due to memory limitations. The system has good performance not only in congested traffic conditions but also at night-time and in urban intersections.

Following the aim of [23], the authors in [37] develop a vision-based surveillance system to monitor traffic flow on a road, but focusing on the detection of cyclists and pedestrians. The system consists of two main distributed processing modules: the tracking module, which processes in real time and is placed roadside on a pole, and the analysis module, which is performed off-line in a PC. The tracking module consists of four tasks: motion detection, filtering, feature extraction using quasi-topological features (QTC) and tracking using first-order Kalman filters. The shape and trajectory of the recognised objects are extracted and stored in a removable memory card, which is transferred to the PC to achieve the analysis process using learning vector quantisation to produce the final counting. This system has some shortcomings. The image algorithms are not robust enough (the background model is not robust enough to cope with changing conditions or shadows) and depend on the position of the camera. The second problem is that even though tracking is performed in real time, the analysis is performed off-line, therefore it is not possible to do flow statistics or monitoring in real time.

In [16] the architecture of a system for surveillance in a maritime port is presented. The system consists of two subsystems: image acquisition and visualisation. The architecture is based on a client/server design. The image acquisition subsystem has a video server module, which can handle four cameras at the same time. This module acquires the images from camera streams, which are compressed, and then the module broadcasts the compressed images to the network using TCP/IP and at the same time records the images on hard disks. The visualisation module is performed by client subsystems, which are based on PC boards. This module allows the selection of any camera using a pre-configured map and the configuration of the video server. Using an internet server module it is possible to display the images through the internet. The system is claimed to have the capability of supporting more than 100 cameras and 100 client stations at the same time, even though the reported implementation had 24 cameras installed mainly at the gates of the port. This is an example of a simple video surveillance system (with no image interpretation), which only consists of image acquisition, distribution and display. The interesting point in this system is to see the use of a client and server architecture to deal with the distribution of the multiple digital images. Moreover, the acquisition and visualisation modules have been encapsulated in a way such that scalability of the system can be accomplished in a straightforward way, by integrating modules into the system in a 'drop' operation.

In [21] a railway station CCTV surveillance system in Italy is presented. Similar to [22], the system has a hierarchical structure distributed between main (central) control rooms and peripheral site (station) control rooms. The tasks that are performed in the central control room are acquisition and display of the live or recorded images. The system also allows the acquisition of images from all the station control rooms through communication links and through specific coding and decoding devices. Digital

recording, storage and retrieval of the image sequences as well as the selection of a specific CCTV camera and the deactivation of the alarm system are carried out in the central room. The main tasks performed in each station control room are acquisition of the images from the local station CCTV cameras, to link with the central control room to transmit the acquired or archived images in real time, and to receive configuration procedures. The station control room also handles the transmission of an image of a specific CCTV camera at higher rate either by request or automatically when an alarm has been raised. The management and deactivation of local alarms is handled from the station control room. Apart from the central control room and the station control rooms, there is a crisis room for the management of railway emergencies. Although this system is a semi-automatic, hierarchical and distributed surveillance system, the role played by human operators is still central because there is no processing (object recognition or motion estimation) to channel the attention of the monitoring personnel.

Ideally, a third generation of surveillance system for public transport applications would provide a high level of automation in the management of information as well as that of alarms and emergencies. That is the stated aim of the following two surveillance system research projects (other projects in public transportation that are not included here can be found in [47]).

CROMATICA [20] (crowd monitoring with telematic and communication assistance) was an EU-funded project whose main goal was to improve the surveillance of passengers in public transport, enabling the use and integration of technologies like video-based detection and wireless transmission. This was followed by another EU-funded project called PRISMATICA [12] (pro-active integrated systems for security management by technological institutional and communication assistance) that looked at social, ethical, organisational and technical aspects of surveillance for public transport. A main technical output was a distributed surveillance system. It is not only a wide-area video-based distributed system like ADVISOR (annotated digital video for intelligent surveillance and optimised retrieval) [19], but it is also a wide-area multi-sensor distributed system, receiving inputs from CCTV, local wireless camera networks, smart cards and audio sensors. PRISMATICA then consists of a network of intelligent devices (that process sensor inputs) that send and receive messages to/from a central server module (called 'MIPSA') that co-ordinates device activity, archives/retrieves data and provides the interface with a human operator. Figure 6

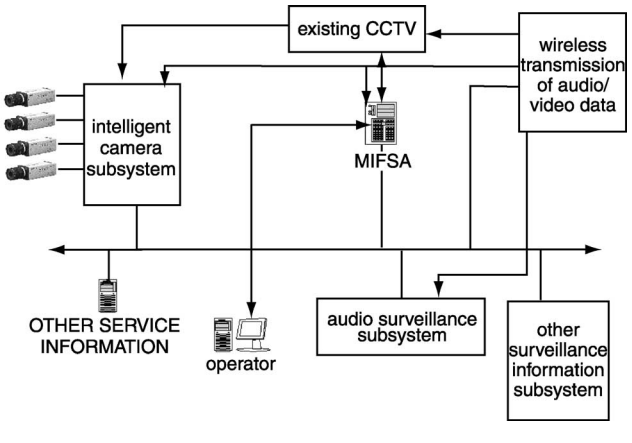

Fig. 6 *Architecture of PRISMATICA system (from B. Ping Lai Lo et al. [12])*

shows the architecture of PRISMATICA. Similarly to ADVISOR (see below), PRISMATICA uses a modular and scalable architecture approach using standard commercial hardware.

ADVISOR was also developed as part of an EU-funded project. It aims to assist human operators by automatic selection, recording and annotation of images that have events of interest. In other words, ADVISOR interprets shapes and movements in scenes being viewed by the CCTV to build up a picture of the behaviour of people in the scene. ADVISOR stores all video output from cameras. In parallel with recording video information, the archive function stores commentaries (annotations) of events detected in particular sequences. The archive video can be searched using queries for the annotation data, or according to specific times. Retrieval of video sequences can take place alongside continuous recording. ADVISOR is intended to be an open and scalable architecture approach and is implemented using standard commercial hardware with an interface to a wide-bandwidth video distribution network. Figure 7 shows a possible architecture of the ADVISOR system. It consists of a network of ADVISOR units, each of which is installed in a different underground station and consists of an object detection and recognition module, tracking module, behavioural analysis and database module.

Although both systems are classified as distributed architectures, they have a significant main difference in that PRISMATICA employs a centralised approach whereas ADVISOR can be considered as a semi-distributed architecture. PRISMATICA is built with the concept of a main or central computer which controls and supervises the whole system. This server thus becomes a critical single point of failure for the whole system. ADVISOR can be seen as a network of independent dedicated processor nodes (ADVISOR units), avoiding a single point-of-failure. Nevertheless, each node is a rack with more than one CPU and each node contains a central computer, which controls the whole node, therefore there is still a single point-of-failure within each node. The number of CPUs in each node is directly proportional to the number of existing image processing modules, making the system difficult to scale and hard to build in cost-sensitive applications.

In [91] the authors report the design of a surveillance system with no server to avoid this centralisation, making all the independent subsystems completely self-contained, and then setting up all these nodes to communicate with each other without having a mutually shared communication point. This approach avoids the disadvantages of the centralised server, and moves all the processes directly to the camera making the system a group of smart cameras connected across the network. The fusion of information between 'crunchers' (as they are referred to in the article) is done through a defined protocol, after the configuration of the network of smart cameras or 'crunchers'. The defined protocol has been validated with a specific verification tool called 'spin'. The format of the information to share between 'crunchers' is based on a common data structure or object model with different stages depending on whether the object is recognised or is migrating from the field of view of one camera to another. However, the approach to distributed design is to build using specific commercial embedded hardware (EVS units). These embedded units consist of a camera, processor, frame grabber, network adapter and database. Therefore, in cost-sensitive applications where a large number of cameras are required, this approach might be unsuitable.

As part of the VSAM project, [76] presents a multi-camera surveillance system following the same idea as [92],

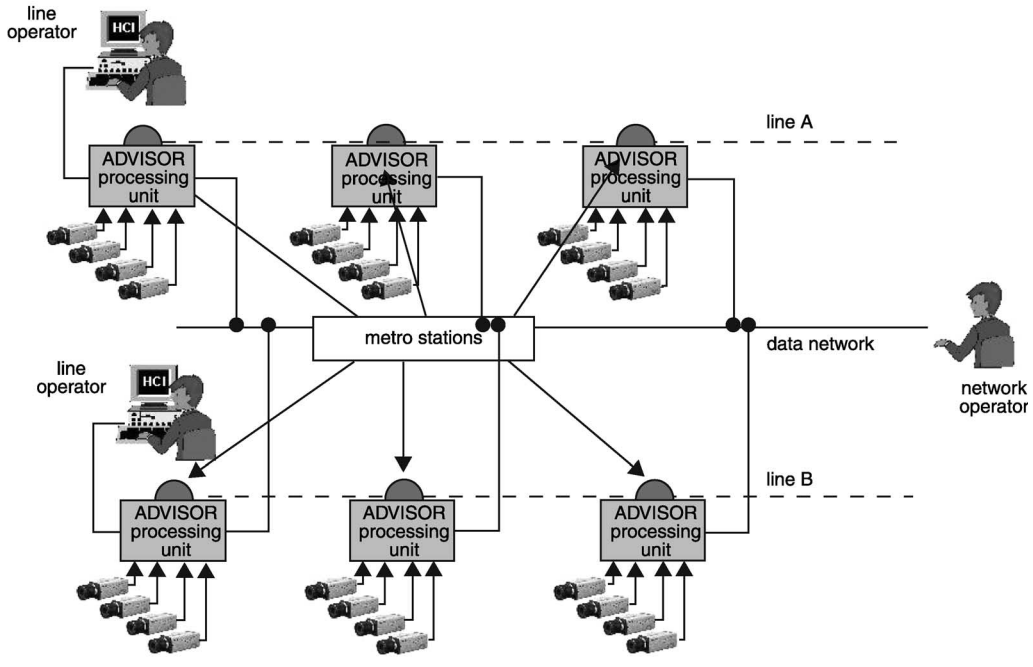

Fig. 7 *Proposed architecture of ADVISOR system (from [19]). Dashed black lines correspond to metro railway and red lines correspond to computer links*

i.e. the creation of a network of 'smart' sensors that are independent and autonomous vision modules. Nevertheless in [76], these sensors are capable of detecting and tracking objects, classifying the moving objects into semantic categories such as 'human' or 'vehicle' and identifying simple human movements such as walking. In [92] the smart sensors are only able to detect and track moving objects. Moreover, the algorithms in [92] are based on indoor applications. Furthermore, in [76] the user can interact with the system. To achieve this interactivity, there are system-level algorithms which fuse sensor data, perform the processing tasks and display the results in a comprehensible manner. The system consists of a central control unit (OCU) which receives the information from multiple independent remote processing units (SPU). The OCU interfaces with the user through a GUI module.

Monitoring wide areas requires the use of a significant number of cameras to cover as much area as possible and to achieve good performance in the automatic surveillance operation. Therefore, the need to co-ordinate information across cameras becomes an important issue. Current research points towards developing surveillance systems that consist of a network of cameras (monocular, stereo, static or PTZ (pan tilt zoom)) which perform the type of vision algorithms that we have reviewed earlier, but also using information from neighbouring cameras. The following Sections highlight the main work in this field.

4.4 Co-operative camera systems

An application of surveillance of human activities for sports application is presented in [35]. The system consists of eight cameras, eight feature server processes and a multi-tracker viewer. Only the cameras are installed on the playing area, and the raw images are sent through optical fibres to each feature server module. Each module realises segmentation, single-view tracking and object classification and sends the results to the multi-tracker module, which merges all the information from the single-view trackers using a nearest neighbour method based on the Mahalanobis distance.

CCN [18] (co-operative camera network) is an indoor application surveillance system that consists of a network of nodes. Each node is composed of a PTZ camera connected to a PC and a central console to be used by the human operator. The system reports the presence of a visually tagged individual inside the building by assuming that human traffic is sparse (an assumption that becomes less valid as crowd levels increase). Its purpose is to monitor potential shoplifters in department stores.

In [33] a surveillance system for a parking lot application is described. The architecture consists of one or more static camera subsystems (SCS) and one or more active camera subsystems (ACS). First, the target is detected and tracked by the static subsystems, once the target has been selected a PTZ, which forms the ACS, is activated to capture high resolution video of the target. Data fusion for the multi-tracker is done using the Mahalanobis distance. Kalman filters are used for tracking, as in [35].

In [36] the authors present a multi-camera tracking system that is included in an intelligent environment system called 'EasyLiving' which aims at assisting the occupants of that environment by understanding their behaviour. The multi-camera tracking system consists of two sets of stereo cameras (each set has three small colour cameras). Each set is connected to a PC that runs the 'stereo module'. The two stereo modules are connected to a PC which runs the tracker module. The output of the tracker module is the localisation and identity of the people in the room. This identity does not correspond to the natural identity of the person, but to an internal temporary identity which is generated for each person using a colour histogram provided by the stereo module each time. The authors use the depth and colour information provided from the cameras to apply background subtraction and to allocate 3-D blobs, which are merged into person shapes by clustering regions. Each stereo module reports the 2-D ground plane locations of its person blobs to the tracking module. Then, the tracker module uses knowledge of the relative locations of the cameras, field of view, and heuristics of the movement of people to produce the locations and identities of the people in the room. The performance of the tracking system is good when

there are fewer than three people in the room and when the people wear different colour outfits, otherwise, due to the poor clustering results, performance is reduced drastically.

In [92] an intelligent video-based visual surveillance system (IVSS) is presented which aims to enhance security by detecting certain types of intrusion in dynamic scenes. The system involves object detection and recognition (pedestrians and vehicles) and tracking. The system is based on a distribution of a static multi-camera monitoring module via a local area network. The design architecture of the system is similar to ADVISOR [19], and the system consists of one or more clients plus a server, which are connected through TCP/IP. The clients connect only to the server (and not to other clients), while the server talks with all clients. Therefore there is no data fusion across cameras. The vision algorithms are developed in two stages: hypothesis generation (HG) and hypothesis verification (HV). The first stage realises a simple background subtraction. The second stage compensates the non-robust background subtraction model. This stage is essentially a pattern classification problem and it uses a Gabor filter to extract features, e.g. strong edges and lines at different orientation of vehicles and pedestrians, and support vector machines (SVM) to perform the classifications. Although this is an approach to developing a distributed surveillance system, there is no attempt at fusing information across cameras. Therefore it is not possible to track objects across clients. Furthermore, the vision algorithms do not include activity recognition and although the authors claim to compensate the simple motion detection algorithm using the Gabor filters, it is not clear how these filters and SVM cope with uncertainties in the tracking stage, e.g. occlusions or shadows.

In [72] a multi-camera surveillance system for face detection is illustrated. The system consists of two cameras (one of the cameras is a CCD pan-tilt and the other is a remote control camera). The system architecture is based on three main modules using a client/server approach as solution for the distribution. The three modules are sensor control, data fusion and image processing. The sensor control module is a dedicated unit to control directly the two cameras and the information that flows between them. The data fusion module controls the position of the remote control camera depending on the inputs received from the image processing and sensor control module. It is interesting to see how the authors use the information obtained from the static camera (the position of the recognised object) to feed the other camera. Therefore, the remote control camera can zoom to the recognised human to detect the face.

An interesting example of a multi-tracking camera surveillance system for indoor environments is presented in [73]. The system is a network of camera processing modules, each of which consists of a camera connected to a computer, and a control module, which is a PC that maintains the database of the current objects in the scene. Each camera processing module realises the tracking process using Kalman filters. The authors develop an algorithm which divides the tracking task between the cameras by assigning the tracking to the camera which has better visibility of the object, taking into account occlusions. This algorithm is implemented in the control module. In this way, unnecessary processing is reduced. Also, it makes it possible to solve some occlusion problems in the tracker by switching from one camera to another camera when the object is not visible enough. The idea is interesting because it shows a technique that exploits distributed processing to improve detection performance. Nevertheless, the way that the algorithm decides which camera is more appropriate is performed using a 'quality service of tracking' function. This function is defined based on the sizes of the objects in the image, estimated from the Kalman filter, and the object occlusion status. Consequently, in order to calculate the size of the object with respect to the camera, all cameras have to try to track the object. Moreover, the system has been built with the constraint that all cameras have overlapping views (if there were topographic knowledge of the cameras the calculation of this function could be applied only to the cameras which have overlapping views). Furthermore, in zones where there is a gap between views, the quality service of tracking function would drop to zero, and if the object reappears it would be tracked as a new object.

VIGILANT [32] is a multi-camera surveillance system (Fig. 8) which monitors pedestrians walking in a parking lot. The system tracks people across cameras using software agents. For each detected person in each camera an agent is created to hold the information. The agents communicate to obtain a consensus decision of whether or not they are assigned the same person who is being seen from different cameras by reasoning on trajectory geometry in the ground plane.

As has been illustrated, in a distributed multi-camera surveillance system, it is important to know the topology of the links between the cameras that make up the system in order to recognise, understand and follow an event that may be captured on one camera and to follow it in other cameras. Most of the multi-camera systems that have been discussed in this review use a calibration method to compute the network camera topology. Moreover, most of these systems try to combine tracks of the same target that are simultaneously visible in different camera views.

In [62] the authors present a distributed multi-camera tracking surveillance system for outdoor environments (its architecture can be seen in Fig. 4). An approach is presented which is based on learning a probabilistic model of an activity in order to establish links between camera views in a correspondence-free manner. The approach can be used to calibrate the network of cameras and does not require correspondence information. The method correlates the number of incoming and outgoing targets for each camera view, through detected entry and exit points. The entry and exit zones are modelled by a GMM and initially these zones are learnt automatically from a database using an EM algorithm. This approach provides two main advantages: no previous calibration method is required and the system allows tracking of targets across the 'blind' regions between camera views. The first advantage is particularly useful because of the otherwise resource-consuming process of camera calibration for wide-area distributed multi-camera surveillance systems with a large number of cameras [19, 21, 22, 47].

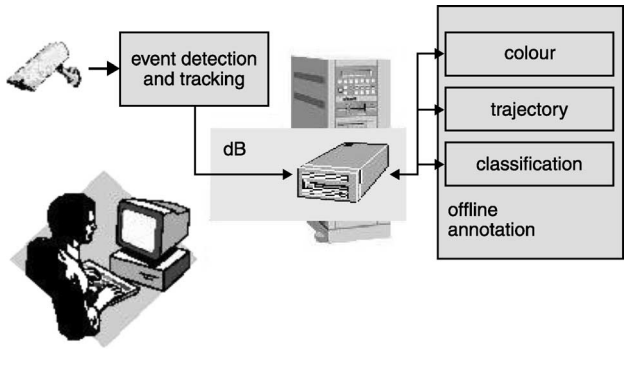

Fig. 8 *Architecture of VIGILANT system (from D. Greenhill et al. [32])*

5 Distribution, communication and system design

In Section 3 we considered different techniques that have been applied to develop more robust and adaptive algorithms. In Section 4 we presented a review of different architectures of distributed surveillance systems. Although the design of some of these systems can look impressive, there are some aspects where it will be advantageous to dedicate more attention to the development of distributed surveillance systems for the future. These include the distribution of processing tasks, the use of new technologies as well as the creation of metadata standards or new protocols to cope with current limitations in bandwidth capacities. Other aspects that should be taken into consideration for the next generation of surveillance systems are the design of scheduling control and more robust and adaptive algorithms. A field that needs further research is that of alarm management, which is an important part of an automatic surveillance system, e.g. when different priorities and goals need to be considered. For example in [93] the authors describe work carried out in a robotics field, where the robot is able to focus attention on a certain region of interest, extract its features and recognise objects in the region. The control part of the system allows the robot to refocus its attention on a different region of interest, and skip a region of interest that already has been analysed. Another example can be found in [19] where in the specification of the system, system requirements like 'to dial an emergency number automatically if a specific alarm has been detected' are included. To be able to carry out these kinds of actions command and control systems must be included as an integral part of a surveillance system.

Other work worth mentioning in the context of large distributed systems is the extraction of information from compressed video [65], dedicated protocols for distributed architectures [69, 94, 95], and real-time communications [96]. Work has also been conducted to build an embedded autonomous unit as part of a distributed architecture [30, 68, 91]. Several researchers are dealing with PTZ [54, 72] because this kind of camera can survey wider areas and can interact in more efficient ways with the end-users who can zoom when necessary. It is also important to incorporate scheduling policies to control resource allocation as illustrated in [97]. Work in multiple robot systems [98] illustrates how limited communications bandwidth affects robot performance and how this performance is linked to the number of robots that share the bandwidth. A similar idea is presented in [71] and [99] for surveillance systems, while in [94], an overview of the state-of-the-art of multimedia communication technologies and a standard is presented. On the whole, the work on intelligent distributed surveillance systems has been led by computer vision laboratories, perhaps at the expense of system engineering issues. It is essential to create a framework or methodology for designing distributed wide-area surveillance systems, from the generation of requirements to the creation of design models by defining functional and intercommunication models as is done in the creation of distributed concurrent real-time systems in other disciplines like control systems in aerospace. Therefore, as has been mentioned earlier in this paper, in the future the realisation of a wide-area distributed intelligent surveillance system should be through a combination of different disciplines: computer vision, telecommunications and system engineering being clearly needed. Work related to the development of a design framework for developing video surveillance systems can be found in [91, 99, 100]. Distributed virtual applications are discussed in [101], and embedded architectures in [102]. For example, much could be borrowed from the field of autonomous robotic systems on the use of multi-agents, where non-centralised collections of relatively autonomous entities interact with each other in a dynamic environment. In a surveillance system, one of the principal costs is the sensor suite and payload. A distributed multi-agent approach may offer several advantages. First, intelligent co-operation between agents may allow the use of less expensive sensors and therefore a larger number of sensors may be deployed over a greater area. Second, robustness is increased, since even if some agents fail, others remain to perform the mission. Third, performance is more flexible, there is a distribution of tasks at various locations between groups of agents. For example, the likelihood of correctly classifying an object or target increases if multiple sensors are focused on it from different locations.

6 Conclusions

This paper has presented the state of development of intelligent distributed surveillance systems, including a review of current image processing techniques that are used in different modules that constitute part of surveillance systems. Looking at these image processing tasks, it has identified research areas that need to be investigated further such as adaptation, data fusion and tracking methods in a co-operative multi-sensor environment, extension of techniques to classify complex activities and interactions between detected objects. In terms of communication or integration between different modules it is necessary to study new communication protocols and the creation of metadata standards. It is also important to consider improved means of task distribution that optimise the use of central, remote facilities and data communication networks. Moreover, one of the aspects that the authors believe is essential in the future for the development of distributed surveillance systems is the definition of a framework to design distributed architectures firmly rooted in systems engineering best practice, as used in other discipline such as control aerospace systems.

The growing demand for safety and security has led to more research in building more efficient and intelligent automated surveillance systems. Therefore, a future challenge is to develop a wide-area distributed multi-sensor surveillance system which has robust, real-time computer algorithms able to perform with minimal manual reconfiguration on variable applications. Such systems should be adaptable enough to adjust automatically and cope with changes in the environment like lighting, scene geometry or scene activity. The system should be extensible enough, be based on standard hardware and exploit plug-and-play technology.

7 Acknowledgments

This work is part of the EPSRC-funded project COHERENT (computational heterogeneously timed networks) grant number is GR/R32895 (http://async.org.uk/coherent/). We would like to thank Mr David Fraser and Professor Tony Davies and the anonymous referees for their valuable observations.

8 References

1 First IEEE Workshop on Visual Surveillance, January 1998, Bombay, India
2 Second IEEE Workshop on Visual Surveillance, January 1999, Fort Collins, Colorado

3 Third IEEE International Workshop on Visual Surveillance (VS'2000), July 2000, Dublin, Ireland
4 First IEE Workshop on Intelligent Distributed Surveillance Systems, February 2003, London
5 Second IEE Workshop on Intelligent Distributed Surveillance Systems, February 2004, London
6 IEEE conference on Advanced Video and Signal Based Surveillance, July 2003
7 Special issue on visual surveillance, *Int. J. Comput. Vis.*, 2000
8 Special issue on visual surveillance, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000
9 Special issue on third generation surveillance systems, *Proc. IEEE*, 2001
10 Special issue on human motion analysis, *Comput. Vis. Image Underst.*, 2001
11 www.cieffe.com
12 Ping Lai Lo, B., Sun, J., and Velastin, S.A.: 'Fusing visual and audio information in a distributed intelligent surveillance system for public transport systems', *Acta Automatica Sinica*, 2003, **29**, (3), pp. 393–407
13 Nwagboso, C.: 'User focused surveillance systems integration for intelligent transport systems', in Regazzoni, C.S., Fabri, G., and Vernazza, G. (Eds.): 'Advanced Video-based Surveillance Systems' (Kluwer Academic Publishers, Boston, 1998), Chapter 1.1, pp. 8–12
14 www.sensis.com/docs/128
15 Weber, M.E., and Stone, M.L.: 'Low altitude wind shear detection using airport surveillance radars', *IEEE Aerosp. Electron. Syst. Mag.*, 1995, **10**, (6), pp. 3–9
16 Pozzobon, A., Sciutto, G., and Recagno, V.: 'Security in ports: the user requirements for surveillance system', in Regazzoni, C.S., Fabri, G., and Vernazza, G. (Eds.): 'Advanced Video-based Surveillance Systems' (Kluwer Academic Publishers, Boston, 1998)
17 Avis, P.: 'Surveillance and Canadian maritime domestic security', *Canad. Military J.*, 2003, pp. 9–15
18 Paulidis, I., and Morellas, V.: 'Two examples of indoor and outdoor surveillance systems', in Remagnino, P., Jones, G.A., Paragios, N., and Regazzoni, C.S. (Eds.): 'Video-based Surveillance Systems' (Kluwer Academic Publishers, Boston, 2002), pp. 39–51
19 ADVISOR specification documents (internal classification 2001)
20 http://dilnxsvr.king.ac.uk/cromatica/
21 Ronetti, N., and Dambra, C.: 'Railway station surveillance: the Italian case', in Foresti, G.L., Mahonen, P., and Regazzoni, C.S. (Eds.): 'Multimedia Video Based Surveillance Systems' (Kluwer Academic Publishers, Boston, 2000), pp. 13–20
22 Pellegrini, M., and Tonani, P.: 'Highway traffic monitoring', in Regazzoni, C.S., Fabri, G., and Vernazza, G. (Eds.): 'Advanced Video-based Surveillance Systems' (Kluwer Academic Publishers, Boston, 1998)
23 Beymer, D., McLauchlan, P., Coifman, B., and Malik, J.: ' A real-time computer vision system for measuring traffic parameters'. Proc. 1997 Conf. on Computer Vision and Pattern Recognition, IEEE Computer Society, pp. 495–502
24 Zhi-Hong, Z.: 'Lane detection and car tracking on the highway', *Acta Automatica Sinica*, 2003, **29**, (3), pp. 450–456
25 Jian-Guang, L., Qi-Feing, L., Tie-Niu, T., and Wei-Ming, H.: '3-D model based visual traffic surveillance', *Acta Automatica Sinica*, 2003, **29**, (3), pp. 434–449
26 Ferryman, J.M., Maybank, S.J., and Worrall, A.D.: 'Visual surveillance for moving vehicles', *Int. J. Comput. Vis.*, 2000, **37**, (2), Kluwer Academic Publishers, Netherlands, pp. 187–197
27 http://www.detec.no
28 http://www.gotchanow.com
29 secure30.softcomca.com/fge_biz
30 Brodsky, T., Cohen, R., Cohen-Solal, E., Gutta, S., Lyons, D., Philomin, V., and Trajkovic, M.: 'Visual surveillance in retail stores and in the home', *in*: 'Advanced Video-based Surveillance Systems' (Kluwer Academic Publishers, Boston, 2001), Chapter 4, pp. 50–61
31 Cucchiara, R., Grana, C., Patri, A., Tardini, G., and Vezzani, R.: 'Using computer vision techniques for dangerous situation detection in domotic applications'. Proc. IEE Workshop on Intelligent Distributed Surveillance Systems, London, 2004, pp. 1–5
32 Greenhill, D., Remagnino, P., and Jones, G.A.: 'VIGILANT: content-querying of video surveillance streams', in Remagnino, P., Jones, G.A., Paragios, N., and Regazzoni, C.S. (Eds.): 'Video-based Surveillance Systems' (Kluwer Academic Publishers, Boston, USA, 2002), pp. 193–205
33 Micheloni, C., Foresti, G.L., and Snidaro, L.: 'A co-operative multi-camera system for video-surveillance of parking lots'. Intelligent Distributed Surveillance Systems Symp. by the IEE, London, 2003, pp. 21–24
34 Boult, T.E., Micheals, R.J., Gao, X., and Eckmann, M.: 'Into the woods: visual surveillance of non-cooperative and camouflaged targets in complex outdoor settings', *Proc. IEEE*, 2001, **89**, (1), pp. 1382–1401
35 Xu, M., Lowey, L., and Orwell, J.: 'Architecture and algorithms for tracking football players with multiple cameras'. Proc. IEE Workshop on Intelligent Distributed Surveillance Systems, London, 2004, pp. 51–56
36 Krumm, J., Harris, S., Meyers, B., Brumit, B., Hale, M., and Shafer, S.: 'Multi-camera multi-person tracking for easy living'. Third IEEE Int. Workshop on Visual Surveillance, Ireland, 2000, pp. 8–11
37 Heikkila, J., and Silven, O.: 'A real-time system for monitoring of cyclists and pedestrians'. 2nd IEEE Int. Workshop on Visual Surveillance, Colorado, 1999, pp. 74–81
38 Haritaoglu, I., Harwood, D., and Davis, L.S.: 'W^4:real-time surveillance of people and their activities', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (8), pp. 809–830
39 Geradts, Z., and Bijhold, J.: 'Forensic video investigation', in Foresti, G.L., Mahonen, P., and Regazzoni, C.S. (Eds.): 'Multimedia video based surveillance systems' (Kluwer Academic Publishers, Boston, 2000), pp. 3–12
40 Hai Bui, H., Venkatesh, S., and West, G.A.W.: 'Tracking and surveillance in wide-area spatial environments using the abstract hidden markov model', *Int. J Pattern Recognit. Anal. Intell.*, 2001, **15**, (1), pp. 177–195
41 Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt P., and Wixson L.: 'A system for video surveillance and monitoring'. Robotics Institute, Carnegie Mellon University, 2000, pp. 1–68
42 www.objectvideo.com
43 www.nice.com
44 www.pi-vision.com
45 www.ipsotek.com
46 www.neurodynamics.com
47 Velastin, S.A.: 'Getting the best use out of CCTV in the railways'. Rail Safety and Standards Board, July 2003, pp. 1–17
48 Haritaoglu, I., Harwood, D., and Davis, L.S.: 'Hydra: multiple people detection and tracking using silhouettes'. Proc. IEEE Int. Workshop Visual Surveillance, 1999, pp. 6–14
49 Batista, J., Peixoto, P., and Araujo, H.: 'Real-time active visual surveillance by integrating'. Workshop on Visual Surveillance, India, 1998, pp.18–26
50 Ivanov, Y.A., Bobick, A.F., and Liu, J.: 'Fast lighting independent background', *Int. J. Comput. Vis.*, 2000, **37**, (2), pp. 199–207
51 Pless, R., Brodsky, T., and Aloimonos, Y.: 'Detecting independent motion: the statics of temporal continuity', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, pp. 768–773
52 Liu, L.C., Chien, J.-C., Chuang, H.Y-H., and Li, C.C.: 'A frame-level FSBM motion estimation architecture with large search range'. IEEE Conf. on Advanced Video and Signal Based Surveillance, Florida, 2003, pp. 327–334
53 Remagnino, P., Baumberg, A., Grove, T., Hogg, D., Tan, T., Worral, A., and Baker, K.: 'An integrated traffic and pedestrian model-based vision system'. BMVC97 Proc., Israel, pp. 380–389
54 Ng, K.C., Ishiguro, H., Trivedi, M., and Sogo, T.: 'Monitoring dynamically changing environments by ubiquitous vision system'. 2nd IEEE Workshop on Visual Surveillance, Colorado, 1999, pp. 67–74
55 Orwell, J., Remagnino, P., and Jones, G.A.: 'Multicamera color tracking'. 2nd IEEE Workshop on Visual Surveillance, Colorado, 1999, pp. 14–22
56 Darrell, T., Gordon, G., Woodfill, J., Baker, H., and Harville, M.: 'Robust real-time people tracking in open environments using integrated stereo, color, and face detection'. 3rd IEEE workshop on visual surveillance, India, 1998, pp. 26–33
57 Rota, N., and Thonnat, M.: 'Video sequence interpretation for visual surveillance'. 3rd IEEE Int. Workshop on Visual Surveillance, Dublin, 2000, pp. 59–68
58 Owens, J., and Hunter, A.: 'Application of the self-organising map to trajectory classification'. 3rd IEEE Int. Workshop on Visual Surveillance, Dublin, 2000, pp. 77–85
59 Stauffer, C., Eric, W., and Grimson, L.: 'Learning patterns of activity using real-time tracking', *IEEE Trans. Pattern Anal. and Mach. Intell.*, 2000, **22**, (8), pp. 747–757
60 Stringa, E., and Regazzoni, C.S.: 'Content-based retrieval and real-time detection from video sequences acquired by surveillance systems'. Int. Conf. on Image Processing, Chicago, 1998, pp. 138–142
61 Decleir, C., Hacid, M.-S., and Koulourndijan, J.: 'A database approach for modelling and querying video data'. Proc. 15th Int. Conf. on Data Engineering, Australia, 1999, pp. 1–22
62 MaKris, D., Ellis, T., and Black, J.: 'Bridging the gaps between cameras'. Int. Conf. Multimedia and Expo, Taiwan, June 2004
63 Black, J., Ellis, T., and Rosin, P.: 'A novel method for video tracking performance evaluation'. The Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, October, France, 2003, pp. 125–132
64 Gavrila, D.M.: 'The analysis of human motion and its application for visual surveillance', *Comput. Vis. Image Underst.*, 1999, **73**, (1), pp. 82–98
65 Norhashimah, P., Fang H., and Jiang, J.: 'Video extraction in compressed domain'. IEEE Conf. on Advanced Video and Signal Based Surveillance, Florida, 2003, pp. 321–327
66 Soldatini, F., Mähönen, P., Saaranen, M., and Regazzoni, C.S.: 'Network management within an architecture for distributed hierarchical digital surveillance systems', in Foresti, G.L., Mahonen, P., and Regazzoni, C.S. (Eds.): 'Multimedia video based surveillance systems' (Kluwer Academic Publishers, Boston, 2000), pp. 143–157
67 Liu, L.-C., Chien, J.-C., Chuang, H. Y-H., and Li, C.C.: 'A frame-level FSBM motion estimation architecture with large search range'. IEEE Conf. on Advanced Video and Signal based Surveillance, Florida, 2003, pp. 327–334
68 Saad, A., and Smith, D.: 'An IEEE 1394-firewire-based embedded video system for surveillance applications'. IEEE Conf. on Advanced Video and Signal based Surveillance, Florida, 2003, pp. 213–219
69 Ye, H., Walsh, G.C., and Bushnell, L.G.: 'Real-time mixed-traffic wireless networks', *IEEE Trans. on Ind. Electron.*, 2001, **48**, (5), pp. 883–890

70 Huang, J., Krasic, C., Walpole, J., and Feng, W.: 'Adaptive live video streaming by priority drop'. IEEE Conf. on Advanced Video and Signal Based Surveillance, Florida, 2003, pp. 342–348
71 Marcenaro, L., Oberti, F., Foresti, G.L., and Regazzoni, C.S.: 'Distributed architectures and logical-task decomposition in Multimedia surveillance systems', *Proc. IEEE*, 2001, **89**, (10), pp. 1419–1438
72 Marchesotti, L., Messina, A., Marcenaro, L., and Regazzoni, C.S.: 'A cooperative multisensor system for face detection in video surveillance applications', *Acta Automatica Sinica*, 2003, **29**, (3), pp. 423–433
73 Nguyen, N.T., Venkatesh, S., West, G., and Bui, H.H.: 'Multiple camera coordination in a surveillance system', *Acta Automatica Sinica*, 2003, **29**, (3), pp. 408–421
74 Jaynes, C.: 'Multi-view calibration from planar motion for video surveillance', 2nd IEEE Int. Workshop on Visual Surveillance, Colorado, 1999, pp. 59–67
75 Snidaro, L., Niu, R., Varshney, P.K., and Foresti, G.L.: 'Automatic camera selection and fusion for outdoor surveillance under changing weather conditions'. IEEE Conf. on Advanced Video and Signal based Surveillance, Florida, 2003, pp. 364–370
76 Collins, R.T., Lipton, A.J., Fujiyoshi, H., and Kanade, T.: 'Algorithms for cooperative multisensor surveillance', *Proc. IEEE*, **89**, (10), 2001, pp. 1456–1475
77 Wren, C., Azarbayejani, A., Darrell, T., and Pentland, A.: 'Pfinder:real-time tracking of the human body', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 780–785
78 Pavlidis, I., Morellas, V., Tsiamyrtzis, P., and Harp, S.: 'Urban surveillance systems: from the laboratory to the commercial world', *Proc. IEEE*, 2001, **89**, (10), pp. 1478–1495
79 Bennewitz, M., Burgard, W., and Thrun, S.: 'Using EM to learn motion behaviours of persons with mobile robots'. Proc. Conf. on Intelligent Robots and Systems (IROS), Switzerland, 2002
80 Oren, M., Papageorgiou, C., Sinham P., Osuna, E., and Poggio, T.: 'Pedestrian detection using wavelet templates'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Puerto Rico, 1997, pp. 193–199
81 Hemayed, E.E.: 'A survey of self-camera calibration'. Proc. of the IEEE Conf. on Advanced Video and Signal based Surveillance, Florida, 2003, pp. 351–358
82 Arulampalam, S., Maskell, S., Gordon, N., and Clapp, T.: 'A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking', *IEEE Trans. on Signal Process.*, 2002, **50**, (2), pp. 174–188
83 Rath, T.M., and Manmatha, R.: 'Features for word spotting in historical manuscripts'. Proc. of the 7th Int. Conf. on Document Analysis and Recognition, 2003, pp. 512–527
84 Oates, T., Schmill, M.D., and Cohen, P.R.: 'A method for clustering the experiences of a mobile robot with human judgements'. Proc. of the 17th National Conf. on Artificial Intelligence and Twelfth Conf. on Innovative Applications of Artificial Intelligence, AAAI Press, 2000, pp. 846–851
85 Nguyen, N.T., Bui, H.H., Venkatesh, S., and West, G.: 'Recognising and monitoring high-level behaviour in complex spatial environments'. IEEE Int. Conf. on Computer Vision and Pattern Recognition, Wisconsin, 2003, pp. 1–6
86 Ivanov, Y., and Bobick, A.: 'Recognition of visual activities and interaction by stochastic parsing', *IEEE Trans. Pattern Recognit. Mach. Intell.*, 2000, **22**, (8), pp. 852–872
87 Regazzoni, C.S., Ramesh, V., and Foresti, G.L.: 'Special issue on video communications, processing, and understanding for third generation surveillance systems', *Proc. IEEE*, 2001, **89**, (10), pp. 1355–1365
88 Gong, S., and Xiang, T.: 'Recognition of group activities using dynamic probabilistic networks'. 9th IEEE Int. Conf. on Computer Vision, France, 2003, Vol. 2, pp. 742–750
89 Buxton, H.: 'Generative models for learning and understanding scene activity'. Proc. 1st Int. Workshop on Generative Model-Based Vision, Copenhagen, 2002, pp. 71–81
90 Ivanov, Y., Stauffer, C., Bobick, A., and Grimson, W.E.L.: 'Video surveillance of interactions'. 2nd IEEE Int. Workshop on Visual Surveillance, Colorado, 1999, pp. 82–91
91 Christensen, M., and Alblas, R.: 'V^2- design issues in distributed video surveillance systems', Demark, 2000, pp. 1–86
92 Yuan, X., Sun, Z., Varol, Y., and Bebis, G.: 'A distributed visual surveillance system'. IEEE Conf. on Advanced Video and Signal based Surveillance, Florida, 2003, pp. 199–205
93 Garcia, L.M., and Grupen, R.A.: 'Towards a real-time framework for visual monitoring tasks'. 3rd IEEE Int. Workshop on Visual Surveillance, Ireland, 2000, pp. 47–56
94 Wu, C.-H., Irwin, J.D., and Dai, F.F.: 'Enabling multimedia applications for factory automation', *IEEE Trans. on Ind. Electron.*, 2001, **48**, (5), pp. 913–919
95 Almeida, L., Pedreiras, P., Alberto, J., and Fonseca, G.: 'The FFT-CAN protocol: why and how', *IEEE Trans. Ind. Electron.*, 2002, **49**, (6), pp. 1189–1201
96 Conti, M., Donatiello, L., and Furini, M.: 'Design and analysis of RT-ring: a protocol for supporting real-time communications', *IEEE Trans. on Ind. Electron.*, 2002, **49**, (6), pp. 1214–1226
97 Jackson, L.E., and Rouskas, G.N.: 'Deterministic preemptive scheduling of real-time tasks', *Computer, IEEE*, 2002, **35**, (5), pp. 72–79
98 Rybski, P.E., Stoeter, S.A., Gini, M., Hougen, D.F., and Papanikolopoulos, N.P.: 'Performance of a distributed robotic system using shared communications channels', *IEEE Trans. Robot. Autom.*, 2002, **18**, (5), pp. 713–727
99 Valera, M., and Velastin, S.A.: 'An approach for designing a real-time intelligent distributed surveillance system'. Proc. of the IEE Workshop on Intelligent Distributed Surveillance Systems, London, 2003, pp. 42–48
100 Greiffenhagen, M., Comaniciu, D., Niemann, H., and Ramesh, V.: 'Design, analysis, and engineering of video monitoring systems: an approach and a case study', *Proc. IEEE*, 2001, **89**, (10), pp. 1498–1517
101 Matijasevic, M., Gracanin, D., Valavanis, K.P., and Lovrek, I.: 'A framework for multiuser distributed virtual environments', *IEEE Trans. Syst. Man Cybern.*, 2002, **32**, (4), pp. 416–429
102 Castelpietra, P., Song, Y-Q., Lion, F.S., and Attia, M.: 'Analysis and simulation methods for performance evaluation of a multiple networked embedded architecture', *IEEE Trans. Ind. Electron.*, 2002, **49**, (6), pp. 1251–1264