第 29 卷 第 3 期
2003 年 5 月
自 动 化 学 报
ACTA AUTOMATICA SINICA
Vol. 29, No. 3
May, 2003

Fusing Visual and Audio Information in a Distributed Intelligent Surveillance System for Public Transport Systems[1)]

Benny Ping Lai Lo Jie Sun Sergio A. Velastin
(Digital Imaging Research Center, School of Computing & Information Systems, Kingston University, Kingston upon Thames, U. K.)
(E-mail: {b. lo, jie. sun, sergio. velastin}@ kingston. ac. uk)

Abstract Extensive research has been conducted in applying video and audio processing algorithms for improving passenger safety and security in public transport systems. However, due to the complex and intense computations involved in the algorithms, most studies focus only on one aspect of the safety or security issues. In addition, as passengers' behaviours and environments are fairly variable and unpredictable, the robustness of some algorithms is still in question and few of the reported results can be applied in all the different scenarios encountered in transport systems. To develop a complete and practical intelligent surveillance system, the EU project, PRISMATICA, is designed to integrate different intelligent detection devices, which includes local camera networks, crowd monitoring devices, intelligent cameras, contactless smart cards, wireless video/audio transmission, and audio surveillance systems, to monitor different safety and security concerns in railways. As different algorithms and techniques are applied in different devices, to fulfil the real-time requirement, the PRISMATICA system is designed as a distributed system where each device is a standalone process, and devices are linked and synchronized using a CORBA network. Although the resulting system is capable of monitoring and detecting different events, certain detected events could represent the same incident. In addition, the system could potentially generate too much information for operators to identify and react to incidents straight away. This paper presents a novel concept of fusing different evidences from a distributed visual and audio processing system to improve the robustness of incident detection and to provide more descriptive events. On-site testing is being carried out in London and Paris to validate the performance of this system.

Key words Video/audio algorithms, intelligent transport system, intelligent camera

1 Introduction

Surveillance cameras have been widely used in public transport systems to improve passenger safety and security. Particularly, in railways, thousands of cameras are installed everywhere in trains and stations. Although CCTV (Closed Circuit Television) extends the monitoring coverage to every corner of the stations and train compartments, incidents are often not seen on-line and recordings have to be used to determine their cause. Shortage of staff is one of the major reasons to the overlooking of incidents, since operators usually have to monitor more than 15 cameras simultaneously, and it is a very laborious task. For example, according to the annual safety performance report published by the Railway Safety organisation in the UK, there were 16 fatalities in year 2001/02 due to trespassing[2)]. In another study conducted by Transport Canada, it was found that 57 trespassers were killed and 40 trespassers were injured in year 1996 in Canada[3)]. If trespassing could be spotted by operators as it happens, the number of such fatalities could be reduced. The annual safety performance report also suggested that 2. 3 fatalities could be prevented per

1) Supported as part by the EU project PRISMATICA(GRD1-2000-10601)
2) Annual Safety Performance Report 2001/02, http://www. railwaysafety. org. uk/pdf/railrepo0102/passngr. pdf
3) Railway Safety Facts 1996 Trespasser Occurrences, http://www. tc. gc. ca/railway/PDF/trespass. pdf
Received November 25, 2002; in revised form April 7, 2003
收稿日期 2002-11-25; 收修改稿日期 2003-04-07

year if the operator can spot the situation from the CCTV and use the PA (public addressing) system to address the situation.

Numerous studies have presented algorithms and systems to automate or semi-automate the surveillance process and assist the operators in detecting incidents. One of the major surveillance tasks for a station operator is to identify abandoned objects, as a small abandoned object could hide a highly destructive bomb placed by a terrorist. Boghossian *et al.* proposed an algorithm in detecting stationary objects or abandoned baggage in railway stations based on motion vectors extracted from the CCTV video images[1]. Similarly, Sacchi *et al.* proposed a distributed system using an image segmentation technique with direct sequence code-division multiple access (DS/CDMA) technology to detect abandoned objects in unmanned railway stations[2].

Regarding the prevention of vandalism and violent offences in railway premises, Fuentes *et al.* proposed a people tracking system for railway stations based on direct and inverse matching matrices[3]. On the other hand, Siebel *et al.* recently proposed a fusion approach to track passenger movements in railway stations[4]. One level above tracking, Rota *et al.* introduces a behaviour analysis system that interprets different scenarios in metro stations based on tracking, scene modelling and contextual information[5].

Apart from incidents involving individuals, overcrowding is one of the major concerns in railway stations. Bouchafa *et al.* developed a system for monitoring the crowd in subway corridors[6]. Boghossian *et al.* introduced a real-time crowd detection system based on motion information[7]. In addition, in a previous study, we proposed a system for measuring crowd levels on station platforms[8].

Due to the complex environment in railway stations and the difficulties involved in studying human behaviour, complicated and intense computations are required in most algorithms, and most studies have only focused on certain aspects of the safety and security issues. In the EU-funded project PRISMATICA ("Pro-active Integrated Systems for Security Management by Technological, Institutional and Communication Assistance"), distributed processing is chosen as the architecture for developing a complete and practical intelligent surveillance system for public transport systems[1)]. Unlike the distributed network of active video sensors proposed by Collins *et al.* , which focus on tracking objects over an extended area by cooperating tracking results from multiple sensors, the PRISMATICA system aims to provide an integrated system which does not only has the tracking ability using multiple cameras, but it also has the function to monitor other safety and security aspect, such as the activation of a mobile panic button alarm and the detection of suspect packages[9].

The PRISMATICA system is comparable to the third generation surveillance system (3GSS) introduced by Marcenaro *et al.* ; however, instead of spreading the processing tasks in a distributed network, the tasks are performed locally in each computer or process within a distributed network, and only the high level information is communicated among processes[10]. In terms of reliability and maintainability, the PRISMATICA system is expected to perform better than the 3GSS, as each process is independent and any failure in one process will only affect a particular sensor, but the system as a whole will still remain in operation. In addition, failures could be identified and located much faster than the 3GSS. Furthermore, as only high-level information is passed between processes, the network load of the proposed system is be much less than the 3GSS, a factor that contributes to achieving the real-time processing requirements.

Although only the high-level information, such as events and alarms, are shown to

1)For more details on the EU Project PRISMATICA please refer to http://www.prismatica.com the user, the PRISMATICA system could potentially generate too much information for operators to identify and react to incidents immediately, as it comprises a considerable number of different sensors. In addition, certain events or alarms detected from different sensors could correspond to the same incident. Therefore, a novel concept of fusing different evidences from the distributed system is presented in this paper for improving the robustness of the incident detection and providing descriptive messages to operators.

2 PRISMATICA system overview

Due to the limitation on the processing power of computers and the limited network bandwidth, traditionally, different detection systems operate independently with their own user interface and control mechanisms. This creates many difficulties for operators, as they have to get familiar with different systems and different operating environment, which leads to expensive operating cost and reluctance to upgrade new safety-related systems. As one of the objectives of the PRISMATICA project is to provide an operationally cost-effective intelligent surveillance system, the PRISMATICA system is designed as a distributed system built upon a CORBA (Common Object Request Broker Architecture) network and comprises different intelligent devices. Each device is a standalone process or system that monitors different safety and security aspect in a railway system, and an operator can monitor and control all the connected devices through a operating console. Fig. 1 shows the organisation diagram of the trial PRISMATICA system.

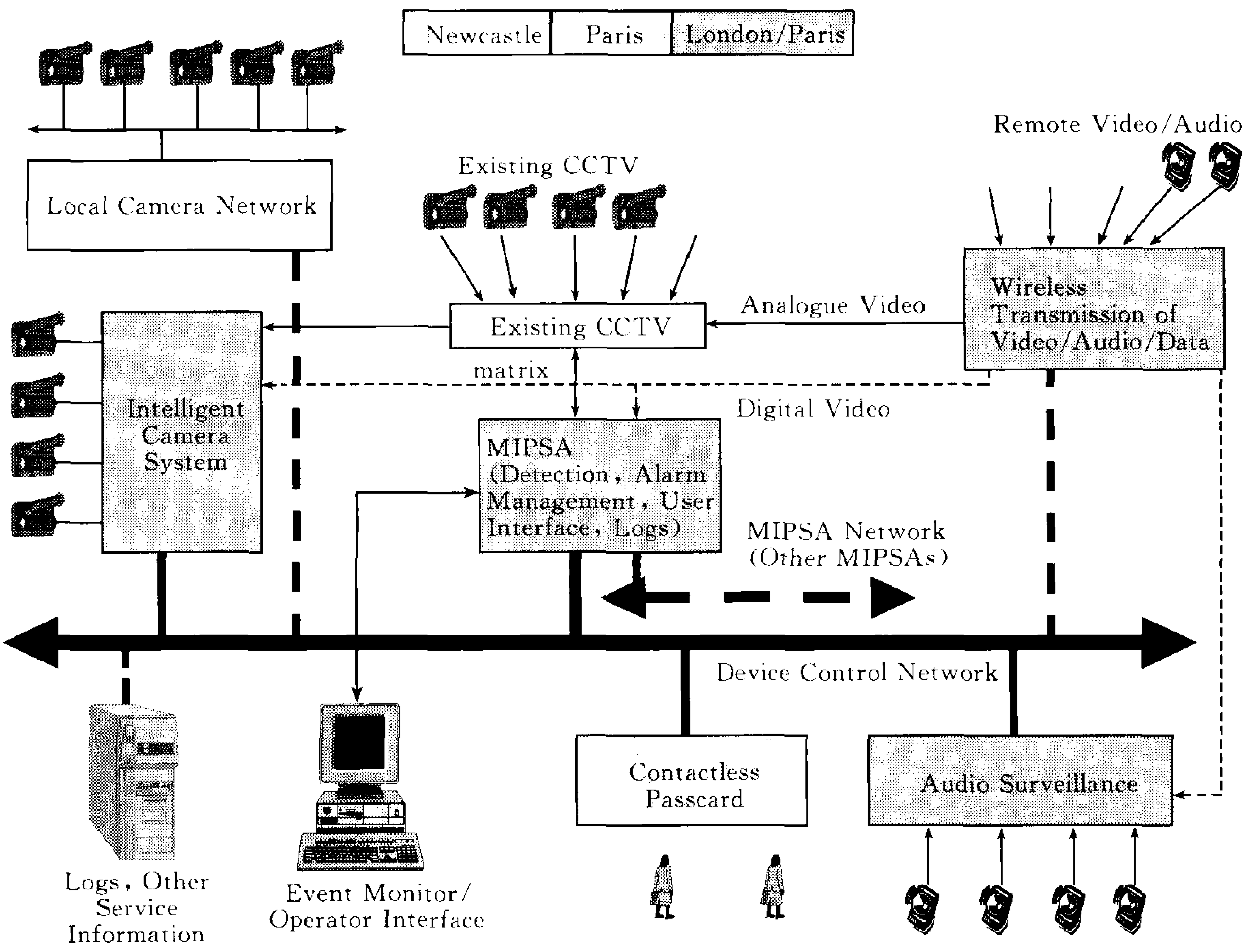

Fig. 1 Overall organisation of the PRISMATICA trial system

As indicated in the diagram, the PRISMATICA system consists of 6 major components:

1) Local Camera Network;

2) Wireless data/video/audio Transmission System;

3) Intelligent Camera System;

4) Audio Surveillance System;
5) Contactless Passcard System;
6) MIPSA.

2.1 Local camera network

The Local Camera Network is a multi-camera industrial surveillance system, developed by ILA (Intelligent Laser Applications GmbH) and INRETS (Institut National de Recherche sur les Transports et leur Sécurité). The system is designed for automating certain monitoring tasks in a transport network, which includes tracking, intrusion detection in forbidden areas, passengers counting at entries or exits, and occupation rate in strategic areas[11]. Due to extensive computation required in those tasks, the Local Camera Network is a distributed processing system, where low-level information processing is carried out locally in each camera unit. High-level information resulting from the camera unit is then sent across a local network to a local substation or central processing computer in order to reduce the network load. A Local Camera Network with eight cameras has been installed in Newcastle Airport (UK) for validating its functions.

2.2 Wireless data/video/audio transmission system

Wireless transmissions in a railway environment is a very difficult problem, since moving trains induce substantial amounts of noise to radio signals, and railway companies often have a strict electromagnetic compatibility (EMC) requirement for electronic equipment, especially radio transmission equipment. In the PRISMATICA project, a wireless transmission system is developed by the CEA(Commissariat á l'Energie Atomique) aiming for temporary video monitoring of sensitive areas (such as delinquency, works, special events, etc). The system is based on direct sequence spread spectrum (DSSS) asynchronous CDMA (A-CDMA) transmissions, which allows for up to 5 radio transmissions sharing the same frequency band (ISM 2.44GHz band)[11]. As real-time video transmission requires high bandwidth, wavelet compression is used to reduce the amount of data. To ensure the transmission quality of the compressed data, the system combines the advantage of DSSS in multi-path propagation with the multiple access capacity of A-CDMA systems.

2.3 Intelligent camera system

Similar to the Local Camera Network, the Intelligent Camera System, developed by INRETS, is a multi-camera video surveillance system; however, instead of distributing the processing in a network, the system is designed to be a single unit which can process multiple cameras. With the size of a small box, the Intelligent Camera System is capable of handling up to four video inputs simultaneously and detect the following events: standing, intrusion, queue lengths, counter-flow pedestrian traffic and people distribution[12]. Whenever a situation is detected, the respective information is sent to the central system via the CORBA network so the operators can be alerted, localise the problem and take appropriate action.

2.4 Audio surveillance system

In addition to the video processing systems, an audio surveillance system, developed by Thales Underwater Systems (France), is incorporated in the PRISMATICA system. The aim of the acoustic surveillance system is to enable the detection of distress calls ("help", "stop thief", etc) or shouts, during security problems, such as assaults and thefts[12]. As per the intelligent camera system, the detection results are sent to the central system for alerting the operators. Together with the video information, the audio information could further assist the operators in assessing situations.

2.5 Contactless passcard system

Most public transport systems have developed the concept of "help-points" whereby a passenger that needs assistance or information presses a button and is then connected to staff (usually in a control centre)[12]. As the "help-point" is fixed, the passenger can be located and assistance sent when necessary. At the same time, many public transport companies are developing the concept of a contactless "smart card" used as a ticket/pass and other value-added services. The Contactless Passcard System, developed by RATP (Régie Autonome des Transports Parisien), in the PRISMATICA project aims to utilise and enhance the smart card technology to provide a personal mobile "help-point" facility to improve the levels of actual and perceived security. As such, each smart card is equipped with a panic button, and in case of an emergency, passengers can press the panic button to alert the staff. Since the system is based on radio transmission, radio receiving stations are set up throughout the station, and according to the overlapping area of the received stations, the location of the signalled passcard will be identified.

2.6 MIPSA

MIPSA (Modular Integrated Pedestrian Surveillance Architecture), developed by the authors, is a technical concept that uses state of the art technology to support human operators in their task to prevent and detect security-threatening situations. Besides being the display and control front end of the PRISMATICA system, the MIPSA also has a small video matrix, a database system, an event recording system, a video acquisition and display system, and a video processing system[12]. The user interface in the MIPSA provides a graphical presentation of the station layout together with the sensors marked at the installed location in order to assist operators in monitoring and controlling the sensors and actuators, as shown in Fig. 2.

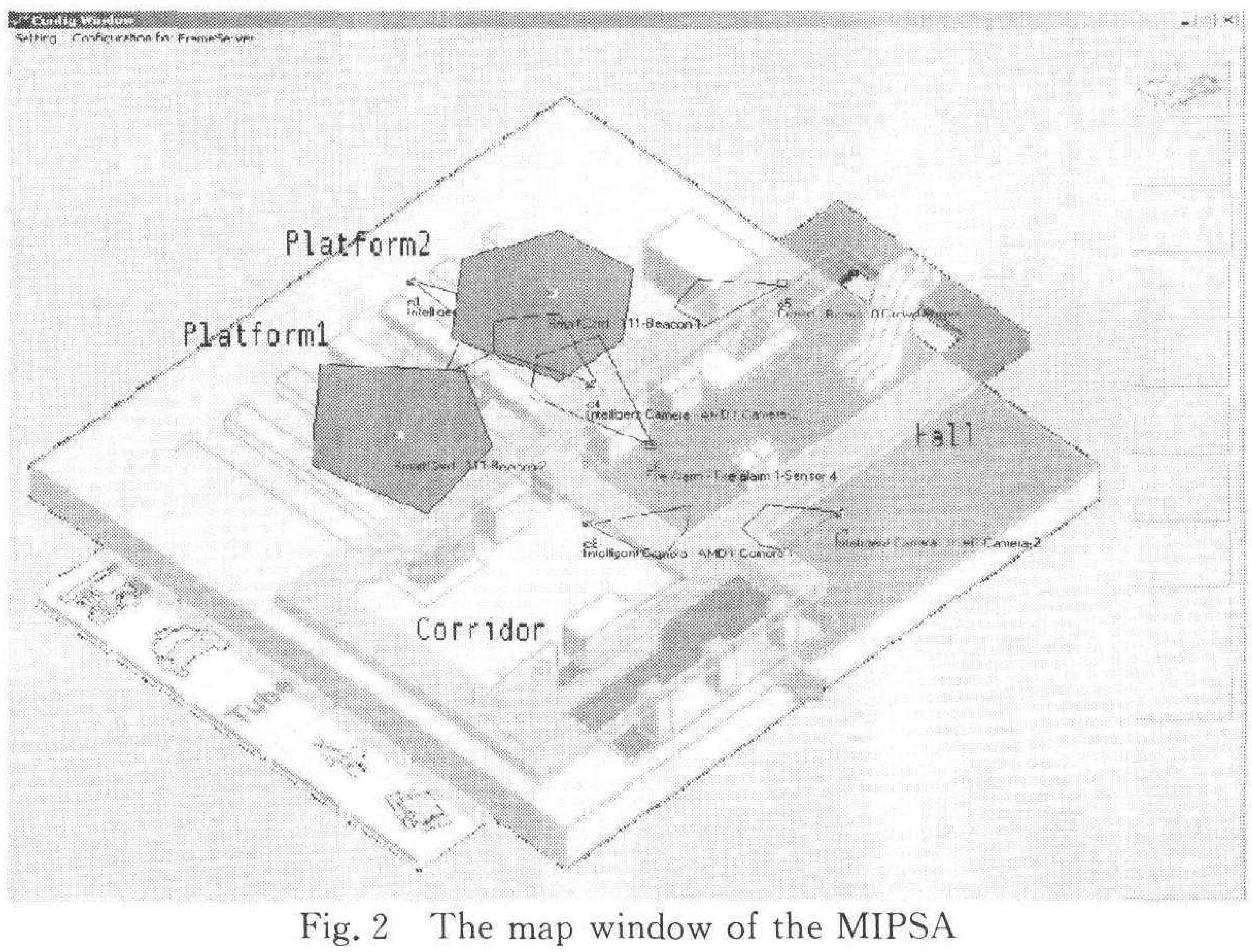

Fig. 2 The map window of the MIPSA

On the other hand, as a frame server (the video matrix) is included in the MIPSA, the MIPSA can capture and display 16 cameras simultaneously using a multiplexed signal. Live video are displayed on the MIPSA for operators to monitor the station, and in case of events a zooming function is provided for the operator to focus on the alarmed area. Fig. 3 illustrates the video display of the MIPSA system.

As shown in Fig. 3, event messages, sent by devices, are logged and listed in the event window below the video displays, and all the logged messages are stored in the database for later retrieval. In addition to logging the event messages, the MIPSA is designed

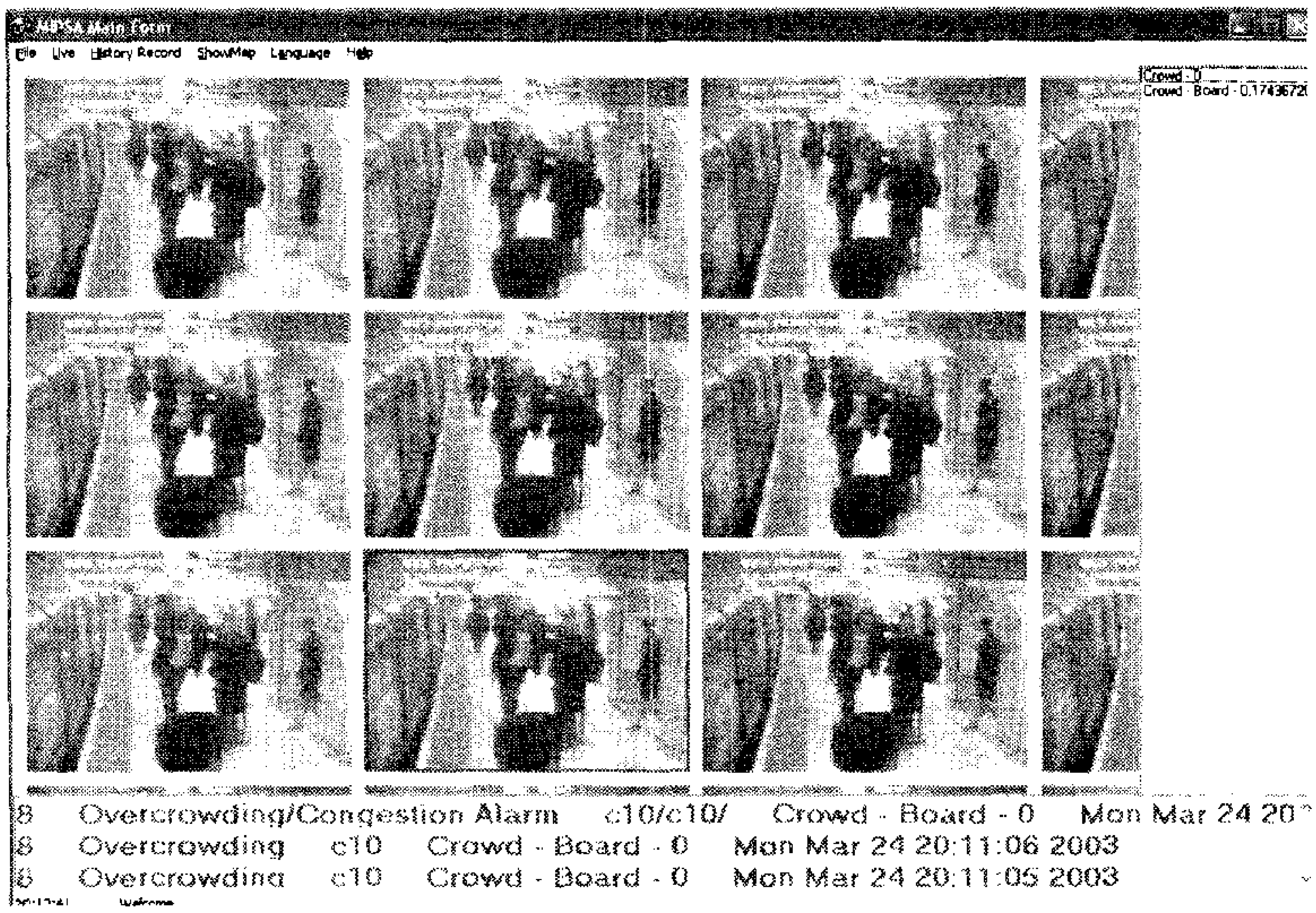

Fig. 3 Video display window of the MIPSA

to be able to record the concerned video images in the database, and play back the recorded video together with the associated event messages, in order to improve evidence gathering processes. Furthermore, a self-contained MIPSA system includes up to 5 video processors for detecting events, such as overcrowding, congestion, stationary objects and motion directions of moving objects. Sample processing outputs are shown in Fig. 4 and Fig. 5.

Fig. 4 (Left) Overcrowding and congestion detection result image where the overcrowded area is highlighted in black rectangle and the congested area in grey. (Right) Stationary object detection result image and the stationary pedestrian is highlighted with a black rectangle

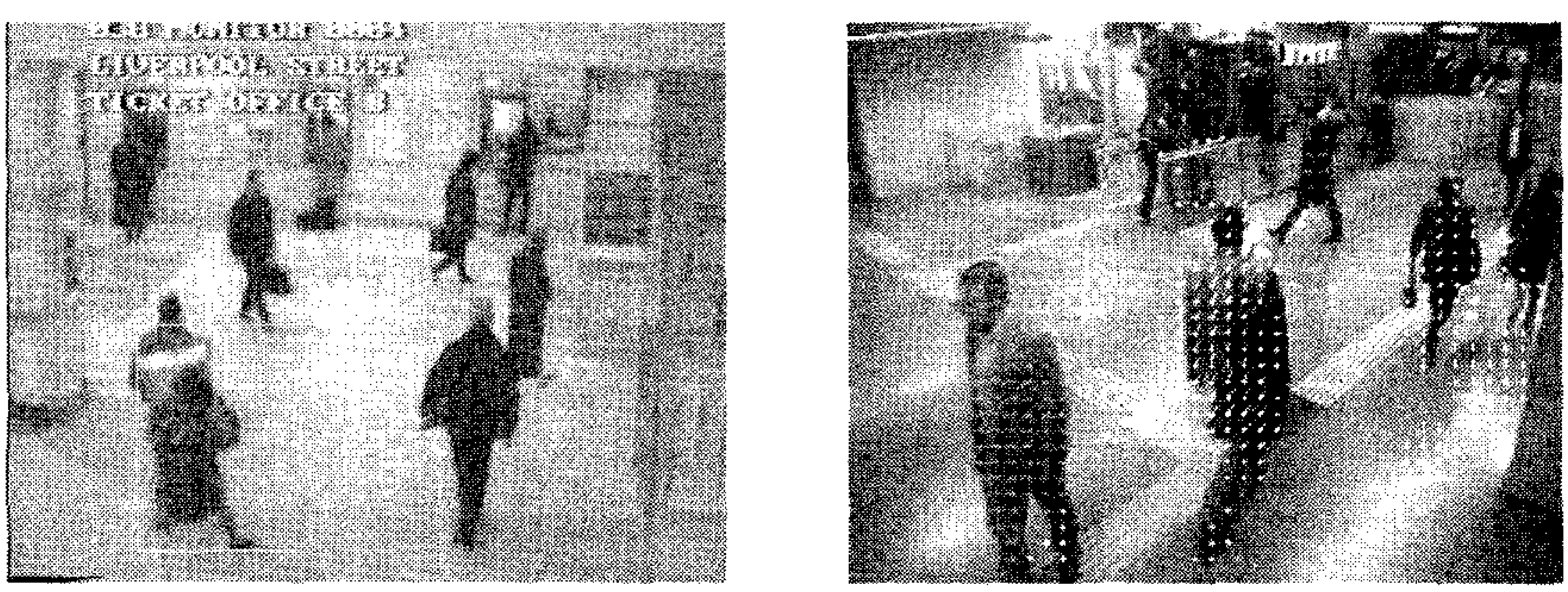

Fig. 5 (Left) Unusual direction of movement indicated with the grey rectangle. (Right) Motion detection result where grey colour vectors show different movement directions

2.7 Communication

CORBA has been chosen as the communication backbone of the PRISMATICA system where each device is represented as a CORBA object in the network. Among different CORBA software, ACE/TAO has been chosen as the CORBA implementation. With the real-time optimisation of the ACE/TAO, the object-oriented nature of CORBA eases the integration of such large scale and diversified system[13]. As ACE/TAO CORBA is in continuous development, the protocols have been wrapped in a Windows™ DLL (Dynamic Linked Library), called MIPSATAO DLL, which uses a conventional C language interface. All the CORBA interface functions have been encapsulated in the DLL. Therefore, by using the DLL, developers can build a device in the PRISMATICA system without needing familiarity with CORBA. On the other hand, since the design is based on CORBA, those developers who are familiar with CORBA can develop their devices independently using any other platforms and the devices can still be integrated into the PRISMATICA system. Fig. 6 illustrates the connection diagram of the PRISMATICA system where the MCO (MIPSA Communication Object) and DCO (Device Communication Object) are the CORBA communication objects created by the DLL instance for identifying devices and links to its functions. In addition to the connection between MIPSA and the devices, separate links (device-to-device links) can be set up between two devices, such that device specific information can be passed between the devices.

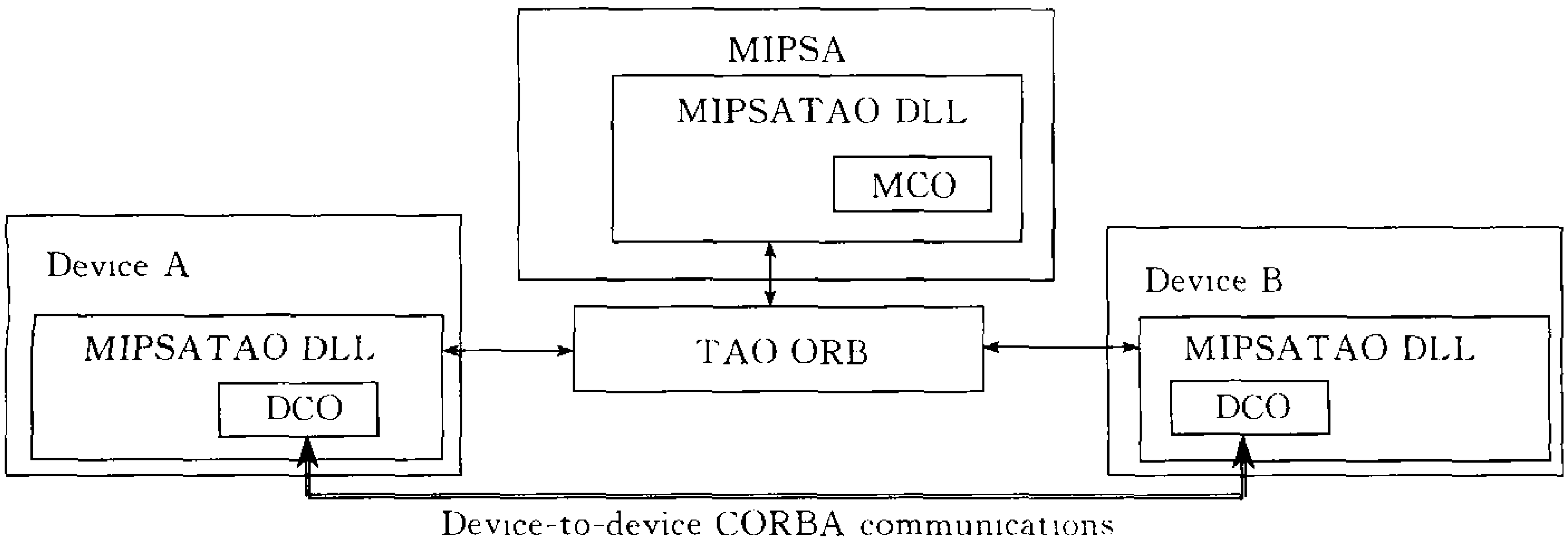

Fig. 6 Connection diagram of the PRISMATICA system

2.8 Communication protocol

Although CORBA provides an object-oriented interface of each device, any changes made in one object will require updates on all other interfacing objects. Besides, the entire system may need to be rebuilt to cope with any new devices. In order to ease the integration of new devices, modification of existing devices, and maintain the CORBA object-oriented features, a generic protocol is defined for the PRISMATICA system. Fig. 7 illustrates the UML diagram of the generic device model used to design the protocol.

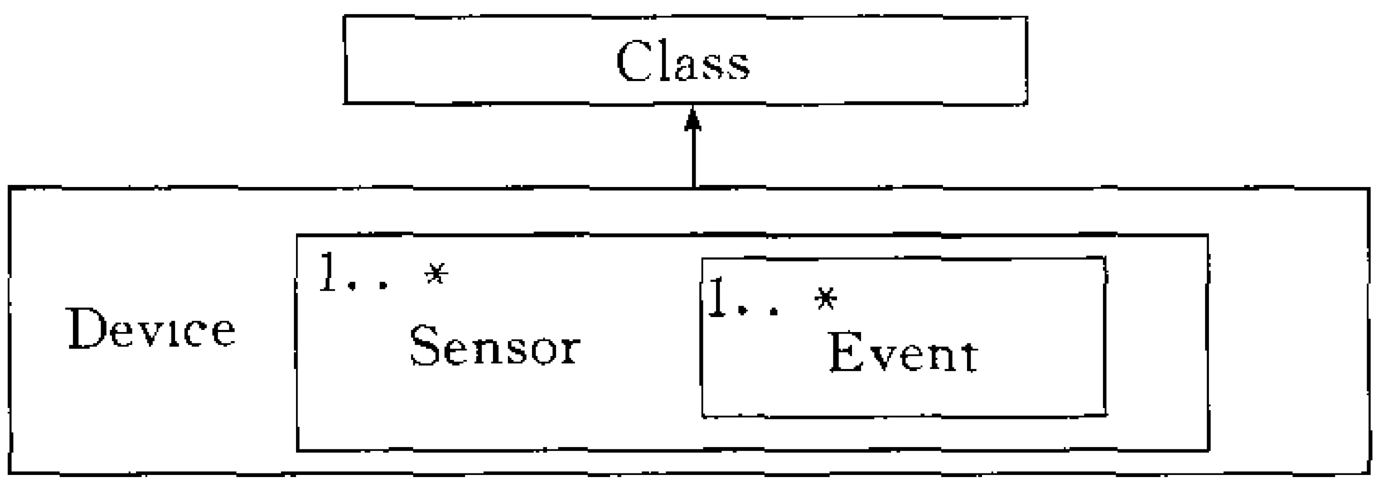

Fig. 7 A generic device model

As shown in Fig. 7, like object-oriented programming, each device instance belongs to a class of devices (e. g. an intelligent camera device belongs to the class 'Intelligent Camera'). Each device consists of one or more (1.. *) sensors (e. g. an intelligent camera device consists of 4 camera sensors), and each sensor defines one or more (1.. *) event objects (e. g. the events monitored by each camera sensor in the intelligent camera device, such as standing, intrusion, queue lengths, counter-flow pedestrian traffic and people distribution). To ease the implementation of the interfaces and to enable the extension of the protocol, the protocol is defined using XML (eXtensible Markup Language). Fig. 8 shows a sample register message defined for registering a device to the PRISMATICA system. For detailed information on the communication and protocol, please refer to [14].

```
<? xml version="1.0"?>
<Register version="1.0" Name="Intelligent Camera" MinorName="2.0" SerialNo="Intell"
        ReleaseDate="23/12/2001" NoSensor="1">
  <Sensor Name="Camera-1" SensorType="Video" NoEvent="1">
    <Event Name="Crowd" CanDisable="1" Type="Measure" SubType="Pulse" Priority="3">
      <Description>Area is crowded</Description>
      <EventConfiguration No="2">
        <Setting Name="Threshold" Title="Overcrowding" Type="NP" Range="0,100">
          <Default No="1"><Item value="30"/></Default>
          <Init No="1"><Item value="40"/></Init>
        </Setting>
        <Setting Name="AOI" Title="Area of interest" Type="AOI" Range="Unknown">
          <Default No="1"><Item value ="10,10"></Default>
          <Init No="1"><Item value="3,3"></Init>
        </Setting>
      </EventConfiguration>
      <EventReading No="0"/>
    </Event>
    <SensorConfiguration No="0"/>
    <Location position="20,30,1"/>
    <Coverage No="3">
      <Cell position="20,30,0"/>
      <Cell position="19,29,1"/>
      <Cell position="20,30,1"/>
    </Coverage>
  </Sensor>
  <Configuration No="1">
    <Setting Name="Site" Title="Geographic location" Type="TEXT" Range="1,40">
      <Default No="1"><Item value="Paris"/></Default>
      <Init No="1"><Item value="London"/></Init>
    </Setting>
  </Configuration>
  <Other No="1">
    <Info name="Designer">Benny Lo</Info>
  </Other>
</Register>
```

Fig. 8 Sample device register message

3 Fusing visual and audio information

The PRISMATICA system is a mass scale system which consists of numerous devices with diverse functions. Although only high level event messages are sent to the MIPSA and presented to the operators, potentially, there could be too much information for the operator to process and s/he could not react to the situation immediately. As devices are monitoring different aspect in the railway, an incident could trigger multiple alarms from different devices. For example, in an assault situation, the audio surveillance device may detect people shouting, passengers may push the panic button on their smart cards, and the intelligent camera may detect certain sudden movement. Correspondingly, the audio surveillance device, the contactless passcard and the intelligent camera will send alarm messages simultaneously to the MIPSA, and three messages will be shown to the operator. Even though the messages denote the same incident, the operator has to go through all the messages, look at the associated video images, interpret the situation and take appropriate action. Instead of assisting the operator, these could complicate the surveillance process.

To simplify the surveillance process and speed up the incident identification, a graphical approach in combination with a Bayesian network is designed and incorporated in the PRISMATICA for fusing evidences from sensors.

3.1 Graphical approach to fuse information from different devices

As devices are specialised in monitoring different aspects in the railway, different information will be generated from the devices. To combine such diverse information, one approach is to gather the detection results based on the geographical location of the sensors. Several studies have been conducted on applying 3D scene modelling to model the environment in order to combine the detection results from different sensors and indicate the location of the incident[5,9]. However, in order to obtain an accurate 3D scene model, extensive measurements have to be carried out and any changes in the camera position or angle will lead to re-calculation of the scene model.

A traditional graphical approach is taken to gather different event information to pin point the exact location of an incident. A coverage area is defined in the protocol for each sensor, and the sensor coverage could be defined in the device locally or assigned manually using the MIPSA. As different stations have different layouts, a grid system is employed to define the coverage area as shown in Fig. 9 where a 1000×1000 grid is overlaid on the station map.

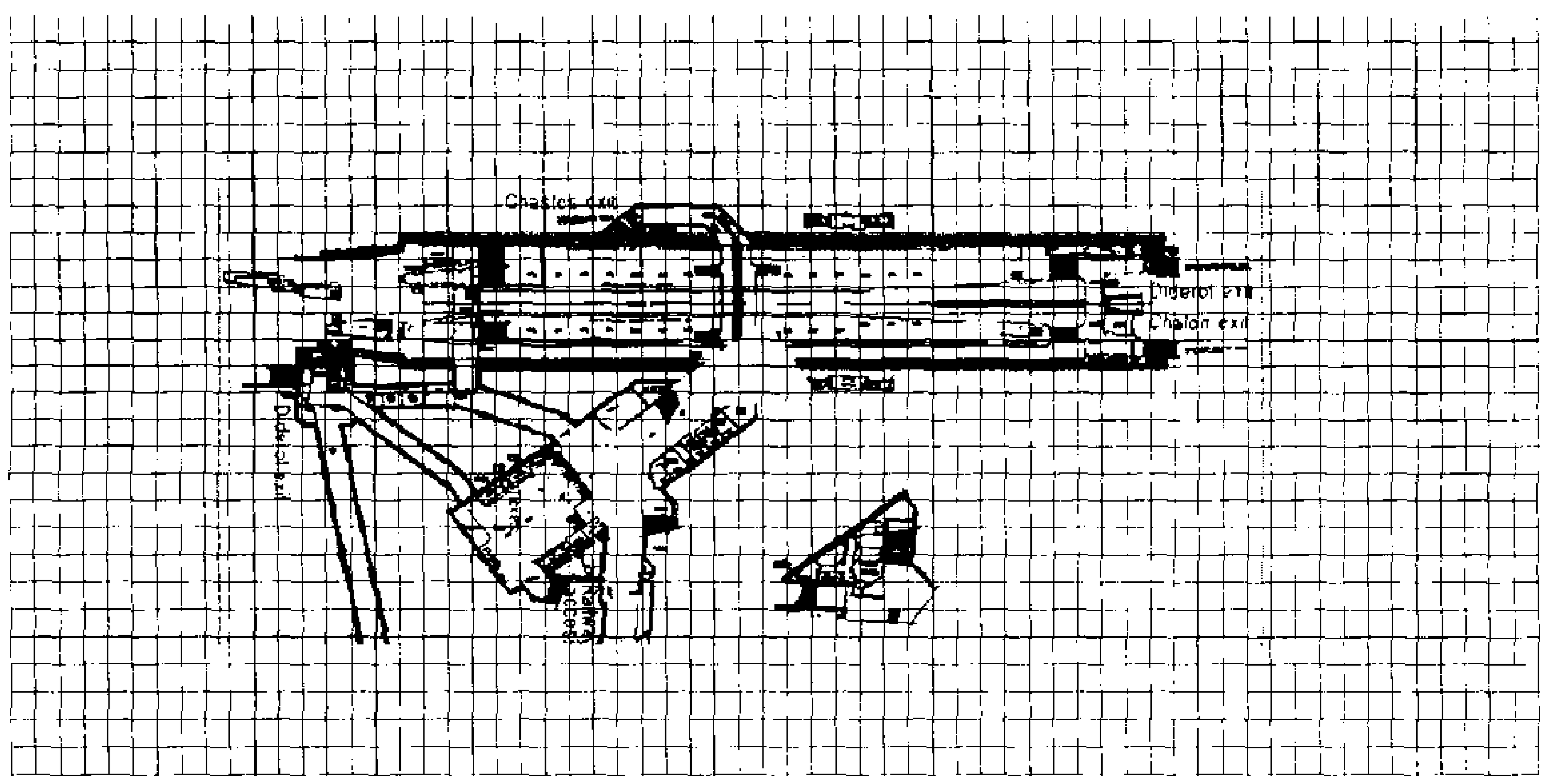

Fig. 9 Grid for defining sensor coverage

As shown in Fig. 9, a coverage area is a polygon with the coordinates according to its position in the grid. To cope with multiple levels in a station, the coordinates are defined as the 3 dimension coordinates where the third parameter (z axis) represents the level in the station.

When an event message is received by the MIPSA, the corresponding sensor and its coverage area are highlighted on the graphical user interface. In case of multiple event messages sent by sensors regarding an incident detected in an area, colour is used to illustrate the likelihood of the location of the incident. As the incident is more likely to occur in the common coverage area of the signalled sensors, brighter or more noticeable colours highlight the overlapping areas, similar to a Venn diagram.

$$P(x,y,z) = \frac{\sum_{i}^{N} SensorCoverage(i,x,y,z) \times SensorSignaled(i)}{\sum_{i}^{N} SensorCoverage(i,x,y,z)} \tag{1}$$

where $P(x,y,z)$ is the probability of an incident occurring in location (x,y,z), and N represents the total number of active sensors connected to MIPSA. In addition, the $SensorCoverage(i,x,y,z)$ represents the coverage of the i^{th} sensor, and $SensorSignaled(i)$ indicates if an event is detected by the i^{th} sensor.

Based on the $P(x,y,z)$, the corresponding colour chosen to highlight coverage area can be calculated as follows:

$$R(x,y,z) = 255 \tag{2a}$$

$$G(x,y,z) = \frac{255a}{\left(P(x,y,z) \times \sum_{i}^{N} SensorCoverage(i,x,y,z)\right)} \tag{2b}$$

$$B(x,y,z) = 0 \tag{2c}$$

where $R(x,y,z)$, $G(x,y,z)$ and $B(x,y,z)$ are the red, green and blue components respectively of the 24 bits colour chosen to draw the coverage area at position (x,y,z) of an alarmed sensor. The parameter α is a constant which denotes the colour levels that indicate the levels of likelihood on the map. An example of highlighting the sensor coverage areas for determining the location of an incident is shown in Fig. 10, where α is set to 3.

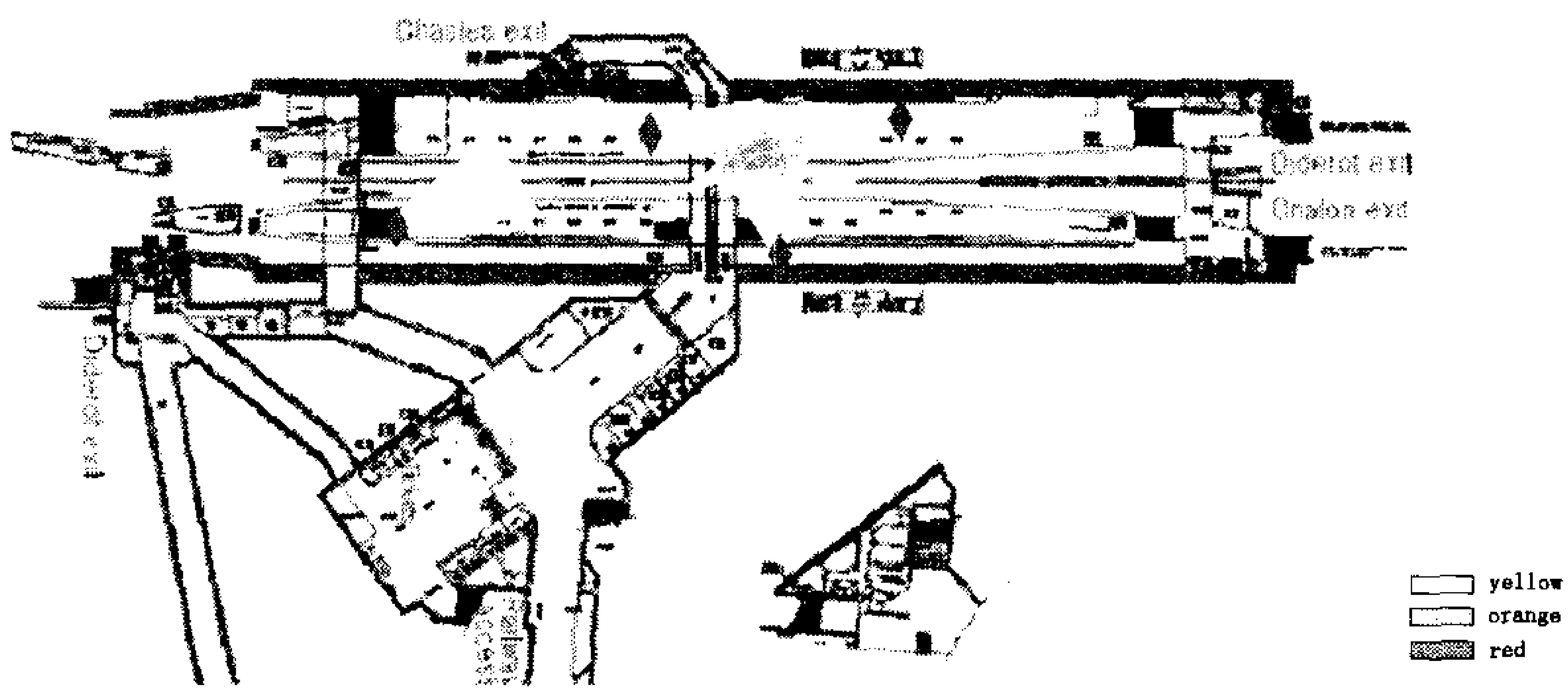

Fig. 10 Highlighting the likelihood of the location of an incident

3.2 Bayesian approach to fuse different evidences

To effectively reduce the amount of information presented to the operator without losing any significant evidence and assist the operator in determining the situation, a more sophisticated method has to be applied to fuse the diverse information from different sensors and identify incidents. As in the previous example where an assault has been detected by three different devices, instead of displaying the three different alarm messages (shouting, panic button alarm, and sudden movement), displaying a more descriptive alarm message, such as "possible assault", would not only assist the operator in identifying the situation, but it would also shorten the reaction time in coping with the incident. As each device is designed independently and it has its own set of alarm messages, to interpret and gather these different messages, a degree of artificial intelligence is required.

In this study, a Bayesian network (BN) approach has been chosen to fuse the different information sent by the devices and sensors. The graphical representation of the BN does not only provide a visual presentation of the reasoning behind the BN, but it also eases the expansion of the network to incorporate new evidences. In addition, the output of the BN is the posterior probability of the hypothesis that can be used to indicate the confidence level of the inference.

Based on the user requirement, several interesting scenarios have been chosen to implement the BN in this study, which are the trespassing, robbery, collision of flows, violence, suspected objects, overcrowding, congestion, rapid increase of crowding level, people too close to a platform edge, and blocking exit/entry. The structure of the BN is illustrated in Fig. 11.

Each child node shown in the graph represents an event detected by the respective device, and several intermediate nodes are designed to group the correlated events and enhance the inference of the network. As indicated in the graph, the BN is constructed based on the logic behind each scenario. For example, in the case of a passenger entering a forbidden area (trespassing), the intrusion alarm of the intelligent camera will be triggered,

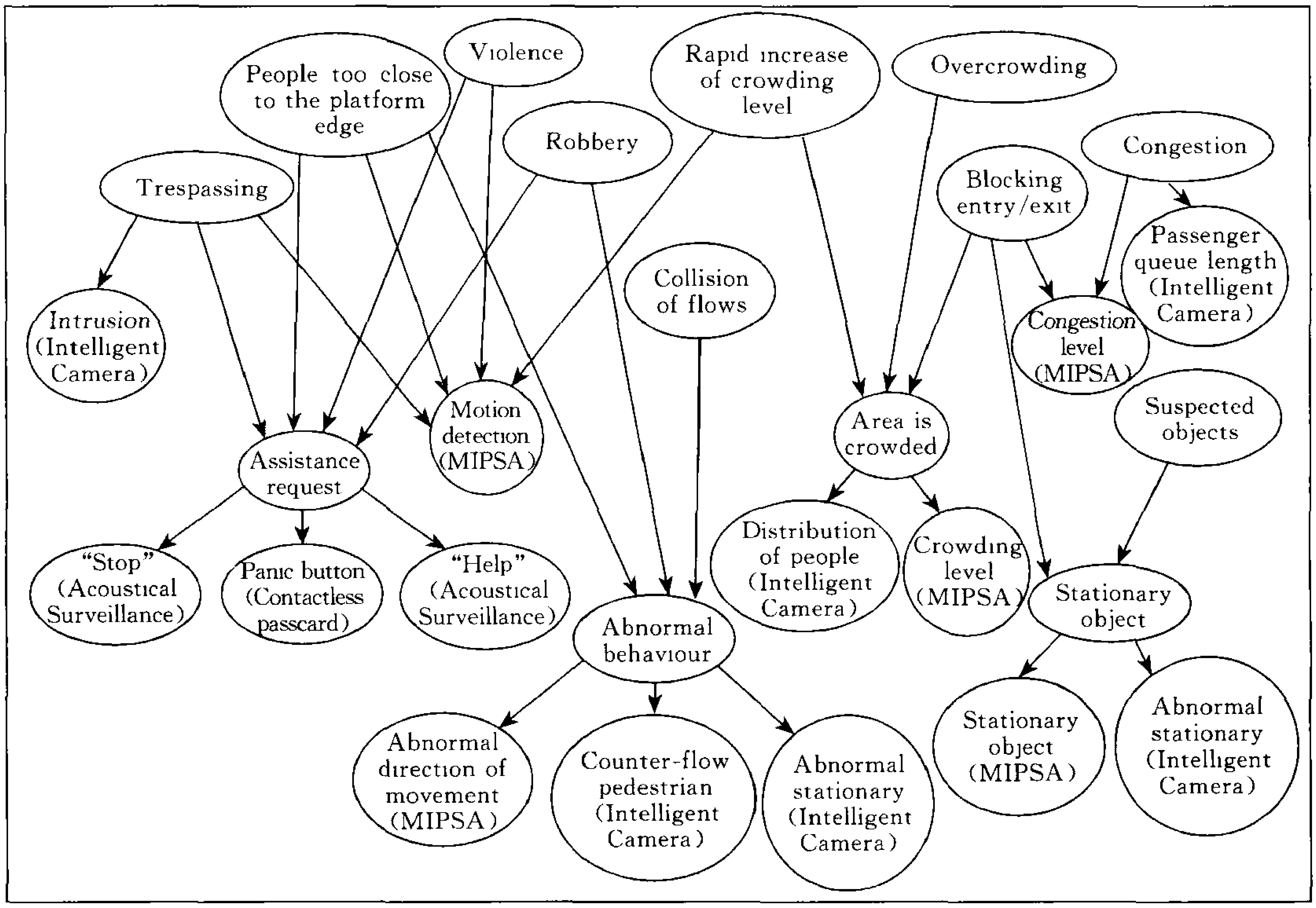

Fig. 11 Bayesian network for fusing multiple evidences

and at the same time, motion in the forbidden area will be detected by MIPSA. In addition, if any staff or passenger spots the trespassing, someone may shout "Stop", which will be detected by the audio surveillance system. As such the node "Trespassing" is assigned to be the parent of the nodes "Intrusion", "Assistance request" and "Motion detection". The posterior probability of the occurrence of trespassing is calculated as follows:

$$P(Trespassing = 1) = \alpha\lambda(Trespassing = 1)\pi(Trespassing = 1) \tag{3}$$

where α is the normalisation constant, and $\lambda(Trespassing=1)$ and $\pi(Trespassing=1)$ are the λ and π evidence of the trespassing node respectively. $\lambda(Trespassing=1)$ and $\pi(Trespassing=1)$ are calculated as follows:

$$\begin{aligned}\lambda(Trepassing = 1) &= \lambda_{Intrusion}(Trepassing = 1) \times \lambda_{Assistance}(Trepassing = 1) \times \\ &\quad \lambda_{Motion}(Trepassing = 1) \\ \pi(Trepassing = 1) &= Prior(Trepassing = 1)\end{aligned} \tag{4}$$

where $\lambda_{\text{Intrusion}}(Trepassing=1)$, $\lambda_{\text{Assistance}}(Trepassing=1)$ and $\lambda_{\text{Motion}}(Trepassing=1)$ are the λ messages sent by the nodes "Intrusion", "Assistance request" and "Motion detection" respectively, and Prior($Trepassing=1$) is the prior probability of the node "Trespassing" with state equal to 1. The λ messages are formulated as follows:

$$\begin{aligned}\lambda_{\text{Intrusion}}(Trepassing=1) &= \sum_{i=0}^{1} P(Intrustion = i \mid Trepassing = 1)\lambda(Intrusion = i) \\ \lambda(Intrustion = i) &= \begin{cases}1, & Intrustion = i \\ 0, & Intrustion \neq i\end{cases}\end{aligned} \tag{5}$$

where $P(Intrustion=i \mid Trepassing=1)$ is the conditional probability between the nodes "Intrusion" and "Trespassing", and $\lambda(Intrustion=i)$ is the λ evidence of the node "Intrusion". If an intrusion alarm is triggered, $\lambda(Intrustion=1)=1$ and $\lambda(Intrustion=0)=0$.

$$\lambda_{Assistance}(Trepassing = 1) =$$

$$\sum_{l=0}^{1}\sum_{k=0}^{1}\sum_{j=0}^{1}\sum_{i=0}^{1}\left[\begin{array}{l}\pi_{\text{Assistance}}(Robbery = l)\pi_{\text{Assistance}}(Violence = k) \\ \pi_{\text{Assistance}}(People = j) \\ P(Assistance = i \mid Trepassing = 1 \;\&\; People = j \;\&\; \\ \qquad Violence = k \;\&\; Robbery = l) \\ \lambda(Assistance = i)\end{array}\right]$$

$$\lambda(Assistance = i) = \lambda_{\text{stop}}(Assistance = i) \times \lambda_{\text{help}}(Assistance = i) \times \lambda_{\text{panic}}(Assistance = i)$$

$$\pi_{\text{Assistance}}(Robbery = l) = \frac{P'(Robbery = l)}{\lambda_{\text{Assistance}}(Robbery = l)}$$

$$\pi_{\text{Assistance}}(Violence = k) = \frac{P'(Violence = k)}{\lambda_{\text{Assistance}}(Violence = k)}$$

$$\pi_{\text{Assistance}}(People = j) = \frac{P'(People = j)}{\lambda_{\text{Assistance}}(People = j)} \tag{6}$$

where $\pi_{\text{Assistance}}(Robbery=l)$, $\pi_{\text{Assistance}}(Violence=k)$ and $\pi_{\text{Assistance}}(People=j)$ denote the π messages sent by the nodes "Robbery", "Violence" and "People too close to the platform edge" respectively to the "Assistance request" node. Like $P(Intrustion=i \mid Trepassing=1)$, the $P(Assistance=i \mid Trepassing=1 \;\&\; People=j \;\&\; Violence=k \;\&\; Robbery=l)$ represents the conditional probabilities between the node "Assistance request" and its parent nodes. In addition, $P'(Robbery=l)$, $P'(Violence=k)$ and $P'(People=j)$ represent the posterior probabilities of the nodes "Robbery", "Violence" and "People too close to the platform edge" respectively. The $\lambda_{\text{stop}}(Assistance=i)$, $\lambda_{\text{help}}(Assistance=i)$ and $\lambda_{\text{panic}}(Assistance=i)$are the λ messages sent by the children nodes "Stop", "Help" and "Panic button" respectively to the node "Assistance request", and the equations to calculate these λ messages are as follows:

$$\lambda_{\text{stop}}(Assistance=i)=\sum_{j=0}^{1}P(Stop=i \mid Assistance=i)\lambda(Stop=i) \tag{7a}$$

$$\lambda(Stop = j) = \begin{cases}1, & Stop = j \\ 0, & Stop \neq j\end{cases} \tag{7b}$$

$$\lambda_{\text{help}}(Assistance=i)= \sum_{j=0}^{1}P(Help=i \mid Assistance=i)\lambda(Help=i) \tag{7c}$$

$$\lambda(Help = j) = \begin{cases}1, & Help = j \\ 0, & Help \neq j\end{cases} \tag{7d}$$

$$\lambda_{\text{panic}}(Assistance=i)= \sum_{j=0}^{1}P(Panic=i \mid Assistance=i)\lambda(Panic=i) \tag{7e}$$

$$\lambda(Panic = j) = \begin{cases}1, & Panic = j \\ 0, & Panic \neq j\end{cases} \tag{7f}$$

where $P(Stop=i \mid Assistance=i)$, $P(Help=i \mid Assistance=i)$ and $P(Panic=i \mid Assistance=i)$are the conditional probabilities between the node "Assistance request" and its children nodes, "Stop", "Help" and "Panic button" respectively. Like the intrusion alarm, if "Stop" is detected by the audio surveillance, $\lambda(Stop=1)=1$ and $\lambda(Stop=0)=0$.

$$\lambda_{\text{Motion}}(Trepassing = 1) = \sum_{l=0}^{1}\sum_{k=0}^{1}\sum_{j=0}^{1}\sum_{i=0}^{1}\left[\begin{array}{l}\pi_{\text{Motion}}(Rapid = l)\pi_{\text{Motion}}(Violence = k) \\ \pi_{Motion}(People = j) \\ P(Motion = i \mid Trepassing = 1 \;\&\; People = j \;\&\; \\ \qquad Violence = k \;\&\; Rapid = l) \\ \lambda(Motion = i)\end{array}\right] \tag{8a}$$

$$\lambda(Motion = i) = \begin{cases}1, & Motion = i \\ 0, & Motion \neq i\end{cases} \tag{8b}$$

$$\pi_{\text{Motion}}(Rapid = l) = \frac{P'(Rapid = l)}{\lambda_{\text{Motion}}(Rapid = l)} \tag{8c}$$

$$\pi_{\text{Motion}}(Violence = k) = \frac{P'(Violence = k)}{\lambda_{\text{Motion}}(Violence = k)} \tag{8d}$$

$$\pi_{\text{Motion}}(People = j) = \frac{P'(People = j)}{\lambda_{\text{Motion}}(People = j)} \tag{8e}$$

where $\pi_{\text{Motion}}(Rapid=l)$, $\pi_{\text{Motion}}(Violence=k)$ and $\pi_{\text{Motion}}(People=j)$ denote the π messages sent by the nodes "Rapid increase of crowd level", "Violence" and "People too close to the platform edge" respectively to the "Motion detection" node. The conditional probability between the node "Motion detection" and its parents is represented by $P(Motion=i \mid Trepassing=1 \,\&\, People=j \,\&\, Violence=k \,\&\, Rapid=l)$. $\lambda(Motion=i)$ denotes the λ evidence of the node "Motion detection". In addition, $P'(Rapid=l)$, $P'(Violence=k)$ and $P'(People=j)$ represent the posterior probabilities of the nodes "Rapid increase of crowd level", "Violence" and "People too close to the platform edge" respectively.

However, as sensors are scattered throughout the stations, the geographical location of the sensors have to be considered in fusing the different sensor detections. In addition, in some areas only certain sensors will be installed, as not all the detection is applicable in all locations. For example, the incident "People too close to the platform edge" will only occur on platforms. In order to incorporate the geographical information of the sensors, sub-networks will be built in different areas to monitor different events. For example, in the stairs, the sub-network will appear as in Fig. 12, where only the events, violence and robbery, will be monitored. Table 1 lists all the events and the areas where the events could occur.

Table 1 Events and the concerning areas

Events	Areas
Violence	Corridor, Hall, Platform, Stairs
People too close to the platform edge	Platform
Trespassing	Hall, Platform
Robbery	Corridor, Hall, Platform, Stairs
Collision of flows	Corridor, Platform
Suspected objects	Corridor, Hall, Platform
Overcrowding	Hall, Platform
Congestion	Hall, Platform
Rapid increase of crowding level	Hall, Platform
Blocking exit/entry	Corridor, Hall, Platform

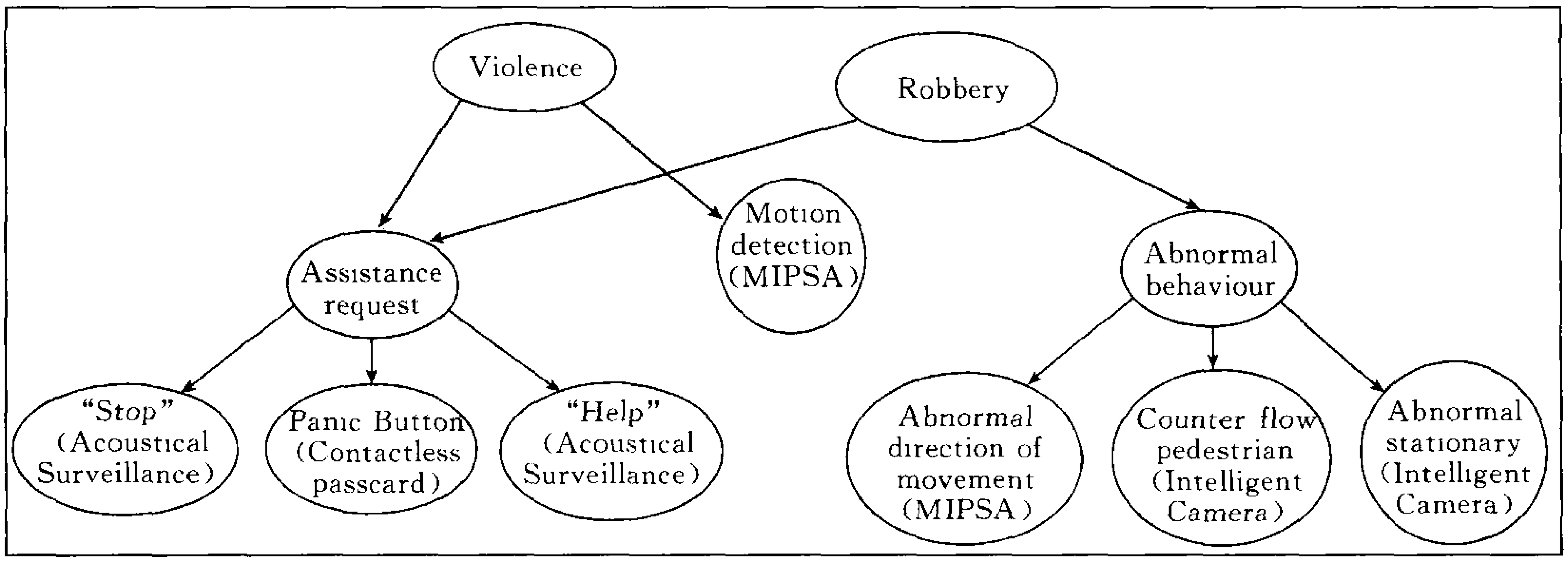

Fig. 12 The sub-Bayesian network for detecting incident in stairs

To ease the integration of the PRISMATICA system and enable supervised training of the BN using the MIPSA, the MIPSA is designed to enable the user to identify different area types (i. e. Corridor, Hall, Platform or Stairs) on the map, as per assigning the coverage areas for sensors where a grid is used to define the coordinates on the map. In addition, based on the area type, different events will be monitored and different BNs will be constructed accordingly. Fig. 13 shows an example of defining the area types on the map using MIPSA.

Apart from the possibility of events occurring in certain locations, the sensor coverage is also one of the major factors to be considered in building the BN. Since only sensors with overlapping coverage could detect and send information regarding an incident, BN is only be applied in the area where overlapping coverage is found.

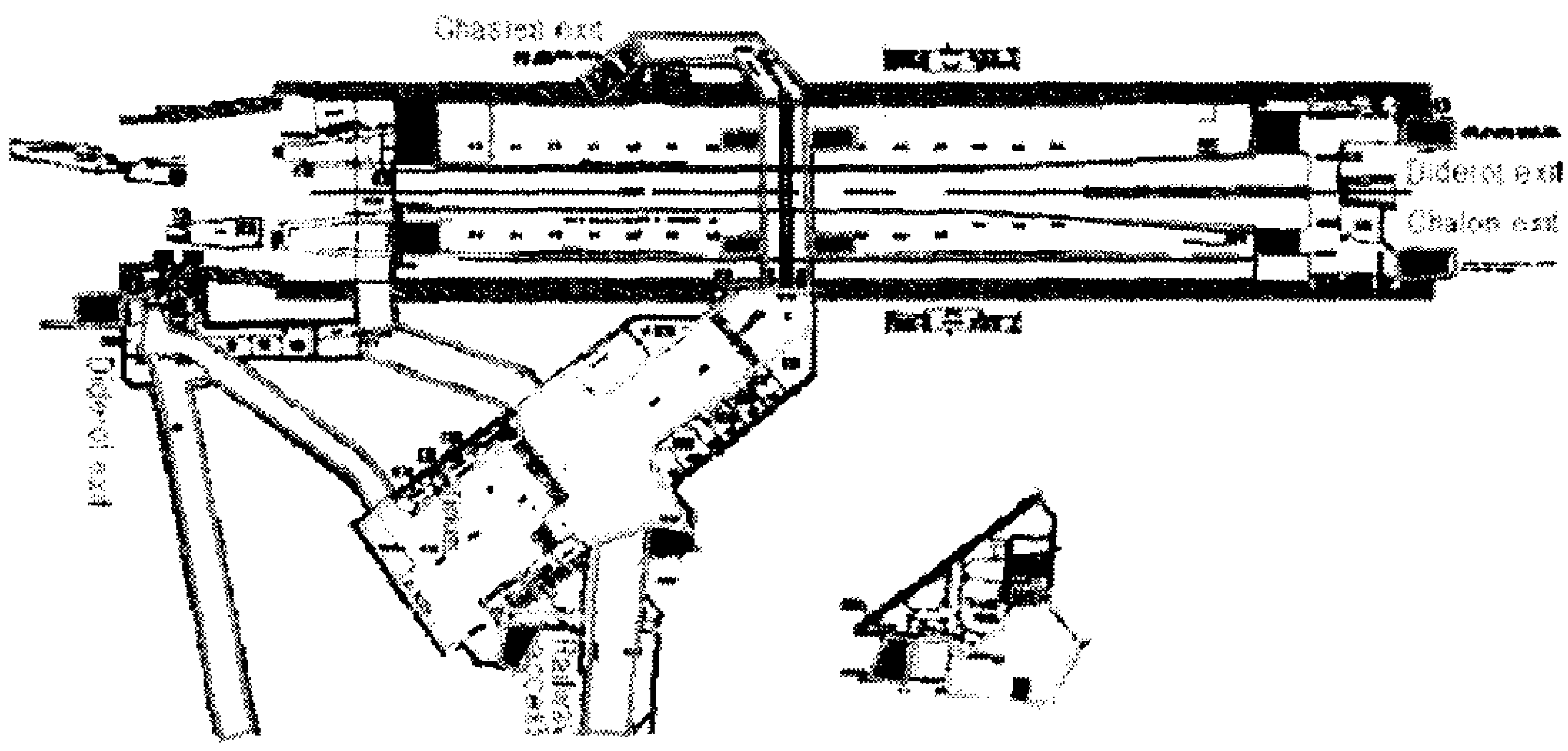

Fig. 13 Defining area types on the MIPSA where the stair, corridor and platform areas are defined by different color polygons respectively

4 Conclusion and discussion

A novel concept of applying a graphical approach together with Bayesian networks is presented for fusing the information detected by visual, audio and other types of devices. The proposed graphical presentation of the sensor coverage area not only indicates the likelihood of the location of an incident, it also enables the incorporation of geographical information in constructing the Bayesian network (BN). On the other hand, the BN provides an inference mechanism to fuse the diverse information from different detection devices and provides more descriptive information for the operator to assess incidents. A synthetic data set has been used to validate the concept of applying BN in fusing different information. At the moment, we are in the process of collecting data for constructing the BN for the real scenarios and updating the MIPSA to apply the concept in the actual operating environment. Recently, the MIPSA system has been installed in the Liverpool Street station in London and Gare de Lyon station in Paris for integrating with other devices and validating its detection functions in the real environment.

Acknowledgements This study has been carried out as part of the EU project PRISMATICA (GRD1-2000-10601). The project involves six major European metro operators (RATP-Paris, LUL-London, ATM-Milan, STIB-Brussels, PPT-Prague, ML-Lisbon), research centres (Kings College London, University College London, Kingston University, INRETS-France, CEA-France) and commercial companies (TIS-Portugal, SODIT-France, FIT-Italy, ILA-Germany, Thales-France). The authors are grateful to London Underground, the Paris Metro and Newcastle International Airport for providing access to their sites and staff.

References

1 Boghossian B A. Motion-based image processing algorithm[Ph. D. Thesis]. London: King's College, 2000

2 Sacchi C, Regazzoni C. A distributed surveillance system for detection of abandoned objects in unmanned railway environments. *IEEE Transactions on Vehicular Technology*, 2000, **49**(5): 2013~2026

3 Fuentes L M, Velastin S A. People tracking in surveillance applications. In: 2nd IEEE International Workshop on Performance Evaluation on Tracking and Surveillance, Hawail, 2001

4 Siebel N T, Maybank S. Fusion of multiple tracking algorithms for robust people tracking. In: Proceedings of the European Conference on Computer Vision 2002, 2002. 373~387

5 Rota N, Thonnat M. Video sequence interpretation for visual surveillance. In: Proceedings of the 3rd IEEE International Workshop on Visual Surveillance, 2000. 59~68

6 Bouchafa S, Aubert D, Bouzar S. Crowd motion estimation and motionless detection in subway corridors by image

processing. In: Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC'97), IEEE Press, 1997. 332~337
7 Boghossian B A, Velastin S A. Real-time motion detection of crowds in video signals. In: IEE Colloquium on High Performance Architectures for Real-Time Image Processing, London: IEE, 1998. 12/1~12/6
8 Lo B, Velastin S A. Automatic congestion detection system for underground platforms. In: Proceedings of the IEEE 2001 International Symposium on Intelligent Multimedia, Video & Speech Processing, Hong Kong: IEEE Press, 2001. 158~161
9 Collins R T, Lipton A J, Fujiyoshi H, Kanade T. Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE*, 2001, **89**(10): 1456~1477
10 Marcenaro L, Oberti F, Foresti G L, Regazzoni C S. Distributed architectures and logical-task decomposition in multimedia surveillance systems. *Proceedings of the IEEE*, 2001, **89**(10): 1419~1440
11 Aime Cabie, Velastin S A, Sun J, Corvee E, Vicencio-Silva M A, Heath C, Sanchez-Svensson M, Lionti R, Aubert D, Khoudour L, Bruyelle J. D9~D10: Verification of tools and processes and their interaction. In: PRISMATICA Project(GRD-2000-10601), Brussels: European Commission, 2002
12 Velastin S A, Sanchez-Svensson M, Sun J, Vicencio-Silva M A, Aubert D, Lemmer A, Brice L, Khoudour L, Kallweit S. Deliverable D7: Innovative tools for security in transports. In: PRISMATICA Project (GRD-2000-10601), European Commission, Brussels: European Commission, 2002
13 Schmidt D C, Gokhale A, Harrison T H, Parulkar G. A high-performance endsystem architecture for real-time CORBA. In: IEEE Communications Magazine Feature Topic Issue on Distributed Object Computing, IEEE Press, 1997. 72~77
14 Lo B, Velastin S A. Communication Specification. In: PRISMATICA Project(GRD-2000-10601), Brussels: European Commission, 2002

Benny Ping Lai Lo Received his bachelor degree in electrical and computer engineering from the University of British Columbia(Canada) in 1995 and master with distinction in electronic research(computer vision) from King's College London(UK) in 2000. From 1995~1997, he worked as a project engineer in Cybermation System Inc. (Canada) in designing SCADA system and GUI for CCTV systems. From 1997 until 1999, he joined the Mass Transit Railway Corporation(Hong Kong) where he worked as a design officer responsible for designing building automation system. He is currently working as a senior researcher in Kingston University(UK) Under two EU projects, ADVISOR and PRISMATICA, where he designed a flexible network-based communication protocols which allows a diverse range of devices (from video processors to other sensors and actuators) to be integrated easily to a distributed visual surveillance system.

Jie Sun Received his master degree in Control Engineering from the Harbin Institute of Technology(PRC) and, after having worked as computer network system engineer for a major company for two years, he has worked in the EU PRISMATICA project since its inception. He was responsible for developing the user interface and the back-end database for the PRISMATICA system and has been coordinating the work of a multinational team on communication protocols and interfacing to the PRISMATICA system.

Sergio A. Velastin Received his Ph. D. degree from the university of Manchester(UK) for research on vision systems for pedestrian and road-traffic analysis. He then worked in industrial R&D and project management before joining the Dept. of Electronic Engineering in Kings College London(University of London). In 2001, he and his team joined the Digital Imaging Research Center in Kingston University, attracted by its size and growing reputation in the field. Currently, he is Technical Coordinator of the EU-funded project PRISMATICA(TREN-Growth) on the synergetic integration of technology(e. g. networking, video/audio processing, wireless transmission) and human-based processes for improving actual and perceived security in public transport systems. His main current interest is distributed intelligent surveillance systems.

用于公众交通系统的分布式智能监控系统中的视听信息融合

Benny Ping Lai Lo　Jie Sun　Sergio A. Velastin

(*Digital Imaging Research Center, School of Computing & Information Systems, Kingston University, Kingston upon Thames*, 英国)

(E-mail: {b. lo, jie. sun, sergio. velastin}@kingston. ac. uk)

摘　要　提出了一个全新的概念,该概念表述了通过融合来自分布式视听处理系统的不同信息来提高事故检测鲁棒性以及提供更多的事件描述. 最后利用来自伦敦和巴黎的现场测试验证了该系统的性能. 本文是以欧盟的 PRISMATICA 项目为基础.

关键词　Video/audio 算法,智能交通系统,智能摄像机

中图分类号　TP391.41