The ADVISOR Visual Surveillance System

Nils T Siebel[1] and Stephen J Maybank[2]

[1] Cognitive Systems Group, Institute of Computer Science and Applied Mathematics,
Christian-Albrechts-University of Kiel, Olshausenstr. 40, 24098 Kiel, Germany
nils@siebel-research.de

[2] School of Computer Science and Information Systems, Birkbeck College,
University of London, Malet Street, London WC1E 7HX, Great Britain
sjmaybank@dcs.bbk.ac.uk

Abstract. ADVISOR is an automated visual surveillance system for metro stations which was developed as part of the project ADVISOR, involving 3 academic and 3 industrial project partners. The ADVISOR system aims at making public transport safer by automatically detecting at an early stage dangerous situations which may lead to accidents, violence or vandalism. In order to achieve this people are tracked across the station and their behaviours analysed. Additional measurements on crowd density and movement are also obtained. Warnings are generated and displayed to human operators for possible intervention.

The article explores the main difficulties encountered during the design and implementation of ADVISOR and describes the ways in which they were solved. A prototype system has been built and extensively tested, proving the feasibility of automated visual surveillance systems. An analysis of test runs at a metro station in Barcelona and several individual experiments show that the system copes with many difficult image analysis problems. The analysis also points the way for future development and ways of deployment of the techniques used in the system.

1 Introduction

The visual surveillance of metro stations is becoming increasingly important from both industrial and social points of view. Many station operators have installed networks of video cameras to make their installations safer and to protect passengers and equipment. The cameras are a means of detecting situations such as accidents, violence and vandalism. Traditionally, human operators watch banks of monitors where selected camera views are displayed. However, out of the large number of cameras (typically around 50–100 for a medium-sized station) only a few can be monitored at any one time. Therefore some accidents are not detected quickly and vandalism and violence do still occur.

The ADVISOR[3] system addresses this issue by providing a means to support surveillance personnel in their difficult and complex task. This is done by the

[3] Annotated Digital Video for Intelligent Surveillance and Optimised Retrieval, project web site http://www-sop.inria.fr/orion/ADVISOR/

realtime analysis of video feeds from surveillance cameras and the generation of warnings for the operators when a situation is detected which requires their attention. ADVISOR is the first system that integrates people tracking, crowd monitoring and behaviour analysis functionalities for this purpose. Related work by previous authors is described in [1–4]. A prototype of the ADVISOR system has been developed over 3 years by three academic and three industrial partners using a total man power of 35 person years. Close links to end-users (transport operators in Barcelona, Brussels and London) and test runs in their stations have ensured that the system is suitable for commercial exploitation.

The remainder of this article is organised as follows. Section 2 provides an overview of the ADVISOR system and its architecture. Section 3 focuses on the central image processing tasks, people tracking and crowd analysis, which obtain the information needed for subsequent analysis and interpretation of the scene. Section 4 details the detection of events and explains how human operators are notified of them. The evaluation of ADVISOR by end-users and their analysis is presented in Section 5. Conclusions derived from this analysis can be found in Section 6.

2 The ADVISOR System

The main actions and features of the ADVISOR system are as follows.

- Tracking people in camera images from multiple cameras
- Estimating crowd density and movement
- Analysing people and crowd behaviours based on these measurements
- Generating alarms based on detected dangerous or criminal behaviours and display these to a human operator within a short time (limited delay)
- Archiving digitised video feeds and event annotations in an auditable and secure database and allowing for searches within this database through a human computer interface (HCI)
- The system is an economical and scalable, fully integrated unit with a single interface to a metro station's camera system.

2.1 System Architecture

Figure 1 shows an architectural overview of the ADVISOR system. The main ADVISOR processing unit is comprised of a number of rack-mounted standard PCs interconnected by a dedicated 100BaseT Ethernet network. The unit is connected to the station's network of surveillance cameras through standard analogue (composite) video connectors. Its outputs are digitised video sequences with annotations describing the detected events. The outputs are displayed to the operators through a Human Computer Interface (HCI) located on a separate standard PC connected to the main processing unit by Ethernet.

The exchange of information between the individual subsystems is realised using the following international standards:

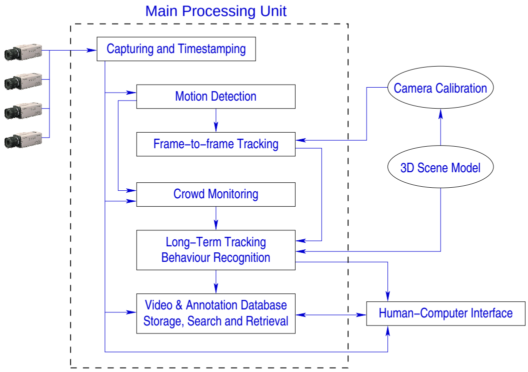


Fig. 1. System architecture with functional units

- Digital video data is exchanged using Baseline JPEG (ISO/IEC IS 10918-1) and the JFIF 1.02 file format (compression ratio approximately 10 : 1).
- Video annotations and other textual data (e.g. control commands) are transmitted using the XML version 1.0 file format. XML Schemas [5] are used to define the syntax of these files (and thereby the interfaces between subsystems), allowing the automatic validation of XML data in these files.
- The module interconnections are realised using the TAO Realtime CORBA implementation with Naming Services.
- On a lower level, high-speed data channels are implemented using a mixture of UDP Unicast and TCP/IP Multicast.

2.2 Processing Units

Processing is broken down into the following sub-tasks:

Capture: The analogue video feeds from the cameras are digitised, subsampled to 384 × 288 pixel images at 5 frames per second (fps) and JPEG compressed. In order to enable a complete audit trail each JPEG file has added to it a time-stamp header which includes the specific time, location and hardware information for the corresponding video image. The digitised image is transmitted to other computers on the Ethernet via IP Multicast.

Motion Detection: A fast motion detection algorithm classifies as "moving" those pixels in the image which do not belong to the background scene. It generates a binary Motion Image where white pixels represent moving pixels and black pixels represent the background. This image is made available to other modules.

Frame-to-Frame Tracker: Based on grouped "moving regions" in the Motion Image and additional measurements in the video image, individuals and groups of people are tracked from frame to frame. Tracking and estimation of the size of tracked objects are both done in 3D using camera calibration.

Crowd Monitoring: The Crowd Monitoring Module estimates crowd density and movement in the image. Data obtained from this analysis are provided to other modules for subsequent analysis.

Behaviour Recognition: Using data from tracking and crowd analysis modules the Behaviour Recognition subsystem detects events and behaviours of individuals and groups which are to be flagged. Warnings are generated and transmitted to subsequent modules.

HCI: The Human Computer Interface (HCI) displays video feeds to the operator, using output from Behaviour Recognition in order to select "interesting" video channels, including those where dangerous situations were detected.

Archive: In the Archive all digital video feeds and annotations are stored and can be queried through the HCI using a range of database search options.

Subsystems implementing the first three tasks are run on the same PC, the others on a separate PC for each task. The following sections provide more details on the most important subsystems.

3 Image Processing

Image processing is done at low level by the Motion Detector. It uses a simple background model which is periodically updated from the live video feed in order to adapt the system to changes in the scene e.g. by lighting changes and camera movements. There are two outputs from the Motion Detector: a binary Motion Image where some pixels are classified as "moving" and the generated background model in form of an "empty" Background Image. Together with the current video image these images are provided to higher-level image analysis modules to obtain more abstract descriptions of images and objects in the scene.

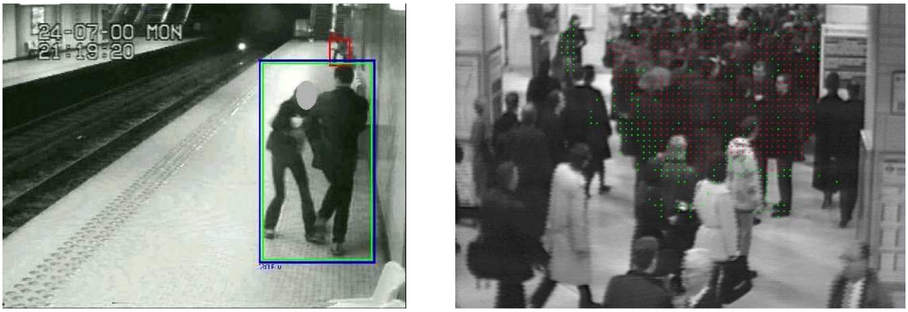
3.1 People Tracking

In order to detect and understand human behaviour it is important that the ADVISOR system locates and tracks individuals as well as groups of people. The following difficulties exist:

- People can occlude each other in the image
- People can have a low contrast to the background so that they cannot be detected easily by the system
- The image quality and lighting conditions (visibility) in metro stations are usually bad.

People tracking was pioneered by O'Rourke and Badler [6] and Hogg [7] in the early 80's and a number of algorithms can be found in the literature [2, 4, 8–14].

The people tracker integrated into the ADVISOR system analyses and tracks moving regions extracted from the Motion Image [14]. Using camera calibration these regions are classified into individuals, groups of people and a few other



(a) Tracked group of people

(b) Analysing dense crowd movement

Fig. 2. Image Analysis: (a) People/Group Tracking; and (b) Crowd Movement Analysis

classes (e.g. train) according to their size and shape. The output of the Frame-to-Frame Tracker is used to construct a tracking graph which facilitates the tracking of individuals over a long period of time, even if they join or leave groups. Robust long-term tracking is achieved by refining the tracking hypotheses over time. People can be tracked across cameras (handover) with the help of camera calibration and a scene model.

3.2 Crowd Analysis

The Crowd Monitoring Module was adapted to ADVISOR from a similar module developed within a previous project [15]. It uses a dedicated video processing board (Sollatek Trimedia STM-1300) to carry out motion estimation on JPEG video feeds sent on the Ethernet. The outputs from the DSP board consist of movements of 8×8 pixel blocks in the image. This data is combined with Motion Data from the background model and analysed for crowd density, movement and direction. Using prior knowledge of the scene observed by a given camera it can detect the following potentially dangerous situations:

- Overcrowding and blocking of areas
- Stationarity of objects and people
- Congestion of pre-defined areas (e.g. exits or escalators)
- Counter-flow, i.e. movement of people against the main flow or in the opposite direction of one-way paths

All output from the image analysis stage is used in the Behaviour Recognition module and stored in the video and annotation database for future access.

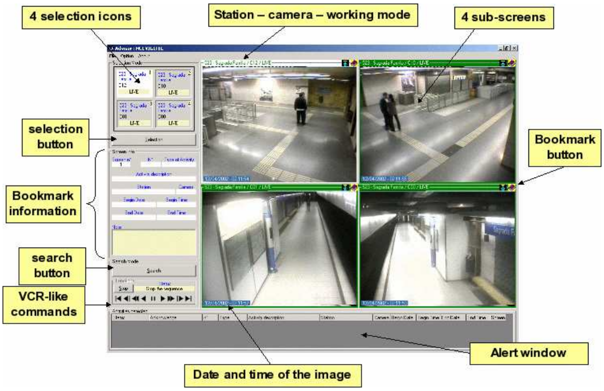


Fig. 3. Surveillance Operator's View (HCI)

4 Interpretation of Scenes

At the interpretation stage image measurements and tracking data are analysed in the short-term and in the long-term, in order to detect and flag situations which might require intervention by the station operators.

4.1 Behaviour Analysis

The Behaviour Analysis subsystem has a number of pre-defined behaviours it has to detect. It analyses the output from the Frame-to-Frame Tracker to detect states (e.g. a person is running), events (e.g. a group starts to be agitated) and scenarios (combinations of states and events). Using this formalism behaviours can be associated with special scenarios (e.g. a group is fighting). Tracking outputs from multiple cameras are combined to overcome problems of occlusion and to support long-term tracking of individuals and groups. Using a Scene Model movements of people are interpreted in the 3D context. Details on the algorithms can be found in [14].

In addition to the crowd behaviours listed in section 3.2 above the following individual behaviours can be detected by the ADVISOR system:

- Violence between people
- Vandalism against equipment (e.g. vending machines)
- Fare evasion (people jumping barriers instead of using a ticket).

4.2 The Human Computer Interface

A Human Computer Interface (HCI) was developed based on requirements derived from discussions with metro operators. Figure 3 shows a screenshot of the system in operation. Four video feeds are displayed to the operator together with

Table 1. Evaluation at Sagrada Família: Detection Rates

Behaviour	Number of Incidents	Detection Rate
Fighting	21	95 %
Blocking	9	78 %
Jumping over Barrier	42	88 %
Vandalism	6	100 %
Overcrowding	35	80 %

any alerts (e.g. warnings about detected events) associated with them. All messages by the system are logged and can be reviewed. Through the HCI the video and annotation database can be accessed e.g. by searching for specific events or bookmarks set earlier. This allows to play back video for post-incident analysis by operators or the Police.

5 Evaluation and Analysis of Experiments

At various points in the project, representatives of two metro companies were invited to evaluate the system from a user viewpoint and provide valuable feedback. For the final evaluation a prototype was installed in the Sagrada Família metro station in Barcelona and demonstrated to various guests including representatives of the Brussels and Barcelona metros. The validation process involved the behaviours fighting, blocking, overcrowding, jumping over the barrier and vandalism, as well as database searches. The evaluation, validation and demonstrations were conducted using both live and recorded video over 4 hours using four parallel input channels. Three channels were pre-recorded video sequences with the behaviours to be recognised and one was a live input sequence from the main hall of the Sagrada Família station. The accuracy of ADVISOR's warnings was measured using ground truth.

Table 1 shows the detection rates during the evaluation process. Only one false alarm (for the *blocking* behaviour) was generated during the whole test. Where detection failed (false negatives) a later analysis showed in most cases inaccurate tracking results and a sensitivity to detection parameters to be causing the problem.

An in-depth analysis of the tests also revealed the problems which still remain for ADVISOR and similar systems:

- Bad image quality and lighting conditions make robust detection and tracking difficult
- Placement and viewing angles of already installed cameras in metro stations often create ambiguities difficult for computer vision systems
- Analysis of behaviours heavily depends on accurate tracking output. Due to the abovementioned problems this is not always available.

From a practical point of view further work is necessary to reduce the complexity of the process of building a 3D scene model (currently very time-consuming) and to reduce the size (number of computers) of the main processing unit.

Overall the results look promising, although for a better analysis of the detection of rare events the system will have to be tested for a much longer time. Detailed feedback from the evaluators showed a very positive response, expressing that a system like ADVISOR could be helpful in their daily task.

6 Conclusions and Future Work

The ADVISOR system introduced in this article successfully integrates people tracking, crowd analysis and behaviour analysis functionalities in a realtime surveillance system for metro stations. Warnings about detected events that require human intervention are automatically generated and displayed to operators. An extensive on-site evaluation of a prototype by experts turned out very positive. It shows the feasibility of the approach and also points the way for future development.

Further work that will need to be performed to address problems encountered during validation includes:

- Behaviour recognition needs to be extended to include a wider range of behaviours and possibly supervised learning algorithms
- An improved people tracker will need to be integrated to make more accurate tracking information available to behaviour analysis. This would also allow the recognition of more types of behaviours. Currently, tests are being carried out to see whether integrating the people tracker given in [16] can improve the overall performance of ADVISOR.
- More use needs to be made of multiple cameras, and to achieve this, better quality images will be required. This would also be needed to meet the requirements of providing satisfactory evidence in court.
- The size of the equipment needed to implement a system must be reduced, so that the cost of installation and maintenance will be more acceptable. This has already been achieved when the complete system was recently implemented on a dual 3.0 GHz PC managing input from 4–8 cameras.
- Tools must be developed to reduce the installation and maintenance costs of scene modelling and camera calibration.

Acknowledgements

The research presented here has been carried out while the authors were working in the Computational Vision Group, Department of Computer Science at The University of Reading, UK. This research was supported by the European Union, grant ADVISOR (IST-1999-11287).[4]

[4] However, this paper does not necessarily represent the opinion of the European Community, and the European Community is not responsible for any use which may be made of its contents.

References

1. Oberti, F., Granelli, F., Regazzoni, C.: Minimax based regulation of change detection threshold in video-surveillance systems. In Foresti, G.L., Mähönen, P., Regazzoni, C.S., eds.: Multimedia Video-Based Surveillance Systems. Kluwer Academic Publishers, Boston, USA (2000) 210–223
2. Ohya, J., Utsumi, A., Yamato, J.: Analyzing Video Sequences of Multiple Humans. Kluwer Academic Publishers, Boston, USA (2002)
3. Chleq, N., Brémond, F., Thonnat, M.: Image understanding for prevention of vandalism in metro stations. In Regazzoni, C.S., Fabri, G., Vernazza, G., eds.: Advanced Video-based Surveillance Systems. Kluwer Academic Publishers, Boston, USA (1998) 106–116
4. Haritaoglu, I., Harwood, D., Davis, L.S.: W^4: Real-time surveillance of people and their actions. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 809–830
5. World Wide Web Consortium (W3C): XML Schema Part 0: Primer. W3C Recommendation. (2001) `http://www.w3.org/TR/xmlschema-0/`.
6. O'Rourke, J., Badler, N.: Model-based image analysis of human motion using constraint propagation. IEEE Transactions on Pattern Analysis and Machine Intelligence **2** (1980) 522–536
7. Hogg, D.: Model-based vision: A program to see a walking person. Image and Vision Computing **1** (1983) 5–20
8. Cai, Q., Mitiche, A., Aggarwal, J.K.: Tracking human motion in an indoor environment. In: Proceedings of the 2nd International Conference on Image Processing (ICIP'95). (1995) 215–218
9. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997) 780–785
10. Baumberg, A.M.: Learning Deformable Models for Tracking Human Motion. PhD thesis, School of Computer Studies, University of Leeds, Leeds, UK (1995)
11. Lipton, A.J., Fujiyoshi, H., Patil, R.S.: Moving target classification and tracking from real-time video. In: Proceedings of the DARPA Image Understanding Workshop (IUW'98), Monterey, USA. (1998) 129–136
12. Sidenbladh, H., Black, M.J., Fleet, D.J.: Stochastic tracking of 3D human figures using 2D image motion. In Vernon, D., ed.: 6th European Conference on Computer Vision (ECCV 2000), Dublin, Ireland, Springer Verlag (2000) 702–718
13. Khan, S., Javed, O., Rasheed, Z., Shah, M.: Human tracking in multiple cameras. In: Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV 2001), Vancouver, Canada, July 9–12, 2001. (2001) 331–336
14. Cupillard, F., Brémond, F., Thonnat, M.: Group behavior recognition with multiple cameras. In: Proceedings of the 6th IEEE Workshop on Applications of Computer Vision (WACV'02), Orlando, USA. (2002) 177–183
15. Yin, J.H., Velastin, S.A., Davies, A.C.: Measurement of crowd density using image processing. In Holt, M.J.J., Cowan, C.F.N., Grant, P.M., Sandham, W.A., eds.: Proceedings of the VII. European Signal Processing Conference (EUSIPCO-94), Edinburgh, UK. Volume III. (1994) 1397–1400
16. Siebel, N.T., Maybank, S.: Fusion of multiple tracking algorithms for robust people tracking. In Heyden, A., Sparr, G., Nielsen, M., Johansen, P., eds.: Proceedings of the 7th European Conference on Computer Vision (ECCV 2002), København, Denmark. Volume IV. (2002) 373–387