Vision Processing in Intelligent CCTV for Mass Transport Security

Abbas Bigdeli[1], Brian C. Lovell[1, 2], Conrad Sanderson[1], Ting Shan[1] and Shaokang Chen[1]
[1]Safeguarding Australia program
NICTA, 300 Adelaide Street, Brisbane
QLD 4000, Australia
[2]ITEE, University of Queensland, Brisbane
QLD 4072, Australia

Abstract- **Intelligent Surveillance Systems is attracting unprecedented attention from research and industry. In this paper, we describe a real-life trial system where various video analytic systems are used to detect events and objects of interests in a mass transport environment. The system configuration and architecture of this system is presented. In addition to implementation and scalability challenges, we discuss issues related to on-going trials in public spaces incorporating existing surveillance hardware.**

I. INTRODUCTION

CCTV surveillance systems are now seen as a major tool for counterterrorism activities after they were shown to be able to successfully track the movements of the four suicide bombers in the days before their attack on the London Underground in July 2005. Despite their usefulness, most current surveillance systems only provide *reactive* security by enabling the analysis of events after the terrorist attack has already occurred — what is really needed by the counterterrorism community is *proactive* security to help prevent future attacks. Intelligent Closed-Circuit TV (ICCTV) systems use powerful computers to analyze video feeds to assist human operators to detect events of interest as they occur — an example might be recognizing the face of a suspected terrorist or person of interest in a crowded railway station. In 2006 NICTA was awarded a research grant to conduct long term trials of advanced ICCTV technologies in important and sensitive public spaces such as major ports and railway stations. The trial will highlight operational and capability deficiencies in current ICCTV systems and will focus NICTA's research on capability gaps. The project is thus a unique collaboration of researchers, vendors, and user agencies aimed at delivering advances in computer vision and pattern recognition for human activity recognition.

A primary objective of this work is the field trialing and on-going development of a system for the robust detection and identification of persons of interest in a crowd. These people will often have non-frontal facial presentation, be photographed under various lighting conditions, and will exhibit natural expressions — such images are typically acquired from CCTV cameras in public spaces as the subjects are not usually aware of camera placement. Other capabilities that are being trialed and developed include 1) robust detection of background changes, 2) tracking and identification of people by their appearance across multiple cameras, 3) detecting suspicious events such as left luggage or the dangerous behavior of people, and 4) video summarization to produce brief video summaries of activity.

In this paper, the configuration of the trial system and some early results from commercial and NICTA research systems is presented. We also discuss the implementation and scalability challenges, as well as issues related to on-going real life trials in public spaces using existing surveillance hardware. The main capabilities that are currently offered by leading Intelligent Surveillance software vendors are demonstrated. Technology gaps are identified and opportunities for computer vision and pattern recognition research in the field of ICCTV are discussed.

II. RELATED WORK

After the 9/11 attack on the World Trade Center in 2001, there has been a significant rush in both the industry and the research community to develop advanced surveillance systems. In particular, developing total solutions for protecting critical infrastructure has been on the forefront of R&D activities in this field[1]. The solution is far much more complex than only video analysis. It must cover activity detection through to control room decision making.

As far as Video Analysis is concerned, today there are a number of products already available in the market from a variety of vendors. Some of the most noticeable ones are iOmniscient, NICE Systems, iSentry and Clarity. In the research arena, one of the closest works to ours was reported by Velastin et al[2] in a project called PRISMATICA. This project was part of a major European initiative on intelligent transport systems. They describe an architecture that takes into account the distributed nature of the detection processes and the need to allow for different types of devices and actuators. The main contribution was a computer vision module used in the system and its particular ability to detect situations of interest in busy conditions. They report that the system components have been implemented, integrated, and tested in real metropolitan railway environments. This is clearly a first step towards providing ambient intelligence in such complex scenarios. In [3] the same group presented some of the key computer vision algorithms that were employed in PRISMATICA and also published a survey on the current state-of-the-art in the development of automated visual surveillance systems [4]. This was quite useful in providing researchers in the field with a summary of progress achieved to date and to identify areas where further research is needed. In this survey, the authors have examined a wide range of capabilities such as the ability to recognize objects and humans — Describing human actions and interactions from information acquired by sensors is essential for automated visual surveillance. However the emphasis of their review was on discussion of the creation of intelligent distributed automated surveillance systems.

Video analysis covers a wide range of applications such as tracking [5], pedestrian detection [6], face recognition[7], as well as more complex detections such as events of interest as reported in [8].

In addition to the video analysis that creates the initial alarms, an immersive 3D visual assessment [1] can be employed for situational awareness and to manage the reaction process. This can then be coupled with wide area command and control capabilities to allow control from a remote location. The Praetorian suit of software packages provides such an environment. The open architecture of Praetorian in essence works as an operating system that can absorb alerts generated from various video analytic systems.

III. System Configuration

Figure 1 shows the basic configuration of one of the real life trial systems. As the existing cameras were all analogue, a video switch was used to digitize the feed and recorded footage from all camera was stored on a digital video recorder. The recorded footage on the DVR can be accessed by the video analytic software. As mentioned earlier, a number of commercially available video analytic software systems were used in this exercise. We also used some of the experimental systems that are being developed at NICTA labs. These include Robust Face Recognition as well as appearance-based tracking of individuals. The 3D Immersive Video Presentation layer uses the followings sources of information: original video feeds, the pre-produced 3D model of the premises, and the generated alarms as they are detected. The output of the Presentation Model is what is displayed on an operator's screen. This is done to put the alarm in the context of the 3D environment — an essential step if responders are to react appropriately in real-time.

The software that creates the 3D Immersive Presentation is quite substantial and a quick overview of the main modules is presented below. Interested readers are referred to [1] for more details.

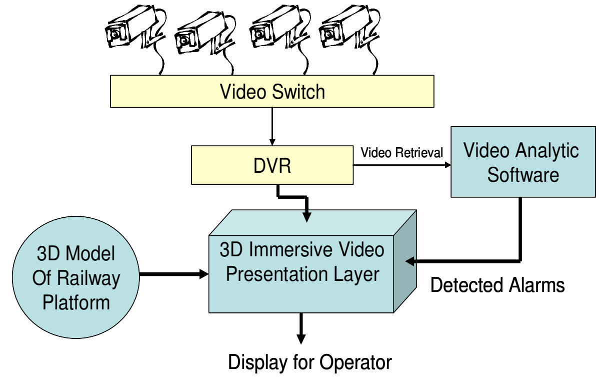

Figure 1. System Configuration

Camera Management Layer

The Camera Management Layer ingests the video feeds from all the cameras connected to the system. This layer is aware of the GIS correct location of each camera with respect to the 3D model. This is needed for correct registration of the camera view on top of the 3D model. The camera management layer also handles the stitching and overlay of the overlapping cameras. Where a PTZ camera is used, this layer manages the camera control as the operator, for example, follows a moving object from one camera to the next.

Alarm Management and Presentation Layer

The second major module is the Alarm Management and Presentation layer. At this layer, various alarms that are generated by one or more of the video analytic software systems are handled. The alarms are overlaid on the 3D model. An incident can be indicated by simply jumping to the camera view that has created the alarm.

IV. Field Trials

In our trials, we used four leading intelligent video surveillance systems to analyze the video sequences obtained from the railway platform. Table 1 shows the various alarms that can be generated from the four systems. It should be noted that certain specialized capabilities are only available in some and not all systems.

TABLE I: Existing and desired Alarms available in Some Intelligent surveillance systems

Common Capabilities	Specialized Capabilities
Left Object Detection	Association of object with owner
Human Tracking	Appearance Model based Tracking (tracking individuals)
Tripwire	
Slip and Fall Detection	
Behavior Analysis	

One of the "test-beds" we are using for our field trials is a railway station in Brisbane (Australia), which provides us with implementation and installation issues that can be expected to arise in similar mass-transport facilities.

Capturing the feed in a real-world situation can be problematic, as there should be no disruption in operational capability of existing security systems. The optimal approach would be to simply use an IP camera feeds. However, in many existing surveillance systems the cameras are analogue and often their streams are fed to relatively old digital recording equipment. Limitations of such systems can include low resolution, recording only a few frames per second, non-uniform time delay between frames, and proprietary codecs. To avoid disruption while at the same time obtaining video streams which are more appropriate for an intelligent surveillance system, it is useful to tap directly into the analogue video feeds and process them via dedicated analogue-to-digital video matrix switches.

Apart from the technical challenges, issues in many other domains may also arise. Privacy laws or policies at the national, state, municipal or organizational level may prevent surveillance footage being used for research even if the video is already being used for security monitoring — the primary purpose of the data collection is the main issue here. Moreover, without careful consultation and/or explanation, privacy groups as well as the general public can become uncomfortable with security research. Some people may simply wish not to be recorded as they have no desire in having photos or videos of themselves being viewable by other people. Plaques and warning signs indicating that surveillance recordings are being gathered for research purposes may allow people to consciously avoid monitored areas, possibly invalidating results.

Research code is often written by scientists/engineers (not necessarily professional programmers) for the explicit purpose of evaluating new methods. While this is sufficient to obtain experimental results which can be published, there can be little incentive to keep the code in a maintainable state or to guarantee that the underlying algorithm implementation is actually stable. Furthermore, research code is often written in Matlab which requires a nontrivial conversion into a language such as C++ to allow faster processing and integration with other software (e.g. via an SDK). The conversion may be quite difficult if, for example, the experimental implementation relies on elaborate functions and toolkits included with Matlab.

Preliminary Observations

In our initial field trial for public dissemination, we recorded 20 minutes of footage from seven cameras on a single platform in a major metropolitan railway station. During the recording time, a range of scenarios were staged on the platform. These include:

- left objects (of various sizes and position at different locations)
- crossing yellow safety lines
- attempting to break the vending machines
- other normal as well as abnormal passenger behavior on the platform

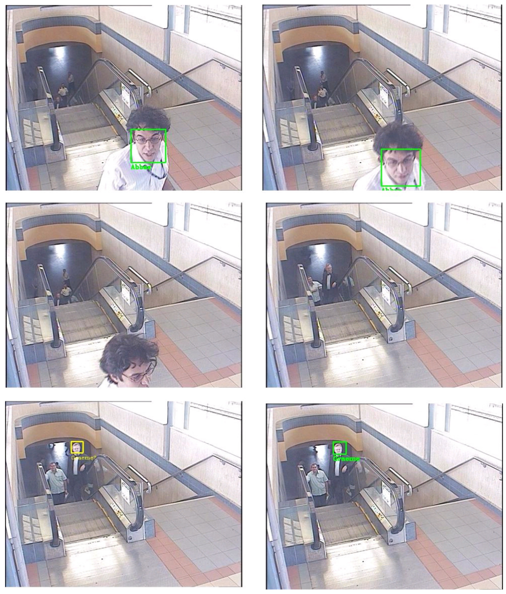

Figure 2. NICTA's real-time face detection and identification of persons of interest in footage from a CCTV camera in a major railway station

result image–15

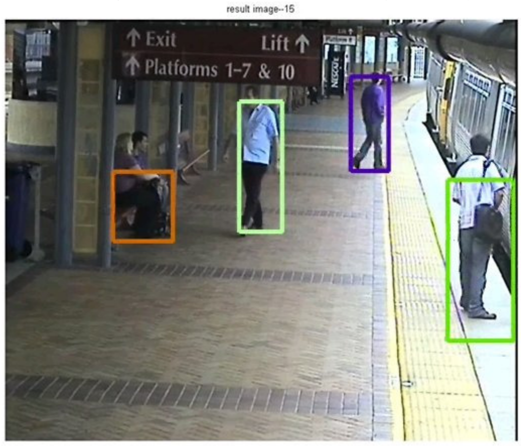

Figure 3. NICTA's tracking and labeling of commuters on a railway platform using CCTV cameras

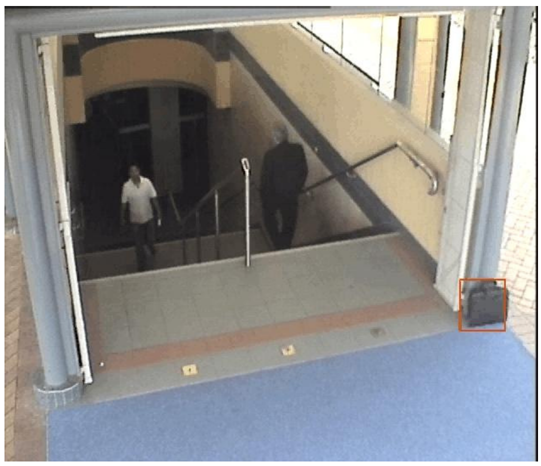

Figure 4. Commercial left object detection on a railway platform using CCTV cameras

The recorded footage was then fed through commercial as well as research video analytic software. Screen shots of some of the examples are given in Figures 2-4. Below we have summarized the initial observation from this exercise:

- Set up and calibration of commercial systems for different cameras and different type of alarms is time consuming and requires a fair bit of trial and error.
- The level of capabilities offered by various commercial software companies varies considerably.
- There are many areas where the research community can make significant contributions. NICTA is currently progressing towards meeting two of these areas: robust face recognition and human behavior analysis.

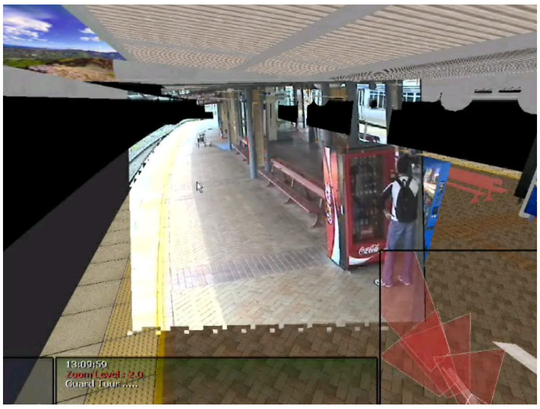

Figure 5. Immersive 3-D Visual Presentation of Camera View and 3-D model on a railway platform

V. Emerging Technologies

Smarter Cameras

Smart Cameras are slowly being introduced in the emerging surveillance systems. They usually perform a set of low-level image processing operations on the input frames at the sensor end. The types of computation performed on Smart Cameras vary between applications. In most cases, it is limited to compressing images and transmission to the host computer. In more recent years however, a wide range of applications have emerged. Algorithms such as face recognition are good examples of a Smart Camera application. Fatemi et al. introduced a real-time face recognition System on a Smart Camera [9] while Kleihost et al [10] took this concept further by developing a dedicated Smart Camera for face recognition. Ozer & Wolf [11] have implemented Human Activity recognition on a Smart Camera.

With growing interest in security and surveillance, Smart Cameras are finding their way into intelligent surveillance systems too. Matsushita et al [12] used Smart Camera for scene capturing and identity recognition. The majority of the video processing and analysis in the existing surveillance systems is executed at a central host using standard workstation racks. Up-and-coming embedded computing technology is being used to implement Smart Cameras with greater processing capabilities. Such capabilities include more robust face and number plate recognition, human activity recognition etc. The adoption of Smart cameras is resulting in a paradigm shift from a central to a distributed control surveillance system. The main motivation for this shift is increasing surveillance systems' functionality, availability, and autonomy.

VI. Conclusion

Intelligent Surveillance Systems is attracting much attention from research and industry. In this paper, we described a real-life trial system located on a public railway platform. We applied various video analytic systems both from commercial vendors as well as those under development in the research community to detect events and objects of interests. Some of the challenges involved in implementation, integration and system configuration were examined.

Acknowledgments

This project is supported by a grant from the Australian Government Department of the Prime Minister and Cabinet. NICTA is funded by the Australian Government's Backing Australia's Ability initiative, in part through the Australian Research Council.

References

[1] Francisco, G., Roberts, S., Hanna, K. and Heubusch, J., "Critical infrastructure security confidence through automated thermal imaging," *In Proceedings SPIE Vol. 6206, Infrared Technology and Applications XXXII*, May 2006.

[2] Velastin, S.A., Boghossian, B.A., Lo B.P.L., Jie Sun, Vicencio-Silva, M.A., "PRISMATICA: toward ambient intelligence in public transport environments," *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, Vol 35 (1), Jan. 2005, pp. 164- 182.

[3] Black, J., Velastin, S., Boghossian, B., "A real time surveillance system for metropolitan railways," *Advanced Video and Signal Based Surveillance, IEEE Conference on*, 15-16 Sept. 2005, pp. 189- 194.

[4] Valera, M. and Velastin, S.A., "Intelligent distributed surveillance systems: a review," *Vision, Image and Signal Processing, IEE Proceedings*, Vol 152(2), 2005, pp.192- 204.

[5] Fuentes, L.M. and Velastin, S.A., "From tracking to advanced surveillance," *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, 2003, Vol 3, pp. III- 121-4.

[6] Coimbra, M. and Davies, M., "Segmentation of moving pedestrians within the compressed domain," *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, Vol 3, 2004, pp. iii- 605-8.

[7] Shan, T., Lovell, B.C., Chen, S. and Bigdeli, A., "Reliable Face Recognition for Intelligent CCTV," *In Proceedings of RNSA Security Technology Conference*, Canberra 2006, pp. 356-364.

[8] Ziliani, F., Velastin, S., Porikli, F., Marcenaro, L., Kelliher, T., Cavallaro, A. and Bruneaut, P., "Performance evaluation of event detection solutions: the CREDS experience" *Advanced Video and Signal Based Surveillance. IEEE Conference on*, 2005, pp. 201- 206.

[9] Fatemi, H., Kleihorst, R., Corporall, H., and Jonker, P., "Real-time Face Recognition on Smart Camera," *Proceedings of Advance Concepts for Intelligent Vision Systems,* Belgium, September 2003, pp. 222-227.

[10] Kleihorst, R., Reuvers, M., Krose, B. and Broers, H., "A Smart Camera For Face Recognition," *Proceedings of International Conference on Image Processing (ICIP 2004)*, 2004, pp. 2849-2852.

[11] Ozer, B. and Wolf, W, "A smart camera for real-time human activity recognition," *In Proceedings, SIPS-01, IEEE.* 2001, pp. 217-225.

[12] Matsushita, N., Hihara, D., Ushiro, T., Yoshimura, S., Remkimito, J. and Yamamoto, Y, "ID CAM: A Smart Camera for Scene Capturing and ID Recognition," *In proceedings of the Second IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '03).* October 2003, pp. 227- 236.