

Advanced Mixed Reality Technologies for Surveillance and Risk Prevention Applications

Daniel Thalmann¹, Patrick Salamin¹, Renaud Ott¹,
Mario Gutiérrez², and Frédéric Vexo¹

¹ EPFL, Virtual Reality Laboratory,
Station 14, 1015 Lausanne, Switzerland
daniel.thalmann@epfl.ch

² ITESM Campus Toluca,
Edo. De México CP, 50110

Key words: Virtual Reality, Surveillance, Security, Teleoperation, Haptic Interfaces, Virtual Environments, Eye-tracking, Handheld Devices, Collaborative Environments

Abstract. We present a system that exploits advanced Mixed and Virtual Reality technologies to create a surveillance and security system that could be also extended to define emergency prevention plans in crowded environments. Surveillance cameras are carried by a mini Blimp which is tele-operated using an innovative Virtual Reality interface with haptic feedback. An interactive control room (CAVE) receives multiple video streams from airborne and fixed cameras. Eye tracking technology allows for turning the user's gaze into the main interaction mechanism; the user in charge can examine, zoom and select specific views by looking at them. Video streams selected at the control room can be redirected to agents equipped with a PDA. On-field agents can examine the video sent by the control center and locate the actual position of the airborne cameras in a GPS-driven map. The aerial video would be augmented with real-time 3D crowd to create more realist risk and emergency prevention plans. The prototype we present shows the added value of integrating AR/VR technologies into a complex application and opens up several research directions in the areas of tele-operation, Multimodal Interfaces, simulation, risk and emergency prevention plans, etc.

1 Introduction

Information technology (IT) plays an important role in security and surveillance initiatives such as security. Security is a major concern for governments worldwide, which must protect their populations and the critical infrastructures that support them [1]. Together with IT, Virtual Reality offers a promising future as key component of surveillance, security systems and emergency plan preparation.

The approach to security that we follow in our work is based on video surveillance. Our system is based on mobile cameras that can be directed to a particular location while keeping an overview of the surveyed site. Airborne cameras are the most flexible solution for a mobile surveillance system.

The prototype we present in this paper shows that efficient cost-effective surveillance systems based on VR technologies can operate in complex environments with

relatively low requirements in terms of equipment and personnel. Moreover, by using Augmented Reality techniques, we can extend this technology to training and emergency plan preparation.

2 Surveillance and Security Systems

A general surveillance and security system is composed of three main parts: *data acquisition, information analysis, on-field operation*. Any surveillance system requires means to monitor the environment and acquire data in the form of video, still images, audio, etc. The current state of our research focuses more on enhancing the interaction infrastructure of the three main parts composing a command and control surveillance system.

Following the *Command and Control* notion, we have designed an surveillance and security system composed of three main parts, covering the basic functions of a general surveillance system presented before: *data acquisition, information analysis, on-field operation*.

2.1 Data acquisition

Data acquisition is performed by means of a set of video cameras. Several kind of cameras can be distinguished: fixed, orientable and mobile [2],[3]. Fixed cameras are an efficient alternative for outdoors use for surveying car traffic or crowds in public sites. They allow for focusing the attention on a particular point. However, there are circumstances under which there is no possibility to fix a camera in advance, due to cost restrictions or lack of appropriate locations for wide visibility. In this case it is necessary to use mobile cameras that are usually airborne.



Fig. 1. left : a UAV; Center: an orientable camera; right: A helicopter Wescam.

In urban environments, blimps are the safest aerial devices for many reasons: they are light, easy to operate, and they fall slowly in case of problem, minimizing the risk of injuring people. We decided to base our system on a teleoperated mini-blimp and focused on implementing an intuitive and flexible interface that could take advantage of VR technologies.

2.2 Information analysis

Information analysis is the central part of a surveillance system. In order to provide an appropriate response to a given incident within reasonable timing, all the information about the whole situation, needs to be gathered in one unique place. A control room for surveillance is composed, in most cases, by a large video wall and multiple screens displaying views from surveillance cameras, for a proper interpretation of the situation. A set of buttons and joysticks are used to select, move and setup appropriate views.



Fig. 2. Various control rooms.

We decided to use VR devices for improving the ergonomics of existing systems. Visualization systems for video surveillance based on an Augmented Virtual Environment (AVE) are an important topic nowadays. AVE fuses dynamic imagery with 3D models in a real-time display to help observers comprehend multiple streams of temporal data and imagery from arbitrary views of the scene [4]. We provide a novel interface based on eye-tracking technologies which minimizes the use of keyboards and other interaction devices. The commander is immersed in a CAVE which displays live video enhanced with graphic overlays.

2.3 On-field operation

On-field operation is the result of decisions taken at the control center and require a team of surveillance agents to take action for controlling the situation on the ground. Common communication devices include: pagers, in-vehicle systems, radios and headsets; etc. Recent security studies and initiatives have pointed out the importance of permanent multimodal communication [5][6].

A common problem with handheld interfaces is that they usually present a scaled-down desktop-based interface composed of menus and buttons which are difficult to use on a small display. On-field agents require to concentrate their attention of the site, not on the interface. There is a need for a dynamic handheld interface that provides multimedia information and minimizes the use of menus and buttons.

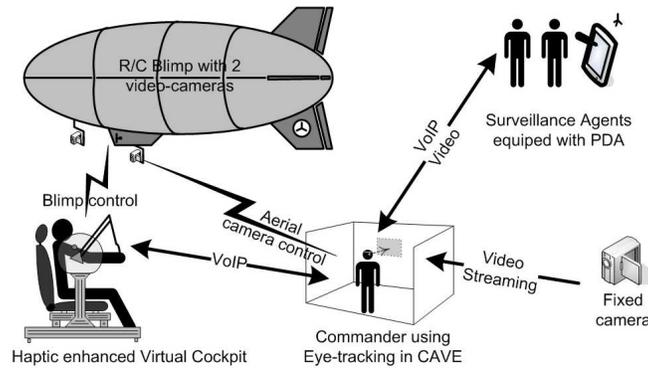


Fig. 3. The overall system architecture.

3 System Architecture

This section describes the overall surveillance, security and training system that we have created. We can distinguish three main parts (figure 3) :

- Control of the aerial device (the Blimp) supporting the video cameras. This task is done by a single pilot seating on Haptic WorkstationTM inside a distant and closed environment. The pilot can control the blimp as if he were inside it.
- On-field agents : They are equipped with handheld devices in order to receive precise orders including multimedia content (text, selected images of a video stream, sound). They also may have a direct control on the cameras to better evaluate the situation.
- Coordinating on-field agents and blimp pilot. A commander communicates with the pilot and every agent and gives them spoken orders. He has also a real-time view of all the cameras (mobile and fixed), and can also send multimedia content to on-field agents.

3.1 R/C Blimp

Our blimp, as shown on figure 4, is a low-cost Unmanned Aerial Vehicle (UAV) that we use in our teleoperation research. The *R/C Blimp* is composed by a 6,75m long and 2,20m diameter. The blimp could fly up to 35km/h. The range of the transmission of the radio controller is 1.5km.

It is carrying two video-cameras and their transmission systems. The first camera can be controlled by the pilot using head movements. The other camera is used by the *Surveillance Control Room*. Thus it must have a good quality, allow for recording and zooming: we have chosen a Panasonic mini-DV TriCCD camera. We use the analogical output of these cameras with two systems of video transmission. The *R/C Blimp* is also equipped with a STXe GPS system, provided by GPS-Flight. It is used to give location,



Fig. 4. Photo of the R/C blimp.

speed and altitude information. Finally, the actuators are controlled by a 40MHz 7-channels Futaba FX-18 and Graupner MC-10 radio controllers. We have used two USB SC-8000 Servo Controllers that allow a pc to control a radio controller.

3.2 Virtual Cockpit

The *R/C blimp* is not so easy to pilot, even with the remote controller. Moreover, the *Virtual Cockpit* is in an isolated room without any direct-view of the *R/C Blimp*. Therefore the interface must be **precise** (for a fine control), **instructive** (to give location information) and **intuitive** (to avoid manipulation errors).

The visual part of the blimp is rendered to the pilot via an Head-Mounted Display (HMD). In order to have a virtual camera that moves according to the head movements, we have used an orientation tracker. The user is virtually seated inside a 3D gondola made of glasses. Finally, the GPS information transmitted to the pilot is overlaid on the window. Concerning haptics, we have used an Immersion Haptic WorkstationTM (see the photo on figure 5).

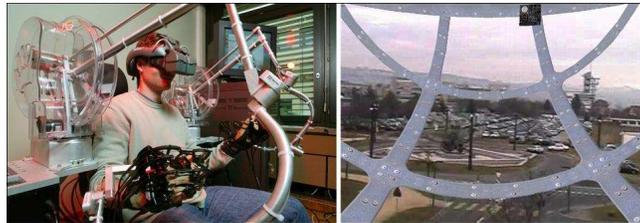


Fig. 5. The Blimp's Virtual Cockpit.

3.3 Surveillance Control Room

In order to place the supervisor in the best disposition for taking decisions, our system displays several video streams in a CAVE. These allows the supervisor to select and send visual information to the *On-field Agents* intuitively and instantaneously.

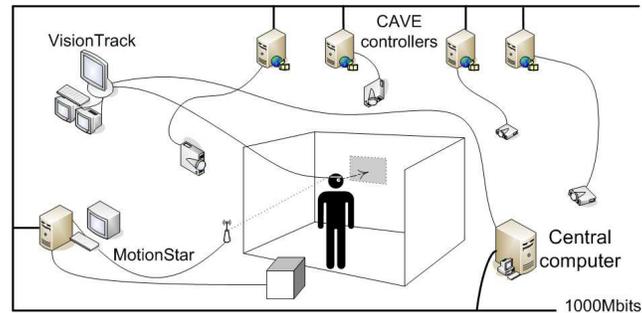


Fig. 6. The control room system.

The CAVE offers the possibility to display multiple video streams on four projection screens (see figure 7), providing full immersion of the user into the scene. The system is composed by four video projectors and four associated computers for recovering the video streams, displaying an eye picking area and sending the selected part of the image to the *On-field Agents*. The video stream of the distant cameras is transmitted and rendered via the local area network using RTP.

A joystick allows the supervisor to move mobile cameras to get the most appropriate point of view (see figure7). By pressing a single button, the user validates his eye selection and sends it to the *On-field Agents*. The picture is sent via internet using virtual private network for security purposes. Passing the right information at the right moment is crucial. Since our camera is aerial, it gets a field of view covering a large area. In contrast, agents on the ground get a local view of the situation. It's important for them to be able to relate the received image to their field of action. For these reasons we needed, in addition to the zoom, to be able to send a cropped part of the full image. Notice that the on-field agents may sometimes require to move by themselves the cameras to better appreciate the situation. In these cases, the supervisor can gave them at any moment the possibility to control the cameras.

In order to select part of the image, the eye tracking system is used to compute the gaze direction and determine the gaze target on the cave. The vision system follows the eye movements and measures the position of the pupils. This system provides the supervisor with powerful tools to send visual and vocal information to *On-field Agents*.

3.4 On-field Agent Equipment

Our handheld communication equipment (PDA) provided a dynamic interface that reacts to the way it was held by the agent. We then had to hold it vertically or horizontally



Fig. 7. The four sided CAVE: Advance Control Room.

depending on the information we wanted. But we finally decided to only use a simple and intuitive interface which the user gets quickly comfortable with.

At first, the user can see the images of every active cameras (four cameras at the maximum for the moment depending on our acquisition card) on his/her PDA screen, as shown hereinafter in figure 8. These pictures are selected at the control room as well as a map of the surveyed site and updated every time there are changes on the picture (i.e. something happened in the field of view of the camera). GPS information is used to point out on the map the current position of the blimp. When held horizontally (figure 8), the interface shows a higher resolution view of the live video stream. This way, *On-field Agents* can better appreciate specific details while maintaining communication with the control room.

On the left of the screen lies a panel with a few buttons the user can push to see one of the pictures on full screen or to see all of them together. Finally, some buttons can be used to control the current shown camera rotation and zoom. The on-field agent can then apply the following basic operations: rotation to the left, right, up, down and zoom in and out. The use of these last buttons generate messages which are sent to the server what controls the camera servos embed on the blimp.

3.5 Modules Intercommunication

Video transport over internet in real-time is a complex problem that has often been addressed. Today, RTP [7], the Real-Time Transport Protocol, fulfilled most of the re-

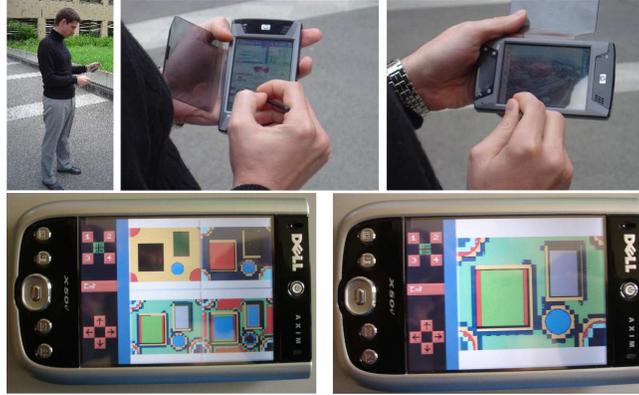


Fig. 8. On-field agent communication equipment.

quirements necessary to this purpose, like asynchrony solving capacity, server/client traffic regulation or multiplexing of different kind of streams [8]. The streamed images are compressed using Motion JPEG [9], which provides a good compromise between image quality and size.

Our system uses a Voice Over IP communication between the *Virtual Cockpit* and the *Surveillance Control Room* and between the *Surveillance Control Room* and the *On-field Agents*. We use the freely available PC and PocketPC versions of Skype.

Moreover, as we presented it before, the PDAs of the on-field agents are also connected to the network and get jpeg pictures of the video streamings shown on the full screen or together. They also can control the camera direction and zoom when they are chosen by the supervisor.

4 Discussion and Results

After several trial sessions involving each of the components taken separately, we have put the whole system together and did some tests in the site of our university.

4.1 Virtual Cockpit

The use of a responsive head orientation tracker to drive the video camera of a tele-operated vehicle is an efficient method. It allows indeed to have mixed reality video stream coming from the *R/C Blimp* that is well positioned into the virtual 3D cockpit. This is also an intuitive way to move such a camera because it reproduces exactly what happens in the reality.

We have observed that the Haptic WorkstationTM is precise enough to offer a fine control on the blimp's actuators. As further work we plan to give extra force feedback to the pilot in order to improve the information that he gets, enhancing the multimodal aspect of the interface.

4.2 Surveillance Control Room

Standing into the CAVE gives you an amazing effect of immersion. You are literally feeling inside the scene. Using large display for aerial video surveillance, offers great benefits by balancing the effects of having a large field of view to detect small details. By surrounding the user with video information, We added orientation information by grouping the display of cameras according to their position. For example, we projected the aerial video from the blimp on the floor, producing the sensation of flying above the scene. The picking, done with the eye tracking system, offers an intuitive and efficient way to select and crop part of the selected image. The target follows smoothly the gaze and offers a visual feedback to the user on the selected zone.

4.3 On-field Agent

We have tested our multimodal communication system with on-field agents. Interface proved to be easy to use and intuitive enough. We thought that minimizing the need for pressing buttons on the handheld device would make the use of this application easier but it almost does not change anything. Moreover, we always had to get the PDA inclination. Finally, we improved our interface Actions from the user are reduced to launching skype and loading the main web page on the PDA. Having a GPS driven map pointing out the position of the airborne camera (blimp) revealed to be very useful. Agents were able to have active interaction with the control room thanks to the visual feedback and the transparent VoIP communication.

Demand is growing among security professionals for systems that transmit video images to hand-held devices. If they have wireless access, security agents don't need to return to the control room to view an image before responding to an incident or rely on a dispatcher's description of a suspect, "A second-hand description of somebody is not necessarily the best" [10]. We believe our system responds well to current needs for mobility and intercommunication. We indeed made some experiments with on-field agents. This helped us to mainly improve the interface of the application on the PDA. We made it more intuitive and comfortable for the user. Moreover, we noticed during these experiments that the possibility for the agent to be able to control the camera is mandatory.

5 Potential Application Risk Prevention and Emergency Training

Based on aerial images, it could be nice to augment them with virtual elements such as crowds. This could be very helpful for the on-field agents training because we could simulate situations at risk which would lead to emergency ones. It would then be easier to manage the crowds for the agents if they know where the problem is and where they can redirect them.

Besides, it would be all the more interesting to be able to coach the agents to these situations. We think that if we could make these training directly on the manifestation terrain, this would improve the agents reactions and they could better evaluate and prevent unsecure and critical situations. As you can see in [11], our laboratory is currently



Fig. 9. Example of mixed reality picture combining virtual crowds with a real aerial picture.

working on a crowds engine. We model the crowd behavior, and how people interact each other while walking or running in several different directions. This could then be very interesting for the agents training to work with mixed reality, e.g. virtual crowds in the real environment to manage.

6 Conclusion

We have described a full surveillance, risk prevention and emergency system based on advanced Virtual Reality technologies. It also can help to prevent some risks by simulating with augmented reality some dangerous situations, e.g. crowds behavior in emergency. The system architecture is divided into three main components: data acquisition, analysis and operation. Our contribution focuses on the successful application of state of the art Virtual Reality and multimedia technologies in a complex system with multiple participants. Our main goal was to maximize interaction and communication between the personnel implied in the tasks. A single user can teleoperate the airship used for data acquisition, while a second one can concentrate on analyzing the video streams. We put in practice the use of haptic interfaces for teleoperating a flying vehicle. Future work consist in taking full advantage of the new dimension provided by haptic feedback for conveying additional information to the pilot and ease the piloting task. In terms of information analysis which take place at the control room, we have obtained satisfactory results with the use of eye tracking technology within a CAVE system. We proposed an innovative interface for picking-up zones of interest from live

video stream. The system is complemented with an efficient multimodal communication device based on a PDA. This solves a common demand of security agents who require more than just voice communication with the control room. This kind of VR and multimedia technology applications fit into the context of current initiatives for security and enhanced surveillance systems. Finally, it would be very interesting to use it, within the framework of augmented reality with our crowds engine, for the on-field agents training.

Acknowledgement

We wish to thank Sylvain Cardin and Achille Peternier for their useful hardware and software contribution to the system presented in this article.

References

1. Reiter, M., Rohatgi, P.: Homeland security. *IEEE Internet Computing* **8**(6) (2004) 16–17
2. Ellis, T.: Multi-camera video surveillance. In: *Proceedings of the 36th Annual International Carnahan Conference on Security Technology*. (2002) 228–233
3. Kumar, R., Sawhney, H., Samarasekera, S., Hsu, S., Hai, T., Yanlin, G., Hanna, K., Pope, A., Wildes, R., Hirvonen, D., Hansen, M., Burt, P.: Aerial video surveillance and exploitation. *Proceedings of the IEEE* (10) (2001) 1518–1539
4. Sebe, I.O., Hu, J., You, S., Neumann, U.: 3d video surveillance with augmented virtual environments. In: *Proceedings of the First ACM SIGMM international workshop on Video surveillance, IWVS'03, New York, USA, ACM Press* (2003) 107–112
5. Yuan, Y., Detlor, B.: Intelligent mobile crisis response systems. *Commun. ACM* (2) (2005) 95–98
6. Sawyer, S., Tapia, A., Pesheck, L., Davenport, J.: Mobility and the first responder. *Communications of the ACM* (3) (2004) 62–65
7. Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V.: Rtp: A transport protocol for real-time applications. In: *Request for Comment RFC-1889 of the Internet Engineering Task Force, IETF*. (1996)
8. Benslimane, A.: Real-time multimedia services over internet. In: *Proceedings of the 1st European Conference on Universal Multiservice Networks, ECUMN'00*. (2000) 253–261
9. G.K. Wallace: The jpeg still picture compression standard. *IEEE Transactions on Consumer Electronics* (1992) xviii–xxxiv
10. Douglas, M.: Surveillance in the palm of your hand. *Mobile Radio Technology Magazine (MRT)*. http://mrtmag.com/mag/radio_surveillance_palm_hand/ (2004)
11. Pettre, J., de Heras, P., Main, J., Yersin, B., Laumaond, J., Thalmann, D.: Real-time navigating crowds: Scalable simulation and rendering. In: *Computer Animation and Virtual Worlds*. Volume 16. (2006) 445–456