

# Advanced Virtual Reality Technologies for Surveillance and Security Applications

Renaud Ott\*  
EPFL VRlab

Mario Gutiérrez†  
ITESM Campus Toluca

Daniel Thalmann‡  
EPFL VRlab

Frédéric Vexo§  
EPFL VRlab

## Abstract

We present a system that exploits advanced Virtual Reality technologies to create a surveillance and security system. 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 PDA interface reacts to the user's gestures. A tilt sensor recognizes the position in which the PDA is held and adapts the interface accordingly. The prototype we present shows the added value of integrating VR technologies into a complex application and opens up several research directions in the areas of tele-operation, Multimodal Interfaces, 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 [Reiter and Rohatgi 2004]. Together with IT, Virtual Reality offers a promising future as key component of surveillance and security systems. VR can be used in a number of tasks: teleoperation of the actual data acquisition systems (cameras, vehicles, etc); providing multimodal interfaces for control rooms where information is analyzed; and finally, empowering on-field agents with multimedia information to ease their tasks of localizing problematic zones, etc.

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. This approach has been successfully applied during public events with large audiences and high security demands such as the Athens 2004 Olympic Games [Airship Management Services ]. The Greek Government awarded a contract to US defence giant Science Applications International Cor-

poration (SAIC) for a Command, Control, Communications, Computer and Intelligence (C4I) infrastructure. A Skyship 600 Blimp (Global Skyship Industries, Inc.) was the central element for providing airborne surveillance over the Olympic sites. The airship was equipped with a highly sophisticated suite of sensors including infra-red cameras for day and night use, etc. With a security budget of more than USD\$ 2 billion, the hi-tech security systems deployed in Athens are still far from being automated and fully optimized. The referred blimp required a 20-person crew to be operational. In consequence, the system is error prone and expensive in terms of human resources.

We believe the basic principles of an airborne surveillance systems have proved their efficacy but there are still issues to be addressed concerning the interaction between the command, control and on-field personnel. Our goal is to enhance the communication channels and operation interfaces by means of VR technology.

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. One of our main contribution is the integration of different VR and multimedia technologies into a coherent surveillance system. The paper is organized as follows: first we present a short overview on surveillance and security systems; we continue by describing the system architecture of our prototype; the paper ends with a discussion on each of the system components and perspectives of future work.

## 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. Such data is to be processed and analyzed by a human, a computer or a combination of both at the command center (analysis). A commander can take the decision of performing an on-field operation to put the environment back into a situation considered as normal. On-field control operations are performed by on-field agents who require efficient communication channels to keep a close interaction with the command center.

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. As stated in the introduction, the system we present in this paper is based on airborne cameras for surveillance of large sites. Aerial monitoring of ground sites using video cameras is playing an increasingly important role in autonomous surveillance applications [Shekhar 2000].

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*. Next subsections present an overview of the state of the art of technologies applied in such areas.

\*e-mail: renaud.ott@epfl.ch

†e-mail: m.gutierrez@itesm.mx; Work done while being at EPFL

‡e-mail: daniel.thalmann@epfl.ch

§e-mail: frederic.vexo@epfl.ch

## 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 [Ellis 2002],[Kumar et al. 2001]. Fixed cameras are an efficient alternative for outdoors use for surveying car traffic or crowds in public sites. They can also be used indoors to survey critical zones such as restricted access areas in a building or office. Orientable cameras (see figure 1) have additional functionalities (zoom, 2DOF orientation) that can be controlled at distance by an operator or in automatic mode. They allow for focusing the attention on a particular point. For instance, for taking a closer look to a particular person in a crowded site such as a stadium. 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.



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

Airborne cameras (see figure 1) can be embedded in helicopters, recognition airplanes, blimps, etc. Flying vehicles can be either manned or unmanned. Unmanned flying vehicles have the advantage of a reduced size and increased performance. They can fly longer and through places that would be difficult or impossible to access with manned vehicles. 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 tele-operated mini-blimp and focused on implementing an intuitive and flexible interface that could take advantage of VR technologies.

There is an important amount of research concerning teleoperation [Fong and Thorpe 2001] [Bruce et al. 1998]. A common approach to teleoperation interfaces consists on implementing physical controls such as joysticks, steering wheels, handles, buttons, and so on. The problem with physical interfaces is that they are expensive to implement, and difficult to reconfigure to match different user requirements and/or applications. Virtual entities (3D models) can solve the problem of reconfiguration and adaptation, but also have some drawbacks. The main disadvantage of an interface based on 3D models is the absence of physical feedback. "Feeling" a control tool is essential, otherwise the manipulation requires too much effort and becomes unprecise. Haptic technologies aim at solving this problem by enabling virtual objects to provide a tangible feedback to the user [Diolaiti and Melchiorri 2002]. Moreover, virtual interfaces allow for implementing a variety of feedback mechanisms to ease the teleoperation, such as vibrating controls and audiovisual signals to inform the user about the vehicle status and the surrounding environment.

## 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.



Figure 2: Various control rooms (top-left: EADS, right: Oakland airport, bottom-left: Willow creek community).

One of the main centers of interest concerns treatment and rendering of various types of data (maps, graphics, text, photo, video, sounds...) through a convivial and ergonomic interface. The US army researches on C3I, study models and interfaces to manipulate and interact with the flow of information. Companies, like the System Design Center (DCS) from EADS, are specialized in building such systems. DCS has developed a multimodal platform, NetCOS, for visualizing, organizing, linking and simulating various types of input data. The complexity inherent to manipulate all the control panels for the application, requires generally several human operators. The supervisor uses mainly oral communication with the control room operators and on-field agents.

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 [Sebe et al. 2003]. 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 [Yuan and Dettlor 2005][Sawyer et al. 2004]. Surveillance agents require efficient means of communication with the commander. Interoperable communications are key to maximize efficacy of on-field agents.

Security personnel look to wireless video systems for critical incident information. Complementary information in the form of maps and live video streaming can help on locating the problematic zone and act faster and with better knowledge of the situation. In [Douglas 2004] the author presents an overview of companies and technologies providing systems that transmit video images to hand-held devices. The need for providing detailed real-time information to the surveillance agents has been already identified and is being addressed by the research community. In [Li et al. 2002] the

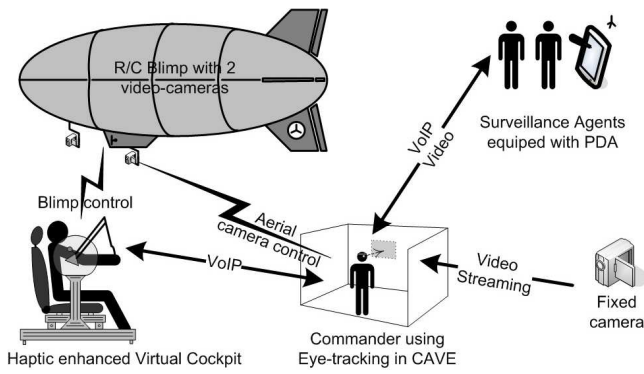


Figure 3: The overall system architecture.

authors present a system which delivers various surveillance services over the Internet for supporting mobile PDA users (*On-field Agents*). Another work which focuses more on video analysis but recognizes the advantages of using PDA in surveillance contexts is described in [Cucchiara et al. 2003].

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. Using this kind of mini-GUI requires too much time and attention for non experienced users. 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. Such an interface should be intuitive, responsive and present just the essential information at the right moment. In section 3.4 we present our proposal of a PDA-based multimodal communication device for on-field agents.

### 3 System Architecture

This section describes the overall surveillance system that we have created. The figure 3 shows the different modules that are used for surveillance. We can distinguish three main parts :

- Control of the aerial device (the Blimp) supporting the video cameras. This task is done by a single pilot seating on Haptic Workstation™ 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, images, sound).
- 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 envelope that is filled with 11m<sup>3</sup> of Helium gas (He). The total weight including its standard flight equipment is 9kg, so there is around 2kg of



Figure 4: Photo of the R/C blimp.

maximum payload for the cameras and the video transmission system. Below, there is a gondola containing the electronics part, the power supply (6600mAh allowing 1h at half-speed) and supporting the two electric motors. Each have 1,5kg of power, allowing the blimp to fly at 35km/h when there is no wind. The range of the transmission of the radio controller is 1.5km, but it can be extended with repeaters.

Figure 5 shows the five actuators controlling this *R/C Blimp*. Aerodynamic stabilizers are used only when the blimp has reached a certain speed (approximately 8km/h). In section 2.1, we have seen that the agility is essential when flying in urban environments. Thus, there is one extra-helix, located on the bottom stabilizer, that can run on both directions. It is used to turn left or right when the Blimp is flies at low speed. The pilot can also change the plane of the main helixes, in order to go up or down at low speed. Finally, there is an actuator for the speed which controls the rounds per minute of the helixes.

Part of the 2kg payload is for carrying two video-cameras and their transmission systems. The first camera can be controlled by the pilot using head movements. The orientation system must be faster than the head, and the field of view has to be large enough to permit a good control of the *R/C Blimp*. Thus, we chose a small (less than 50g) wide-angle CCD-camera. 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 (1.2kg). We use the analogical output of these cameras with two systems of video transmission. On the ground, we receive two PAL/B signals that are digitalized. Both are connected to a workstation located on the network that broadcasts the streams.

The *R/C Blimp* is also equipped with a STXe GPS system, provided by GPS-Flight. It is used to give location, speed and altitude information. This module contains a common GPS antenna but also a 2.4GHz transmission system which can transmit in real-time at 10Hz. The receptor of this transmission system is plugged via USB to a pc connected to our network.

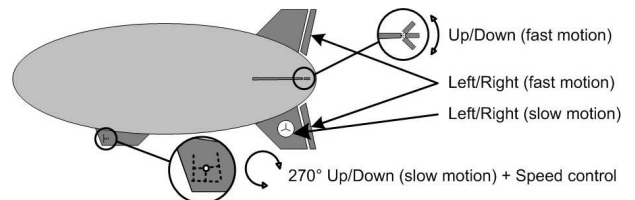


Figure 5: Actuators of the blimp: aerodynamic stabilizers, left/right motor, motors for changing helixes plane and speed.



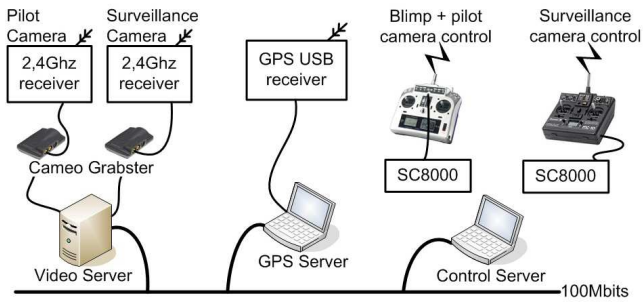


Figure 6: The Blimp communications system: three pc connected on a 100Mbps network for receive video and GPS and controlling servos.

Finally, the actuators are controlled by a 40MHz 7-channels Futaba FX-18 radio controller. On board, the reception system is completely duplicated to prevent failures. Each of the two airborne cameras can be oriented along the pitch and the yaw angle (two channels), and the Panasonic camera can also zoom (one channel). The wide-angle camera that is used by the pilot is also connected to the Futaba radio, so the pilot has to control seven channels. There is also a second radio controller, a Graupner MC-10, that is used by the *Control Room* to move the high-quality camera. We have used two USB SC-8000 Servo Controllers that allow a pc to control a radio controller.

To put in a nutshell, figure 6 describes the computers that are used to control and gather information from the blimp. All these equipments are located on the roof of our laboratory in order to communicate without interferences. All these communications are near real-time because everything is done in hardware: the video is delayed for less than 50ms, and the control of the blimp via the servo controller is in real-time. In the next subsection we will describe the *Virtual Cockpit* system used to control this Blimp.

### 3.2 Virtual Cockpit

In this subsection, we will describe how we have implemented the cockpit using VR devices. The *R/C blimp* is not so easy to pilot, even with the remote controller, which is usually the classic device for this purpose. 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). In the first part, we deal with the visual rendering of the cockpit, after we describe the haptics part of the interface. Both of them tend to achieve the three goals mentioned above.

The visual part of the blimp is rendered to the pilot via a Proview™ XL50 Head-Mounted Display (HMD). This HMD gives a 1024 × 768 resolution on both eyes, with an horizontal field of view of 40°. In order to have a virtual camera that moves according to the head movements, we have used an orientation tracker, the InertiaCube3, provided by Intersense. This tracker is very precise and has a refresh rate of 180Hz.

Software side, we are using real-time OpenGL viewer created in our laboratory. It allows for rendering stereo 3D models and videos at the same time by mapping streams on a polygon. Figure 7 shows the representation of the blimp inside the virtual environment. The user is seated inside a 3D gondola made of glasses. This video is mapped on a polygon with the stream coming from the pilot camera. Since the camera follows head movements and there is less than 5ms of

ping between the Control Server computer (see figure 6) and the Haptic Workstation™ server, there is a good matching between the head orientation and the real scene displayed. Finally, the GPS information transmitted to the pilot is overlaid on the window. We choose to not integrate it inside the 3D virtual environment in order to keep it always available to the pilot. This information is represented as a 2D map indicating the blimp's location, heading, altitude, and speed (horizontal and vertical) are displayed as well.

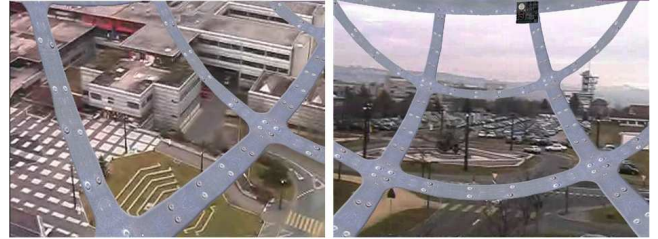


Figure 7: The Blimp's Virtual Cockpit.

Concerning haptics, we have used an Immersion Haptic Workstation™ (see the photo on figure 8). It is composed of DataGloves, 6DOF hand tracking and force-feedback on the wrists. In [Ott et al. 2005b], the authors have shown that the teleoperation of a vehicle using haptic devices is more efficient when having a gesture interface. Thus, we have chosen to use this technique: the right hand is used to control the aerodynamic stabilizer (by moving the hand forward/backward and left/right), whereas the left hand is used to control the engine power. However, when there is no power, a left/right hand movement doesn't move the stabilizers but controls the rear engine allowing the pilot to do fast about-turn, useful in urban environments. One disadvantage of the Haptic Workstation™ is that it is uncomfortable during long sessions, because the arm must be kept outstretched while supporting the weight of the exoskeleton. In [Ott et al. 2005a], the authors have presented a software for improving the comfort when using a haptic device by creating a zero-gravity illusion. We have used such system in order to compensate both the user arms and the exoskeleton. Moreover, the force feedback constraints the user hands to the neutral position, i.e. no power and straight ahead. When the user wants to move a hand, the force feedback intensity increases. The pilot can thus feel how fast he is going by evaluating the force applied on his hands. In the next subsection, we will study the *Surveillance Control Room* where orders to the pilot are given.

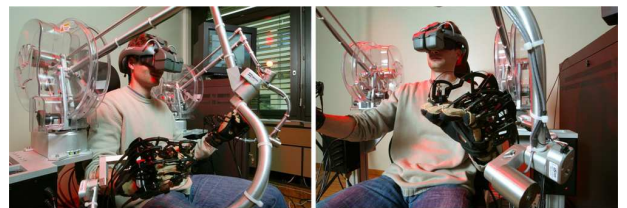


Figure 8: Haptic Workstation™ and HMD used for rendering the virtual cockpit.

### 3.3 Surveillance Control Room

In order to place the supervisor in the best disposition for taking decisions, we focused on designing an ergonomic interface, efficient and intuitive, using Virtual Reality devices. Our system displays

several video streams in a CAVE, and allows the supervisor to select and send visual information to the *On-field Agents* intuitively and instantaneously.

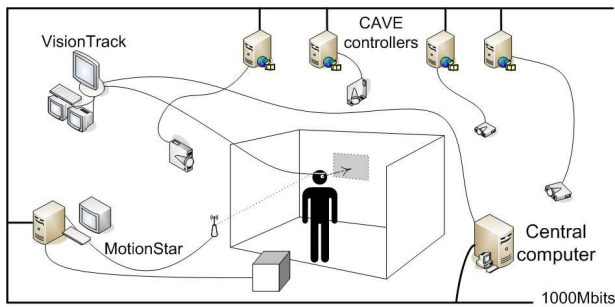


Figure 9: The control room system.

The CAVE offers the possibility to display multiple video streams on four projection screens (see figure 10), 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.

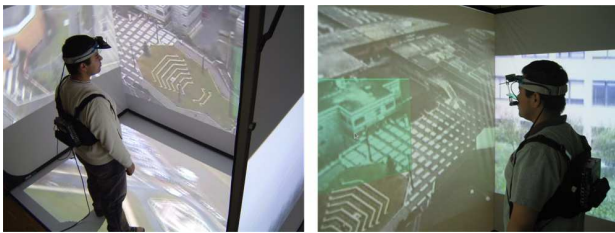


Figure 10: The four sided CAVE.

A joystick allows the supervisor to move mobile cameras to get the most appropriate point of view (see figure 11). Orientation is controlled by the pad and two buttons are used to zoom in and out. 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. We choose to send a single relevant picture rather than directly the video stream viewed by the supervisor for readability reasons and network resources management. 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.

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. This device VisionTrack™ (see figure 11), is composed by a head-mounted assembly with two cameras, a dichroic mirror and an infrared eye illuminator. The vision system follows the eye movements and measures the position of the pupils and the corneal reflection with some specialized video processing hardware. Based on those two positions, the system computes the gaze direction and sends it continuously to the central computer through the serial port. Since the user is able to move inside the cave, we cannot only rely on the eye orientation to determine his gaze target. For taking into



Figure 11: Top and left: The Eye-tracking system. Bottomright: The mounted rotating camera and the joystick controller.

account the orientation of the head, we added a 6DOF magnetic sensor, MotionStar™. The system sends position and orientation to the central computer via the local network.

The central computer gathers the information from the eye tracker and from the motion capture. The system corrects the inherent instability of the measurements by filtering the input data. The exact gaze vector is given by combining the orientation basis from the magnetic sensor and the deviation of the gaze from the head. Knowing the exact configuration of the CAVE, the system detects which plane is observed and the exact pixel looked at. Some threshold values and a time delay avoid the target from blinking between two planes on the singularities. For smoothing the movement of the selection zone, the cropping box is attached to the gaze target with a damped string algorithm. Its size is also fixed at  $320 \times 240$  pixels, the maximum resolution of the PDA screen.

This system provides the supervisor with powerful tools to send visual and vocal information to *On-field Agents*.

### 3.4 On-field Agent Equipment

Building upon the concepts and requirements outlined in the section 2.3, we have designed a handheld interface that minimizes the use of buttons and menus and responds to intuitive gestures from the user. Our handheld communication equipment (PDA) provides a dynamic interface that reacts to the way it is held by the agent.

The system is based on commercial PDA devices using the PocketPC operating system with Wireless LAN capabilities. By means of a tilt sensor, the PDA detects whether it is positioned horizontally or vertically and presents different information. Parts of the live video stream received at the control room could be sent on the handheld, as well as map of the site enhanced with GPS information coming from the aerial monitoring system (R/C blimp).

When held vertically, see figure 12, the interface shows a reduced



Figure 12: On-field agent communication equipment.

image of the live video stream selected at the control room as well as a map of the surveyed site. GPS information is used to point out on the map the current position of the blimp. When held horizontally (figure 12), 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.

The PDA application is loaded as a web page using the default Internet browser pre-installed on the PDA. One of the two web pages is automatically loaded according to the PDA position detected by the tilt sensor. An application running on the PDA reads the tilt sensor values and generates a web page that redirects the internet browser to the corresponding web page generated at the web server. A PC-based application receives information from the control room (selected region of the video stream) and data from the airborne GPS. It generates two web pages corresponding to each of the two modalities (horizontal, vertical). Web pages are published through a conventional Apache Web server. This way, the alternative data-display interfaces are accessible for handheld devices and any PC system connected to the Virtual Private Network of the site under surveillance.

### 3.5 Modules Intercommunication

Video transport over internet in real-time is a complex problem that has often been addressed. Today, RTP [Schulzrinne et al. 1996], the Real-Time Transport Protocol, fulfilled most of the requirements necessary to this purpose, like asynchrony solving capacity, server/client traffic regulation or multiplexing of different kind of streams [Benslimane 2000]. Many implementations of RTP are available, but we choose the one that works in the Sun Java Media Framework (JMF) for three reasons: it is free, it contains image compression utilities useful for reducing the bitrate of the stream (this is not in the RTP standard) and finally its easier to program with it than with other free libraries (we have also tested LiveMedia). On our 100Mbits ethernet network, we have realized two applications (client and server) that encode, transport and decode a PAL video stream (it means a resolution of  $720 \times 576$  24bits-pixels at 25Hz) with less than 50ms lag. The streamed images are compressed using Motion JPEG [G.K. Wallace 1992], which provides a good compromise between image quality and size. This processing has also been used for the fixed aerial cameras controlled by the *Surveillance Control Room*.

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. We selected this software for our prototype due to the following reasons: First of all, Skype usually provides better performance in terms of transmission quality than similar products such as Microsoft's MSN messenger. Moreover, current version of MSN messenger for PocketPC does not provide voice communication functionality.

As shown on figure 3, a communication system is used to link the *Virtual Cockpit* and the *R/C Blimp* Servo Controller Server. It performs in real-time with a latency of less than 1ms with 5 bytes packets (one for the channel number, and four for the servo position which is represented by a 4 bytes float value). Figure 13 resumes the lag in the communication streams.

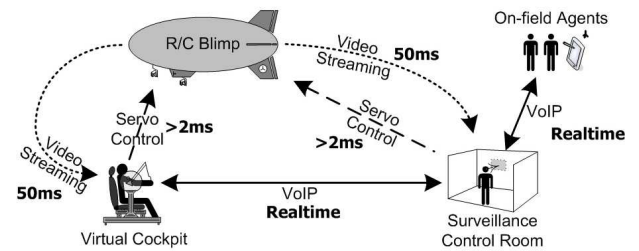


Figure 13: The lag times per communication channel.

## 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 *Virtual Cockpit* proposed in section 3.2 uses three main VR devices: the Haptic Workstation<sup>TM</sup>, an Intersense orientation tracker and a Proview XL50 HMD. We have tested successfully our teleoperation interface to pilot the *R/C Blimp*. The goals were to have a precise, intuitive and instructive interface.

The use of a responsive head orientation tracker to drive the video camera of a teleoperated 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. Thus, it increases the believability of the *Virtual Cockpit* by giving to the pilot the impression that he is inside a real Blimp. Nevertheless, our system presents a limitation: the head in the virtual cockpit can be oriented on three axis (yaw, pitch and roll) whereas the camera moves only along two axis (yaw and pitch). Thus, if the pilot rolls his head too much, there will be a slight image shift. Fortunately, we do not usually roll our head significantly. Concerning the Head-Mounted Display, it is also useful for improving the immersion of the pilot because it allows a stereo display and head movements do not obstruct the 3D images as a screen does. Usually, this kind of devices has a narrow field of view, but, in our case, it is not an issue because the pilot camera does not has a wide FOV.

We have observed that the Haptic Workstation<sup>TM</sup> 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. For example, we can provide collision alerts trough haptic stimuli in addition of visual. We can do this by gathering GPS data and comparing it to a 3D model of the surveyed area. Another example: When teleoperating an aerial vehicle, some researches [Fuchs and Moreau 2004] demonstrate that most of the crashes happen because the pilot has no idea of the external factors where the real vehicle is immersed (wind, temperature, noise, etc.). Usually it comes from a lack of audio, visual or haptic feedback. In our case, the *R/C Blimp* is really dependent of the wind force and direction, but we did not address this problem: we have only flied with good weather conditions. However, as further work, we plan to equip it with an anemometer and to send appropriate force feedback to the user in order to give him an approximation of the wind. These are examples to demonstrate the potential benefits of Haptic Workstation<sup>TM</sup> and VR devices as a teleoperation platform.



## 4.2 Surveillance Control Room

The CAVE, the eye tracker and the joystick are used in our control room to immerse the user into the scene and interact in the most natural manner with this virtual environment. We discuss in this subsection the use of VR device for improving the ability of the user to detect abnormal events and to communicate intuitively visual information.

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.

In order to improve the current system, it will be interesting to calibrate and analyze the whole process. This can be done by studying the sensitivity of the VisionTrack™ and MotionStar™ systems. Then correlate the given data to the calculated coordinates on the CAVE, in order to compute correction matrices. It will be possible in such case to reduce significantly the observed error. The transition among the different planes of the CAVE suffers from a small delay but provides a good compromise between response time and stability. This system could be improved by allowing the target to overlap different planes at the same time. Future work needs also to focus on tools for modulating, sorting and linking the displayed data dynamically, using mixed reality environments.

## 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 achieved the goal of minimizing the need for pressing buttons on the handheld device. 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" [Douglas 2004]. We believe our system responds well to current needs for mobility and intercommunication.

Future improvements consist on enabling full interoperability regardless of the current communication network. This implies, achieving transparent roaming between WLAN and digital mobile networks (GPRS, HSCSD, EDGE, etc). Deeper analysis on the interface requirements and design should be carried out to improve the efficiency of the interface. Detailed studies could lead to identify needs for additional information to be displayed on the PDA. Nevertheless, we believe the general architecture we have proposed based on web clients is a solid foundation for future improvements.

## 5 Conclusion

We have described a full surveillance and security system based on advanced Virtual Reality technologies. Our system can be applied to the surveillance of public sites such as stadiums, universities, parks, etc. 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 secondary goal was reducing the amount of people needed for operative tasks that were not directly related to the actual surveillance and security activities. A single user can teleoperate the airship used for data acquisition, while a second one can concentrate on analyzing the video streams. VR technologies eliminate the need for secondary/auxiliary operators and provide intuitive and more natural interfaces. We put in practice the use of haptic interfaces for teleoperating a flying vehicle. Our results are encouraging in the sense that the haptic interface works as well as a classical control console. Future work consist on 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 provides *On-field Agents* with valuable information and facilitates their task. This kind of VR and multimedia technology applications fit into the context of current initiatives for security and enhanced surveillance systems.

## 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

- AGAH, A., WALKER, R., AND ZIEMER, R. 1998. A mobile camera robotic system controlled via a head mounted display for telepresence. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, vol. 4, 3526–3531.
- AIRSHIP MANAGEMENT SERVICES. Athens 2004 olympics. olympic security airship programme. <http://www.airshipman.com/newsarticles/olympic>
- AVANT, D., BAUM, M., BERTRAM, C., FISHER, M., SHETH, A., AND WARKE, Y. 2002. Semantic technology applications for homeland security. In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, ACM Press, New York, NY, USA, 611–613.
- BENSLIMANE, A. 2000. Real-time multimedia services over internet. In *Proceedings of the 1st European Conference on Universal Multiservice Networks, ECUMN'00*, 253–261.

- BOULT, T. 2003. Geo-spatial active visual surveillance on wireless networks. In *Proceedings of 32nd Applied Imagery Pattern Recognition Workshop*, 244–249.
- BRUCE, S., RICE, C., AND HEPP, R. 1998. Design and test of military cockpits. In *Proceedings of IEEE Aerospace Conference*, vol. 3, 5–14.
- COLLINS, R., LIPTON, A., FUJIYOSHI, H., AND KANADE, T. 2001. Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE* 89, 10 (October), 1456–1477.
- CUCCHIARA, R., GRANA, C., PRATI, A., AND VEZZANI, R. 2003. Computer vision techniques for pda accessibility of in-house video surveillance. In *First ACM SIGMM international workshop on Video surveillance, IWVS '03*, ACM Press, New York, NY, USA, 87–97.
- DIOLAITI, N., AND MELCHIORRI, C. 2002. Teleoperation of a mobile robot through haptic feedback. In *Proceedings of IEEE International Workshop on Haptic Virtual Environments and Their Applications*, 67–72.
- DOUGLAS, M. 2004. Surveillance in the palm of your hand. *Mobile Radio Technology Magazine (MRT)*. [http://mrtmag.com/mag/radio\\_surveillance\\_palm\\_hand/](http://mrtmag.com/mag/radio_surveillance_palm_hand/) (October).
- ELLIS, T. 2002. Multi-camera video surveillance. In *Proceedings of the 36th Annual International Carnahan Conference on Security Technology*, 228–233.
- FONG, T., AND THORPE, C. 2001. Vehicle teleoperation interfaces. In *Autonomous Robots*, Kluwer Academic Publishers, vol. 11, 9–18.
- FONG, T., THORPE, C., AND GLASS, B. PDADriver: A handheld system for remote driving. In *Proceedings of IEEE International Conference on Advanced Robotics*.
- FUCHS, P., AND MOREAU, G. 2004. *Le Traité de la Réalité Virtuelle, deuxième édition*, vol. 1. Les Presses de l'Ecole des Mines de Paris.
- G.K. WALLACE. 1992. The jpeg still picture compression standard. *IEEE Transactions on Consumer Electronics* 38 (February), xviii–xxxiv.
- GUTIERREZ, M., OTT, R., THALMANN, D., AND VEXO, F. 2004. Mediators: Virtual haptic interfaces for tele-operated robots. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2004)*, 515–520.
- INFORMATION TECHNOLOGY JPEG2000 IMAGE CODING SYSTEM PART 1: CORE CODING SYSTEM, 2000. ISO/IEC IS 15444-1, ISO/IEC JTC1/SC29/WG1, december.
- KUMAR, R., SAWHNEY, H., SAMARASEKERA, S., HSU, S., HAI, T., YANLIN, G., HANNA, K., POPE, A., WILDES, R., HIRVONEN, D., HANSEN, M., AND BURT, P. 2001. Aerial video surveillance and exploitation. *Proceedings of the IEEE* 89, 10 (October), 1518–1539.
- LEMOINE, P., GUTIERREZ, M., VEXO, F., AND THALMANN, D. 2004. Mediators: Virtual interfaces with haptic feedback. In *Proceedings of EuroHaptics 2004*, 68–73.
- LI, S.-T., HSIEH, H.-C., SHUE, L.-Y., AND CHEN, W.-S. 2002. Pda watch for mobile surveillance services. In *Proceedings of IEEE Workshop on Knowledge Media Networking*, 49–54.
- LIU, P. X., MENG, M. Q., AND GU, J. J. 2001. A study on the interface of the internet-based teleoperation. In *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 326–331.
- MARZENARO, L., OBERTI, F., AND REGAZZONI, C. 2000. Change detection methods for automatic scene analysis by using mobile surveillance cameras. In *Proceedings of International Conference on Image Processing*, vol. 1, 244–247.
- NGUYEN, L. A., BUALAT, M., EDWARDS, L. J., FLUECKIGER, L., NEVEU, C., SCHWEHR, K., WAGNER, M. D., AND ZBINDEN, E. 2001. Virtual reality interfaces for visualization and control of remote vehicles. In *Autonomous Robots*, Kluwer Academic Publishers, vol. 11, 59–68.
- OTT, R., GUTIERREZ, M., THALMANN, D., AND VEXO, F. 2005. Improving user comfort in haptic virtual environments through gravity compensation. In *Proceedings of WorldHaptics'05*, 401–409.
- OTT, R., GUTIERREZ, M., THALMANN, D., AND VEXO, F. 2005. Vr haptic interfaces for teleoperation : an evaluation study. In *Proceedings of the IEEE Intelligent Vehicles Symposium, IV'05*.
- REITER, M., AND ROHATGI, P. 2004. Homeland security. *IEEE Internet Computing* 8, 6, 16–17.
- SAWYER, S., TAPIA, A., PESHECK, L., AND DAVENPORT, J. 2004. Mobility and the first responder. *Communications of the ACM* 47, 3, 62–65.
- SCHULZBINNE, H., CASNER, S., FREDERICK, R., AND JACOBSON, V. 1996. Rtp: A transport protocol for real-time applications. In *Request for Comment RFC-1889 of the Internet Engineering Task Force, IETF*.
- SEBE, I. O., HU, J., YOU, S., AND NEUMANN, U. 2003. 3d video surveillance with augmented virtual environments. In *Proceedings of the First ACM SIGMM international workshop on Video surveillance, IWVS'03*, ACM Press, New York, USA, 107–112.
- SHEKHAR, C. 2000. Semi-automatic video-to-site registration for aerial monitoring. In *Proceedings of the International Conference on Pattern Recognition, ICPR'00*, IEEE Computer Society, Washington DC, USA, vol. 4, 736–739.
- SHREINER, D., WOO, M., NEIDER, J., AND DAVIS, T. 2003. *The OpenGL Redbook: OpenGL Programming Guide, Fourth Edition*. Addison-Wesley Professional.
- WANG, J., KANKANHALLI, M. S., YAN, W., AND JAIN, R. 2003. Experiential sampling for video surveillance. In *Proceedings of the First ACM SIGMM international workshop on Video surveillance, IWVS'03*, ACM Press, New York, USA, 77–86.
- YEN, J. 2004. Introduction. *Commun. ACM. Special Issue: Emerging technologies for homeland security* 47, 3, 32–35.
- YOON, W. K., GOSHOZONO, T., KAWABE, H., KINAMI, M., TSUMAKI, Y., UCHIYAMA, M., ODA, M., AND DOI, T. 2004. Model-Based Space Robot Teleoperation of ETS-VII Manipulator. *IEEE Transactions on Robotics and Automation* 20, 3 (June), 602–612.
- YUAN, Y., AND DETLOR, B. 2005. Intelligent mobile crisis response systems. *Commun. ACM* 48, 2, 95–98.
- ZHANG, P., TANAKA, K., SHIMIZU, E., AND ITO, M. 2003. A teleoperating system for underwater manipulator with virtual model aid. In *Proceedings of The 29th Annual Conference of the Industrial Electronics Society*, vol. 1, 882–886.