A Novel Graphical Interface and Context Aware Map for Incident Detection and Monitoring

Brett Hall, Mohan Trivedi
Computer Vision and Robotics Research Laboratory (CVRR)
University of California, San Diego (UCSD)
La Jolla, CA 92093-0407
Phone (858) 822-0002 Fax (954) 301-8172
bretthall@ieee.org, trivedi@ece.ucsd.edu

SUMMARY

Recent advances in video sensors, computer vision, graphics, mobile robotic systems, distributed computing, and high-speed computer networks have inspired a new generation of visualization and telepresence systems for Intelligent Transportation Applications. This paper focuses on the development of and experiments with a visualization and telepresence environment that together present an interactive, immersive, and context-preserving display of outdoor information. The developed visualization integrates a variety of types of information about a section of interstate such as aerial photographs, maps, and live outdoor videos into a single interactive environment. The persistence of context in the visualization aids the remote viewer in more easily comprehending and exploring the outdoor space. All the environments systems and algorithms are integrated and are being extensively tested in a real-world ITS application context using several novel testbeds.

INTRODUCTION

Several important requirements must be considered when developing an outdoor tele-exploration (remote space exploration) environment. These requirements need to be satisfied to give the environment user a visually appealing and effective sense of the remote space. First, a variety of types and resolutions of content-rich information is needed to adequately represent the space and the appropriate details within the space. Second, the application should be interactive and allow the user to specify which areas, scales, and information they are interested in seeing at that instant. Third and most importantly, the information needs to be presented together with its surrounding environmental context to give the user a good comprehensive sense of the space. Finally, the visualization environment needs to transition and adapt smoothly between the requested views to most effectively convey this large quantity of information and maintain an overall sense of togetherness.

A common example of a tool that responds to a similar need is a paper map used for navigation and orientation within a new area. This medium presents the information of the space in a static, symbolic manner that is useful in part because it simplifies a large amount of data down to a subset that the user is interested in such as roads, buildings, parks, hospitals, etc. However once a map is created the user cannot view new information or update or change the information already on the map if a new need arises. Additionally the paper format cannot generally meet both a need for wide-area context display and local area detail, i.e. a map for a state will not be helpful in navigation around a university campus or recreation area. In this sense the paper map is not interactive or dynamic in its view of the information, and clearly these elements are useful and desirable in a map implementation.

The environment that has been developed meets these needs with a context aware interactive map framework and application. This powerful framework provides a way to methodically acquire, process, and intuitively represent a variety of information types to the user based on their interest. The resulting application allows a user to interactively view diverse information of different types and resolutions in an interactive, integrated visualization application. Both static and dynamic information are represented within the context of the nearby environment. Context-preservation assists the user in fully understanding the space by increasing their comprehension using visual, symbolic, permanent, and temporal aspects of the space viewed interactively from a remote location.

In many ITS applications a wide area of roadbed needs to be continuously monitored for multiple conditions including incidents, congestion, weather, road damage, and other hazardous conditions. With the trend of rapidly increasing highway usage in urban areas these conditions need to be detected and resolved quickly to minimize the health and congestion impacts of delayed emergency and hazard response. Roadside video sensors are excellent for monitoring these and other conditions because of the large amount of information available about the region in view of the camera. However one of the difficulties in using video sensors is the installation and use of the large number of sensors required to cover a large area. Viewing a large number of video streams manually in a sequential manner on a monitor or side-by-side on a bank of video monitors makes it unintuitive to understand the physical relationships (position, orientation, etc) between the video sensors and the area being monitored. In order to better understand the relative camera positions enable operators to better determine the potential wide-area effects it is useful to consider a viewing environment that preserves the context of the surrounding outdoor area enabling intuitive viewing of relevant information.

CONTEXT AWARE MAP

An architecture has been developed for the organization and development of the system that has been introduced, this architecture is pictured in figure 1. For this research to be possible, a lot of information must flow behind the scenes including video, audio, images, and control signals. It is useful to begin to understand the relationships between the functional elements and the information to get a sense for the system organization as well as its inherent reliance upon communication.

The core element of this system is the context aware map, which stores and presents a variety of information about the environment being visualized interactively to the user. This information includes maps, but is also able to include aerial images, CAD drawings, and live video information.

An important assumption of this environment is that the ground is represented as a flat, 2-D surface within the application. This assumption reduces the computational complexity and enables incorporation of live information into an interactive visualization that runs on standard computer platforms. By using two dimensions to represent a 3-D space it is logical to present the outdoor space as it would appear from an overhead perspective. This "top down" view is useful for many different tasks, and is also frequently used in paper maps and other formats. It is also important to note that accurate three-dimensional models of outdoor environments can be very difficult and time-consuming

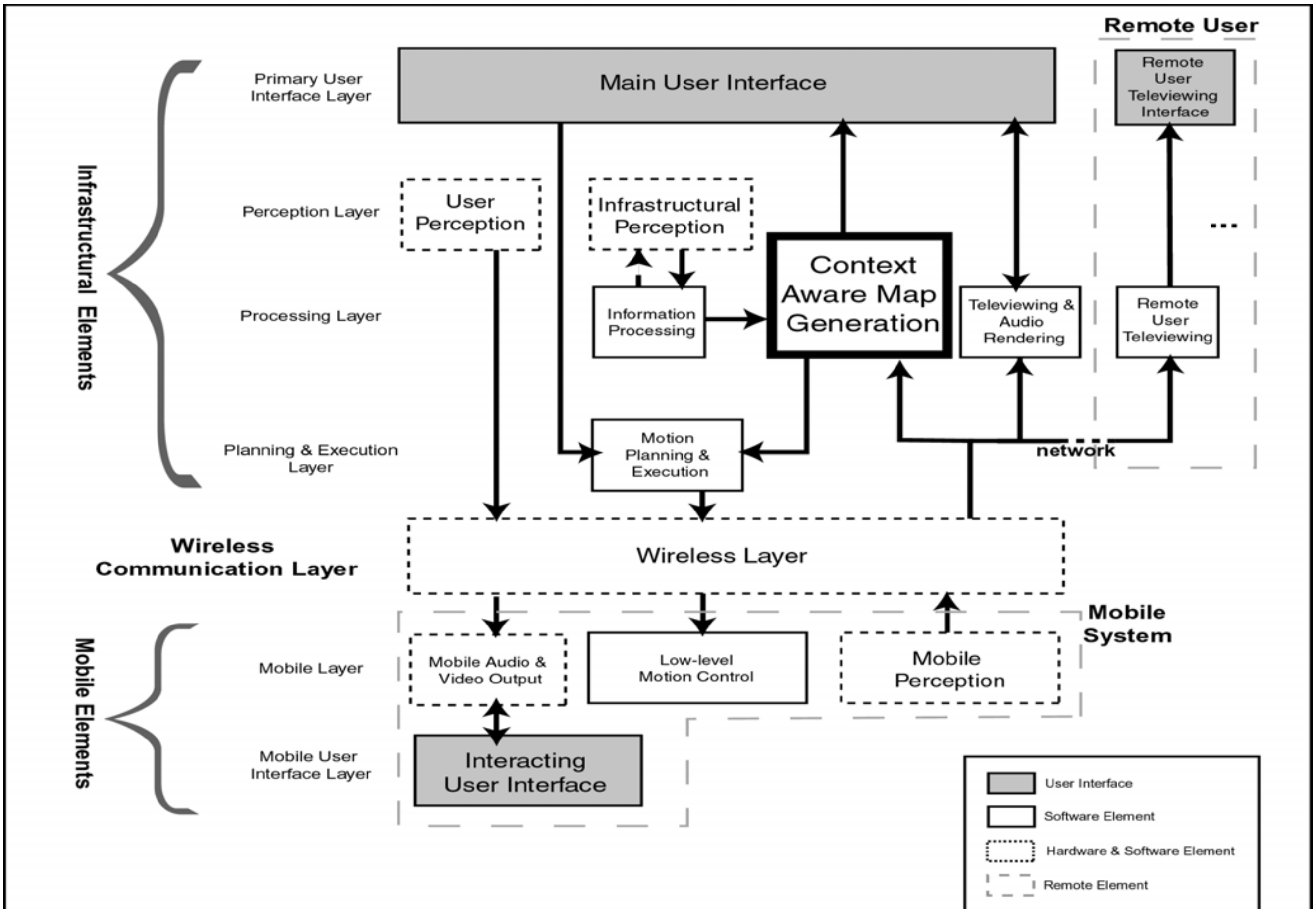

Figure 1: The system architecture has distributed sensing and processing elements and uses the context aware map to maintain and display information to the user

to produce accurately. By requiring simply a 2-D geo-registered aerial image to begin, the process of registering the information is faster and the resulting visualization is effective.

The context aware map is labeled as context aware because it can use a diverse array of information to present the context for a space. Context is defined as the circumstances in which an event occurs; a setting. For use in this visualization, representation of the environmental context can be achieved through the use of multiple pieces and resolutions of information about an area, each of which will have a natural resolution or scale associated with it. By using multiple pieces of information at different scales, the context can be captured and represented. A multi-layer approach is very practical to represent and display these different types of 2-D information within a single interface. In this layered approach, each set of flat information has a “depth” value that will control its position in Z relative to the other layers. A layer with a higher Z value will appear on top of the other information, a layer with a lower Z value will appear occluded by a higher layer if one is present. However to provide this information in a comprehensible format a physical ground basis is needed to set the relative information positions and scales. This process of geo-registration can be very complex; however using an orthographically registered aerial image of the area can make a good approximation. Correct information registration with respect to this background image and ground basis is accomplished using a simple 4-point registration method.

An important novel aspect of this visualization is that it incorporates live information about a space within the same interface as the other information. The process of displaying live video updates the images when a new image is available from the video source. The display and acquisition of live video is also synchronized with the environment visual display loop to avoid flickering video.

TELE-EXPLORATION FRAMEWORK

A framework and system organization was necessary to begin system development. This framework establishes the modularity and flexibility of the resulting system as well as illustrating how the elements are assembled into a working system. The mobile aspect of this framework incorporates the control and bi-directional information transmission of mobile platforms into the strategy. The use of mobile elements further expands the ability of the system to display remote spaces from the user's desired viewpoint. This organization relies heavily upon the context aware map to display the detail and context information to the coherent information to the main user, shown at the top of figure 1. Remote users are shown at the bottom of the figure (in the mobile elements section) and on the right hand side. These users are able to communicate and interact with the main user using multi-modal interfaces.

A key component of the organization is the decision on where to separate the elements to decrease the dependence on a central coordination agent resulting in a more flexible and robust system. Dividing and distributing key system elements results in the ability to run the user interfaces and computationally complex algorithms from different locations based on available resources and needs. The context aware interactive map can be run in multiple locations for distributed users because it relies on a network connection to connect to the various elements. The implementation of this framework is object-oriented and modular such that the elements can be re-used in different implementations.

An important feature of this organization is the distribution of the video sensing elements; the video for the visualization is captured and transmitted over the network. This removes a conventional constraint of machine vision systems by eliminating the need for a video cable connected directly from the camera to the computer.

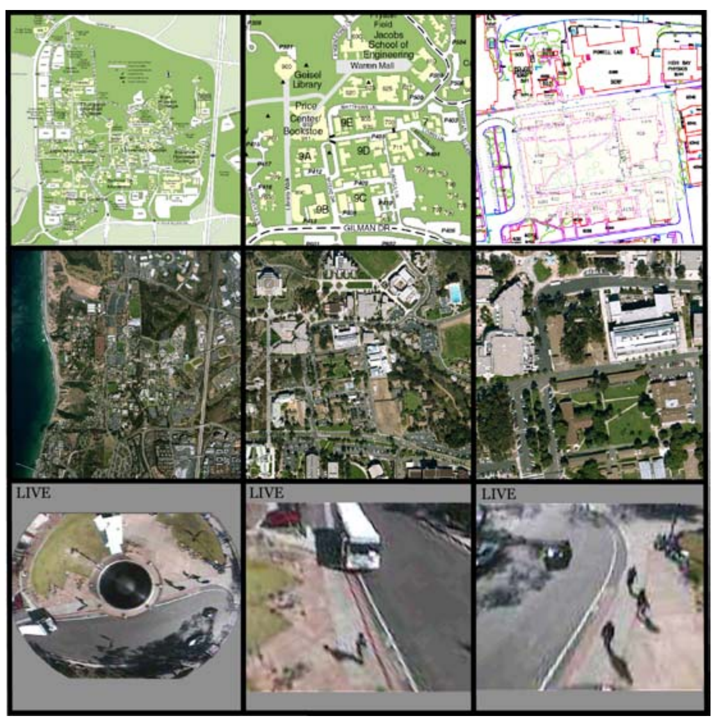

Figure 2 : Aerial images, maps, and live video contain content-rich information

STATIC & DYNAMIC STREAMS OF INFORMATION

Achieving a good understanding of a remote space requires using appropriate content-rich information as a representation of the space. There are a couple distinct categories of information that are specifically considered for use in the representation of the space:

static and dynamic. Static information includes many different and readily available types including perspective and aerial images, road maps, CAD drawings, etc. This kind of information provides both appearance-based background and foreground information as well as symbolic and other non-appearance based information. For this to be represented in the visualization the information should be stored in an image format for display. Dynamic information is information that changes over time such as live video, weather, wireless network coverage, etc. Primarily live video is being utilized as dynamic information, typically this information is displayed as a foreground layer.

To develop and test these powerful algorithms a research testbed of video clusters has been designed and installed at the UCSD CVRR. Currently several sensor clusters exist outside our lab, two are located on streetlights near an busy intersection on campus where buses, cars, bikes, and people are regularly in motion. Other sensors are mounted on the SERF building with coverage of nearby roads and courtyards. Our newest cluster has a direct view of I-5 from on campus. Each sensor cluster contains a high-speed pan/tilt/zoom (PTZ) rectilinear camera and Omni-Directional Video Sensor (ODVS) both mounted in weatherproof housings. All of this video is digitized by special video server units and streamed to the lab using direct fiber optics and a pair of one-gigabit switches. This special high-speed network connection has the capability of carrying sixteen full-size video streams simultaneously. This video is also switched to the public internet from the lab, allowing access to the

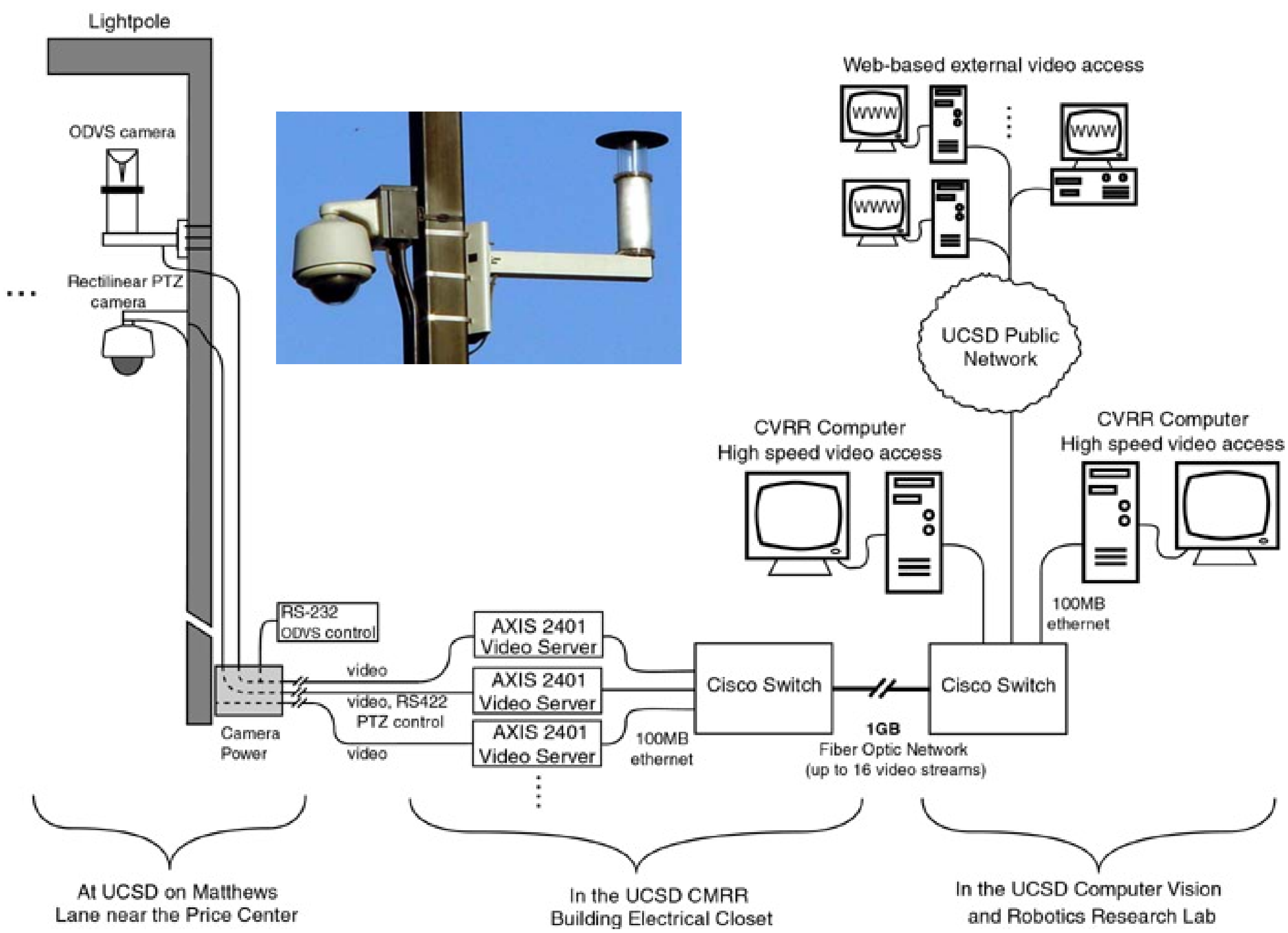

Figure 3: Outdoor Video Clusters provide continuous streaming video of roads and other areas over standard or high-speed networks

video cluster information and PTZ camera control from anywhere in the world. This distributed sensing element enables the visualization application to be run with live video anywhere network is available, increasing the flexibility and expandability of the visualization.

Two different methods for video capture within the visualization application have been implemented. The first method uses an ActiveX control supplied by the video server hardware vendor, the second method is an abstract and non-window based direct http access method. These methods each had different advantages, speed of implementation and video frame rates turned out to be the most important. Method 1 was able to obtain speeds up to 30 fps and maintained good frame rates when multiple videos were viewed simultaneously, Method 2 had slower overall frame rates (maximum 15fps) but scaled better to higher numbers of videos. Method 1 was chosen for this implementation because of its consistently high frame rates (up to 30 fps at 50% JPEG compression) and ease of integration with the application. These experiments were performed with a dual PIII-700 MHz PC with 640MB RAM, a GeForce3 graphics card, and running Windows 2000 Professional.

The development of this visualization framework and application displays a concept of a useful exploration environment that will be available in future generations of portable mapping devices. With the rate at which current computing power, mobile systems, and graphics acceleration are increasing, future mobile devices will have system resources beyond what is currently available on stationary, larger computers. With the future possibilities in mind, the existing implementation is targeted for powerful desktop computing technology to enable demonstration of real-time and interactive implementations.

EXPERIMENTS

The developed environment has been demonstrated to be useful in outdoor visualization for ongoing experiments in several application contexts. Usability studies are being conducted on this research with several real-world applications, most notably traffic incident detection and management studies. A new video cluster installation along I-5 where it passes through UCSD campus has been used to develop a visualization including live video of the I-5 corridor. (See figure 4 & 5) As a component of ITS, this visualization enables viewing of live traffic conditions within context of the surrounding roadways and areas. The results from this implementation will be the focus of the work presented here.

Figure 4: Demonstration of the Interactive Zoom functionality within the visualization focusing in on a live traffic view using infrastructural cameras

Multiple new video clusters are being installed along interstates as infrastructure to improve the accuracy and speed of traffic incident management. As part of that activity, this research is being used to provide the ability to monitor and explore large areas of outdoor roadbed. Additionally, the mobile elements of this framework that were not focused on in this paper are being developed and used to interact with traffic incidents.

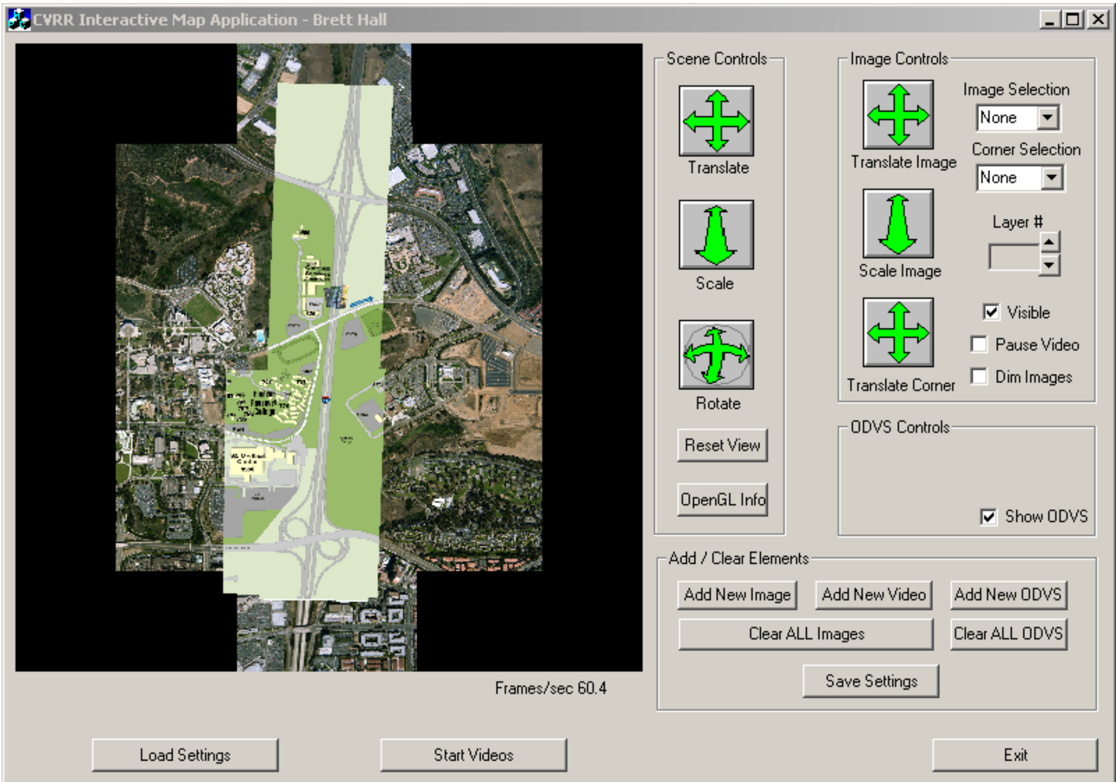

Figure 5: A visualization of I-5 around UCSD campus demonstrates the interactive display of live video within context including maps and aerial images

This implementation of the visualization runs on standard PC hardware under Windows using conventional networks for information transmission. The resulting visualization is dynamic and interactive, which is very difficult to summarize and describe in a static document. Some screen shots of the application running are included to attempt to demonstrate some of the capabilities of the application.

The application gives the user the ability to interactively view the environment easily by smoothly zooming, translating, and rotating the scene providing continuity to the visualization as the view changes. Multiple layers of static and dynamic images convey both visual and symbolic information within the context of the surrounding area. Each individual image/video can be fully manipulated using this interface by clicking and dragging on the right hand side icons. Additionally the images are selectable by clicking directly in the interface window.

Figure 6: Live video overlay is accomplished using a 4-point registration process and aligning the road correctly with the background

CONCLUDING REMARKS

The environment that has been developed demonstrates effective interactive visualization of outdoor spaces and highways using multiple types of static information as well as multiple live video elements. This visualization is particularly effective because of its presentation of the context around the area of interest. The framework has provided a flexible and robust methodology in the implementation of this environment. This research is currently being extended to several new applications including bridges, multi-sensor cluster interstate coverage, and border-crossing.

ACKNOWLEDGEMENTS

Our research is supported in part by the California Digital Media Innovation Program (DiMi) in partnership with the California Department of Transportation (Caltrans). We also wish to thank our colleagues from the lab who are also involved in related research activities.

REFERENCES

(1) B. Hall. “Visual Tele-Exploration using Multiple Cameras and Mobile Avatars: An Integrated Framework and Experiments” Masters Thesis, Electrical Engineering, University of California San Diego, December 2001

(2) K. Huang, M. Trivedi, "Networked omnivision arrays for intelligent environment", Proceedings of the Applications and Science of Soft Computing IV, August 2001

(3) M. Trivedi, S. Bhonsle, A. Gupta, "Database Architecture for Autonomous Transportation Agents for On-scene Networked Incident Management (ATON)," International Conference on Pattern Recognition, Barcelona, Spain, September 2000

(4) R. Capella. " Real-Time System for Integrated Omni-Directional and Rectilinear Image-based Virtual Walkthroughs," Masters Thesis, Electrical Engineering, University of California San Diego, 1999.

(5) K. Ng, H. Ishiguro, M. Trivedi, T. Sogo. “Monitoring Dynamically Changing Environments by Ubiquitous Vision System,” IEEE Visual Surveillance Workshop, Fort Collins, Colorado, June 1999.

(6) S. Hsu, S. Samarasekera, R. Kumar, H.S. Sawhney. “Pose Estimation, Model Refinement, and Enhanced Visualization Using Video,” Proceedings of CVPR 00, Hilton Head Is., SC, vol. I, pp.488-495, June 2000.

(7) S. Teller, M. Bosse, D. de Couto. ”Eyes of Argus: Georeferenced Imagery in Urban Environments,” GPS World, pp. 20-30, April 2000.

(8) B. Hall, K. Huang, M. Trivedi. "A Televiewing System for Multiple Simultaneous Customized Perspectives and Resolutions," Proceedings of the Intelligent Transportation Systems Council, Oakland, August 2001.

(9) M. Trivedi, B. Hall, G. Kogut, S. Roche. “Web-based teleautonomy and telepresence,” Proceedings of SPIE Vol. 4120, p. 81-85, San Diego, August 2000

(10) L. Lee, R. Romano, G. Stein. “Monitoring activities from multiple video streams: establishing a common coordinate frame,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, (no.8), IEEE Computer Society, August 2000.

(11) R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa. “A System for Video Surveillance and Monitoring,” tech. report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000.

(12) R. Kumar, H.S. Sawhney, J. C. Asmuth, A. Pope, S. Hsu. "Registration of video to geo-referenced imagery," Proceedings of ICPR 98, Brisbane. Australia, Aug. 1998.