# Cooperative Multi-Sensor Video Surveillance[*]

**Takeo Kanade, Robert T. Collins, and Alan J. Lipton**
Carnegie Mellon University, Pittsburgh, PA
E-MAIL: {kanade,rcollins,ajl}@cs.cmu.edu
HOMEPAGE: http://www.cs.cmu.edu/~vsam

**P. Anandan, Peter Burt, and Lambert Wixson**
David Sarnoff Research Center, Princeton, NJ
E-MAIL: {panandan,pburt,lwixson}@sarnoff.com

## Abstract

Carnegie Mellon University (CMU) and the David Sarnoff Research Center (Sarnoff) have begun a joint, integrated feasibility demonstration in the area of Video Surveillance and Monitoring (VSAM). The objective is to develop a cooperative, multi-sensor video surveillance system that provides continuous coverage over large battlefield areas. Image Understanding (IU) technologies will be developed to: 1) coordinate multiple sensors to seamlessly track moving targets over an extended area, 2) actively control sensor and platform parameters to track multiple moving targets, 3) integrate multi-sensor output with collateral data to maintain an evolving, scene-level representation of all targets and platforms, and 4) monitor the scene for unusual "trigger" events and activities. These technologies will be integrated into an experimental testbed to support evaluation, data collection, and demonstration of other VSAM technologies developed within the DARPA IU community.

## 1 Introduction

The recent growth in diverse imaging sensors and deployment platforms opens exciting new possibilities for Video Surveillance and Monitoring (VSAM) systems that provide continuous battlefield awareness. Future military scenarios will involve multiple sensors mounted on maneuverable ground and air vehicles cooperating with stationary ground sensors to monitor large battlefield areas for enemy troop movements (Figure 1).

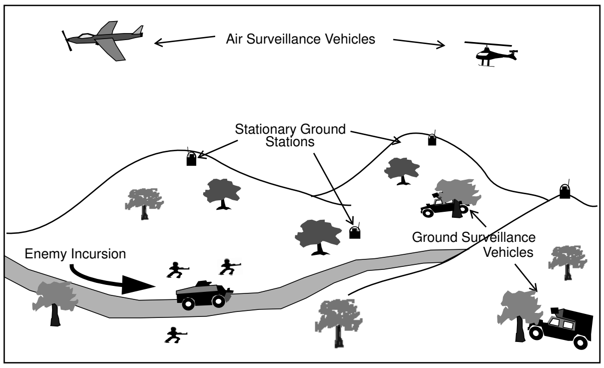


**Figure 1:** Multiple sensors cooperate to provide broad battlefield coverage.

Carnegie Mellon University (CMU) and the David Sarnoff Research Center (Sarnoff) have begun an integrated feasibility demonstration (IFD) to develop image understanding (IU) technologies to support this cooperative, multi-sensor, battlefield VSAM scenario. This report describes the overall objectives of the CMU-Sarnoff VSAM IFD project, their relevance to battlefield situational awareness, the key scientific and technology challenges to be addressed, and plans for the development, demonstration, and evaluation of the new VSAM technologies.

## 2 Objectives and Military Relevance

The major object of the CMU-Sarnoff IFD team is to develop a suite of VSAM technologies that enable a single human operator at a worksta-

[*]This work is funded under DARPA BAA 96-14.

tion to supervise a network of remote VSAM platforms (stationary, moving on the ground, or airborne), having multiple, steerable sensors. Platform surveillance operations will be mainly autonomous, notifying the operator only of salient information as it occurs, and engaging the operator minimally to alter platform operations. The network of sensors will cooperate to perform broad-area monitoring and continuous target tracking over large areas that can not be viewed continuously by a single sensor alone. The IFD team will integrate this technology suite into an experimental testbed system that will additionally support evaluation, data collection, and demonstration of other VSAM technologies developed within the DARPA IU community.

Cooperative multi-sensor surveillance will significantly enhance battlefield awareness, by providing the commander with complete and continuous coverage of troop movements and target activities within a broad area. Examples of military scenarios that can use the VSAM technologies include:

- *perimeter monitoring*, in which a continuous watch is maintained over a familiar facility such as a warehouse, a military base, or a sensitive building. The major objectives of the monitoring task are to be alert to potential incursions by enemy troops or other suspicious activity,
- *forward observer*, in which ground and air-based surveillance vehicles are sent ahead of the troops to determine potential hazards for intended troop movements,
- *border patrol*, in which border areas are monitored for potential drug and/or weapon trafficking,
- *point reconnaissance* of a location such as a bridge, weapon storage site, an entry gate, or a suspected terrorist hangout for unusual movements and loitering by people or vehicles, and
- *cantonnement facility monitoring*, in which video observations of a weapons cantonnement facility collected over multiple days are analyzed to detect potential weapon movements.

The prototype testbed system that will be developed by the CMU-Sarnoff team will facilitate growth in the area of VSAM IU by supporting development and evaluation of component technologies. Potential military users will be able to observe field demonstrations, guide the selection of problems, and provide feedback on the utility of the developed components. In the optional out-years of the program, an integrated system will be delivered for testing and evaluation by military users, enabling the transfer of VSAM technologies to the DOD community.

In addition to the military applications mentioned above, this effort will also spur technology transfer to commercial applications, such as building and parking lot security, warehouse guard duty, and monitoring restricted access areas in airports. Combined ground and air surveillance capabilities also have promising applications in civilian law-enforcement operations.

## 3 Scientific and Technical Challenges

The major scientific and technical challenges of the CMU-Sarnoff VSAM approach are to: 1) coordinate multiple sensors to seamlessly track moving targets over an extended area in a visually complex environment, 2) actively control sensor and platform parameters to track multiple moving targets, 3) provide scene-level representations of targets and their environment by integrating evolving visual, geometric, and symbolic sensor observations together with collateral scene data, and 4) monitor the scene for unusual "trigger" events and activities that should cue further processing or operator involvement. This section outlines the technical challenges that IFD research must address in order to meet the above objectives.

### 3.1 Coordinating multiple sensors

Central to the goal of the VSAM IFD program is real-time detection and tracking of targets over a wide area using multiple distributed sensors. To perform this task, the following technical areas will be addressed. Note that all of the operations described must be performed in real-time.

**Robust target detection and tracking:** Targets must be detected and continuously followed as they move through a large cluttered area, even when they disappear behind occluding surfaces and later reappear, or when they stop and later resume moving. Tracking must be maintained as the camera pans, tilts, and zooms in to obtain a closer look, and in the presence of image motion containing 3D parallax induced by movement of the sensor platform. A combination of motion and appearance cues will be used to achieve robust target tracking.

**Continuous target following using multiple distributed sensors:** Targets must be continuously followed as they move out of the field-of-view of one sensor into that of another. This requires establishing the correspondence of the fields-of-views of the different cameras to achieve target "hand-off". It also requires appearance matching of the target as seen by sensors with significantly different viewpoints.

**Cooperative ground-and-air surveillance:** Targets detected in airborne views can be used to cue local ground sensors, and vice versa. This requires geo-registering airborne views with a set of ground-based views. In order to achieve the geolocation accuracy required for air-to-ground (or ground-to-air) hand-off, visual pose refinement using cultural landmarks and terrain features will be performed to refine initial pose estimates based on platform ephemeris data.

## 3.2 Active sensor control

Active camera control will be performed to maximize system performance and maintain target pursuit over large areas. This involves controlling sensing parameters (e.g. view direction, zoom, panning speed, vergence angles), processing resources (resolution, focus of attention, load balance), and mobile sensor deployment.

**Sensor planning and control:** Sensor hand-off for cooperative, multi-sensor surveillance will be achieved using standard visibility and occlusion analysis. This requires using collateral terrain maps and 3D site models containing man-made features to perform visibility analysis from each sensor position, to determine, based on current estimates of target trajectory, which sensor will have the closest, unoccluded view. This work will also involve task-based planning of new camera views, while imposing physical constraints on sensor platform mobility.

**Multi-tasking for multiple target tracking:** Occassionally, a single camera resource must be used to track multiple moving objects, not all of which fit within a single field of view. This problem will be addressed by introducing sensor multi-tasking, meaning that the camera field of view will be periodically switched between two (or more) targets that are being monitored. This requires continuously locating and updating the target positions within a panoramic reference mosaic image or a map, and using a combination of visual and inertial information to perform the scans.

## 3.3 Scene-level representation

An important component of the VSAM testbed is an interface that allows the human operator to visualize all available scene information, and to control the sensor suite to achieve mission objectives. To do this, information from multiple sensors will be integrated with collateral site information to provide an evolving scene-level representation (Figure 2).

**Multi-sensor information integration:** Information in the form of estimated target locations and appearances will be gathered from many difference sensors, possibly of different sensing modalities, and redundant data must be correlated and merged. This will be handled by transforming all target and platform locations and trajectories into a georeferenced coordinate system, either by locating them with respect to calibrated reference imagery, or solving for 3D position directly using known constraints such as terrain elevation.

**Dynamic scene visualization:** Comprehending a vast flow of incoming information from multiple sensors, regarding multiple targets, is a challenging task for any human operator. To make the task easier, a comprehensive, graphical visualization of the dynamic scene will be presented to the user that combines elements of

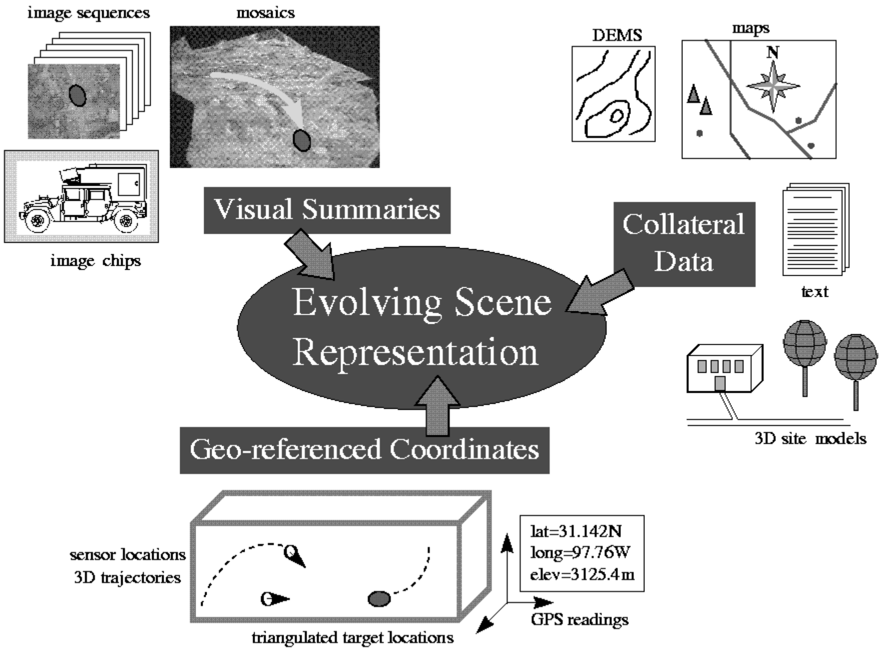


**Figure 2:** Components of an evolving, dynamic, scene-level representation.

visual sensor imagery, prior geometric models of the scene and targets, other collateral information such as maps, and symbolic depictions of activities of interest.

**Collateral data integration and update:** Prior collateral information about the scene will be maintained in the form of annotated maps, digital elevation models, reference imagery (e.g. satellite photos) and symbolic 3D site models. These will all be tied to the common geospatial scene coordinate frame. Incoming imagery will be used to not only update the positions of dynamic targets in the evolving scene model, but also to refine these prior models based on close-range views from the deployed sensor platforms.

### 3.4 Activity Monitoring

By broadening the scope of VSAM technology beyond simple 2D image-level tracking into dynamic, scene-level descriptions co-registered with 3D collateral data, the CMU-Sarnoff approach will enable research into high-level activity and event monitoring. For example, the system could be tasked to monitor sensitive areas for such "suspicious" activities as:

- vehicles going the wrong way down a one-way street,
- vehicles (or people) entering a restricted access area,
- vehicles that repeatedly circle the block around a sensitive building,
- people coming and going from the front door of a suspected drug hideout,
- pedestrians who loiter in front of a building for a long time,
- pedestrians trying to look over a fence, or peer through windows.

Many of these tasks would be difficult, if not impossible, to perform with 2D visual image data alone, but are enabled by having co-registered scene models to provide regions of interest and expected patterns of motion.

## 4 The VSAM testbed

The CMU-Sarnoff team is developing a testbed architecture that will support the design, evaluation, and demonstration of VSAM IU technologies developed by the IFD team and the rest of the DARPA VSAM community. The testbed architecture consists of multiple sensor processing units (SPUs) in the field, communicating with an operator control unit (OCU) connected to an operator console (see Figure 3). The goal

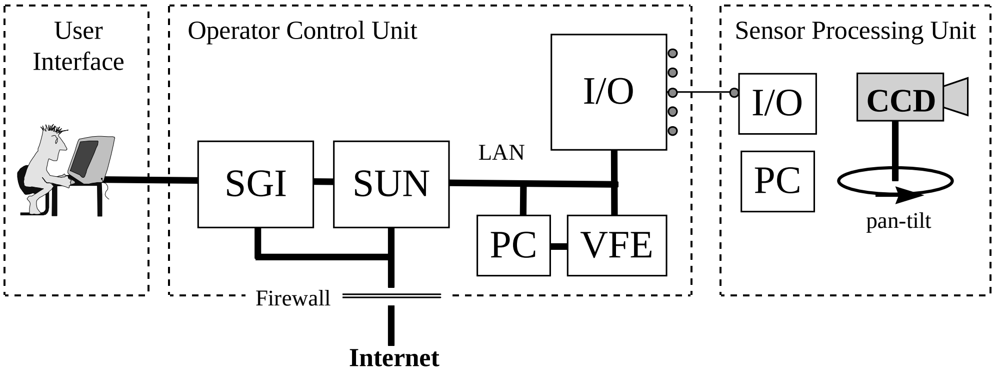


**Figure 3:** The VSAM testbed architecture

has been to design a testbed that is both rich enough (in terms of equipment and computational power) and flexible enough (in terms of functionality) to support a wide range of VSAM research.

**Sensor processing units.**

Multiple sensor processing units (SPUs) are mounted in fixed locations on hills or rooftops to provide distributed coverage over a wide area. At least two sensors will be mounted on mobile platforms – one ground vehicle (NavLab), and one airborne vehicle (autonomous helicopter or chartered flight).

The specification of what constitutes an SPU is intentionally left open-ended within the testbed architecture, in order to encompass a wide variety of sensor types such as monocular visible light and IR cameras, stereo heads, LADAR, and acoustic sensors. However, a typical SPU will consist of a color CCD camera with a motorized zoom lens, mounted on a controllable pan-tilt head. An onboard controller (e.g. Pentium PC) is responsible for collecting and managing sensor data, communicating with the OCU and generating the appropriate signals to control sensor hardware. Sensors mounted on mobile platforms will have access to real-time video processing hardware (Sensar VFE) for frame-rate video stabilization, and to onboard pose sensors for providing estimates of SPU location and orientation.

**Operator Control Unit.**

The operator control unit (OCU) is responsible for integrating the results produced from multiple sensors with a database of collateral scene information, in order to form and maintain an evolving, dynamic scene representation. The core of the OCU consists of two workstations (SGI and/or Sun), one dedicated primarily to the graphical user interface and the other handling information fusion and tasking control. Input from sensors in the field comes in via communication links ranging from radio ethernet and cell phone for discrete packets of symbolic information, to microwave links and coax cable for higher-bandwidth transmission of video streams. A Sensar VFE real-time video processor controlled by a PC host is provided to stabilize video streams from sensors that don't have enough onboard processing power. A local area network (LAN) connects all components to each other, and to an external internet connection, protected by a firewall.

**Graphical User Interface.**

One of the technical goals of the VSAM project is to demonstrate that a single human operator can effectively monitor a large battlefield area. Towards this end, the test system will have a graphical user interface for battlefied visualization and sensor suite tasking. Through the inerface, the operator can task individual sensor units, as well as the entire testbed sensor suite, to perform surveillance operations such as generating a quick summary of all target activities in the area. The operator may choose to see a map of the area, with all target and sensor platform locations overlaid on it. Alternatively, the operator may select a more immersive display (with a more limited field of view) by interacting with a texture-mapped 3D model of terrain and cultural features (buildings and roads), within which dynamically updated sensor and target locations are displayed (Figure 4).

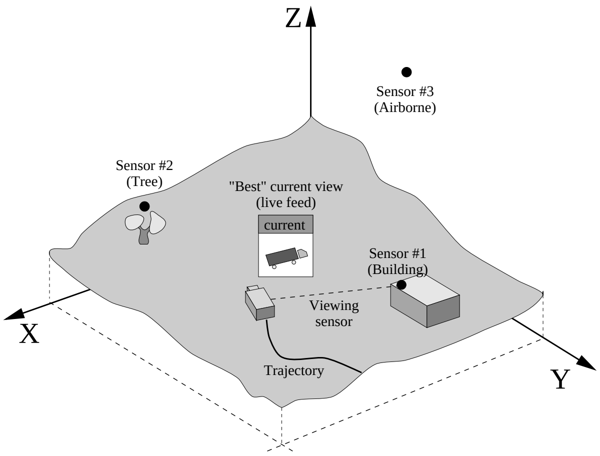


**Figure 4:** Sample user interface visualization.

The user interface will minimize operator typing by employing a graphical screen interface with "hot" areas that can be selected with a mouse or touch screen. For example, pointing to a sensor icon on the screen could bring up an overlay window showing the stabilized video output from that sensor viewpoint.

## 5 Demonstration Plan

Technology developed under the VSAM program will be demonstrated to the user community and the DARPA IU community through annual demonstrations. The Year 1 demonstrations will emphasize individual ground and air-based surveillance capabilities, whereas the Year 2 demonstration will emphasize combined ground and air surveillance. The Year 1 demonstrations are described in more detail below.

### 5.1 The Bushy-Run Site

The CMU "Bushy Run" site is a decommissioned chemical and nuclear research facility that sits on 140 acres of land in Penn township, Westmoreland county (Figure 5). The site is currently unoccupied, and ideal for research experiments and realistic demonstrations of the VSAM IFD testbed system, using both ground-based and airborne sensors to cooperatively track vehicles and people moving through an outdoor environment.

Bushy Run is 30 minutes from the CMU campus, and has expansive open spaces, tree lined fields with varying degrees of ground vegetation, and two empty two-story buildings along paved roads. The buildings, roadways, and natural terrain at the site, combined with the facility's limited access to the public, make it an ideal location for controlled experiments and demonstrations involving moving object detection and tracking, as well as for conducting potentially dangerous flight tests involving experimental aerial platforms without endangering human bystanders.

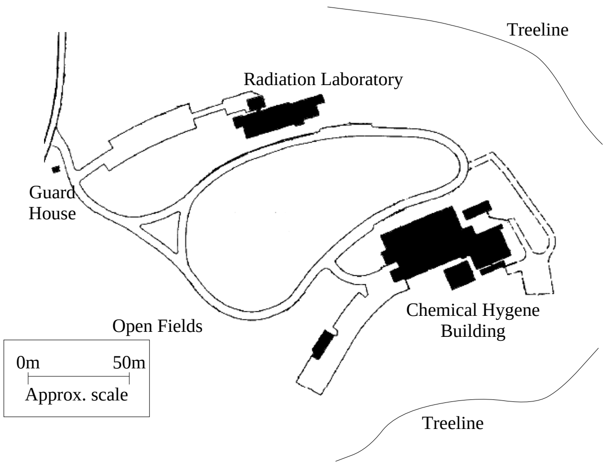


**Figure 5:** The CMU Bushy Run demo site.

### 5.2 Year 1 Demonstrations

Two IU capabilities will be highlighted during the first year: cooperative ground-based surveillance, and multi-target tracking by an airborne sensor. Below is a brief description of the objectives of each, coupled with tentative military scenarios that set the stage for the demonstrated IU capabilities.

#### Cooperative Ground-Based Surveillance

Consider a facility monitoring scenario, which involves continuous surveillance and monitoring of a military facility such as a base or warehouse complex. The site is assumed to be familiar, and detailed site-specific information (e.g. site models) are available. The site is also assumed to be too large to monitor with a single camera.

Several ground-based stationary sensors are mounted throughout the facility, and along its perimeter. A central operator control unit allows security personnel to analyze information gathered by the sensors. The operator is alerted if vehicles or pedestrians attempt to breach the perimeter in a location other than a normal facility entrance. The system also monitors for suspicious activity within the compound, presenting video clips of interesting events to the operator for review. The operator may designate targets of interest, and the system automatically tracks them through the course of their movements. The key IU capability to be demonstrated occurs as the system hands-off control from one stationary sensor to another, following the target as it enters and exits the fields-of-view of the different sensors. The goal is to maintain a continuous visual lock on the target, as it travels through the compound.

**Multi-Target Tracking by a Single Sensor**

Consider the need to provide instantaneous situational awareness on the battlefield, where multiple friendly and enemy forces are simultaneously engaged over a large area. A single unmanned air vehicle (UAV) is deployed to circle the battlefield in order to send back timely information on the locations of the combatants. The battlefield is too large to fit in a single field of view when the sensor is focussed at a resolution high enough to distinguish friendly from enemy forces. Nonetheless, it is desired to detect and track as many moving objects as possible, given the limited resources available.

To handle this situation, the UAV VSAM system is instructed to operate in multi-tasking mode, and the sensor begins to scan the scene. As the field of view passes each moving target, it's location is noted with respect to a reference mosaic in which pixel locations are directly related to geographic coordinates (using known transformations calibrated previously). The sensor continously pans and tilts around the scene, noting new targets as they become visibile for the first time. After a quick scan to summarize the positions of moving objects in the scene, the positions of targets of interest are continuously updated by switching the sensor field of view between each of them in turn, using a combination of visual and inertial information to determine where to scan. When returning to update the position of an object, the search begins from its expected new location, given its last known position and trajectory.

## 6 Evaluation Plan

Key features of the IFD VSAM research program are 1) cooperative use of multiple sensors and 2) moving platforms to provide 3) broad area surveillance and 4) real-time tracking in 5) cluttered and urban environments. We will evaluate the IFD testbed architecture and component IU technologies along several dimensions to measure system competence with respect to each of these features.

False alarm rates for target detection and cueing will be measured with respect to a number of varying factors such as size and distance of the target from the sensor, speed and direction of target trajectory, amount of scene clutter, and number of targets that are simultaneously in view. The sensitivity of moving object detection and tracking processes to ego-motion of the sensor platform will be evaluated for both pan-tilt systems and general vehicular (ground and air) motion. The effectiveness of multi-sensor VSAM integration will be measured by quantifying spatial and temporal discontinuity induced in perceived object trajectories as tracking control is passed between adjacent sensors. We will experimentally determine how large an area can be reliably monitored by a given number of fixed and moving sensor platforms, and how each sensor should be deployed to maximize VSAM performance. The accuracy with which sensor occlusion can be predicted using static scene models and dynamic target models will also be addressed.

The main use of multi-sensor integration in this system is to accurately localize targets within the 3D scene. Geolocations of observed targets will be computed in a number of ways: by multi-image triangulation if the target is viewed simultaneously by multiple sensors, by range-from-size computations or backprojection of target center of mass onto a collateral terrain map if the target is viewed by a single sensor only, and by extrapolating from the last known trajectory if the target is currently occluded from all sensor viewpoints. In each case, accuracy for the computed target location and trajectory will be evaluated by measuring the deviation between estimated and actual locations of ground truth targets with respect to the number and configuration of sensor platforms.

Beyond these systematic tests of system capabilities, the IFD testbed will also be exercised under a variety of weather conditions and at night (using infrared and laser ranging sensors) in order to assess how these environmental elements and sensor modalities affect system performance.

## 7 Conclusion

Carnegie Mellon University and the David Sarnoff Research Center are developing a cooperative, multi-sensor video surveillance and monitoring system. Multiple sensors on stationary and moving platforms will cooperate to continuously track moving targets through large, cluttered environments. Extracted target and ephemeris data is collected at an operator control station, and combined with prior collateral information to build and maintain an evolving, dynamic representation of the scene. A single human operator will be able to interact with this scene representation through a graphical user interface, allowing him or her to effectively task the multiple sensors and monitor targets over a large area. An experimental testbed system is being built to support evaluation and demonstration of these and other VSAM technologies being developed within the DARPA IU community.

## References


[1] J. Costeira and T. Kanade, "A multi-body factorization method for motion analysis," *Proc. ARPA Image Understanding Workshop,* 1996, pp.1013–1025.

[2] L.S. Davis, R. Bajcsy, M. Herman, and R. Nelson, "RSTA on the move: detection and tracking of moving objects from an autonomous mobile platform," *Proc. ARPA Image Understanding Workshop,* 1996, pp.651–664.

[3] M. Hansen, P. Anandan, G. van der Wal, K. Dana, P. Burt. , "Real-time scene stabilization and mosaic construction," *IEEE Workshop on Applications of Computer Vision,* 1994.

[4] M. Irani and P. Anandan, "A unified approach to moving object detection in 2D and 3D scenes," *Proc. ARPA Image Understanding Workshop,* 1996, pp.707–718.

[5] R. Kumar, H. Sawhney and J. Asmuth, "Geospatial registration," *Proc. DARPA Image Understanding Workshop,* 1997, this proceedings.

[6] R. Kumar, P. Anandan and K. Hanna, "Shape recovery from multiple views: a parallax based approach," *Proc. Darpa Image Understanding Workshop,* 1994.

[7] L. Matthies, R. Szeliski and T. Kanade, "Kalman filter-based algorithms for estimating depth from image sequences," *International Journal of Computer Vision,* Vol.3, 1989.

[8] H. Sawhney and R. Kumar, "True multi-view registration with application to automosaicing and lens distortion correction," *Proc. DARPA Image Understanding Workshop,* 1997, this proceedings.

[9] J. Shi and C. Tomasi, "Good features to track," *IEEE Conference on Computer Vision and Pattern Recognition,* 1994, pp. 593–600.

[10] C. Tomasi and T. Kanade, "Shape and motion from image streams: factorization method," *International Journal of Computer Vision,* Vol. 9(2), 1992.