2010 13th International IEEE
Annual Conference on Intelligent Transportation Systems
Madeira Island, Portugal, September 19-22, 2010

Map Interface for Geo-Registering and Monitoring Distributed Events

Brendan Morris and Mohan Trivedi

Abstract— **While there have been many advances in intelligent monitoring, it is still difficult to understand complex environments without human assistance. Rather than focus on fully automated monitoring, this work advocates user-centered analysis. A standardized analysis environment for visual fusion and embedding of information is developed called CANVAS (Contextual Activity Notification Visualization Analysis System). CANVAS provides a user interaction interface for instantaneous feedback of contextual processing units which enables high level semantic extraction and understanding. This assistive tool utilizes advanced monitoring techniques to provide the desirable context necessary for decision making and planning. In addition, it takes advantage of web-based technology for ubiquitous accessibility.**

I. INTRODUCTION

Intelligent monitoring of environments has progressed rapidly in the past 10 years [1]. Multiple cameras are now utilized to monitor complex environments because of improved video compression and network transmission. Monitoring goals have transitioned from low level surveillance tasks (*e.g.* detection and tracking) to higher level environmental and situational awareness.

Accurate environment understanding requires incorporation of the needs of the monitoring system user. This user must be included in the analysis loop for critical decision because these decisions are based on a deep understanding of the environment and the monitoring situation. Unfortunately, due to vasts amounts of streaming information, limited attention, and distributed awareness, a human operator can not accurately monitor large areas and networks effectively. Automatic computational techniques are vital for the monitoring process in order to highlight and guide user attention to relevant areas. The large volumes of monitoring data must be condensed and presented to a user in an accessible format suitable for quick decision making.

This work presents a surveillance and monitoring system called CANVAS. CANVAS is a Contextual Activity Notification Visualization Analysis System that spatially integrates distributed sensors. It is used to develop advanced monitoring techniques, integrate cameras and GPS enabled devices, and centralize information [2]. It provides a flexible backbone which allows improvements to vision algorithms while providing a seamless visualization interface. The visualization provides a user with environmental context for the distributed analysis modules in a customizable web interface for improved environmental awareness.

B. Morris and M. Trivedi are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA {b1morris, mtrivedi}@ucsd.ed

II. SYSTEM DESCRIPTION

CANVAS is a monitoring system capable of integrating spatially distributed sensors into a single unified environment for activity understanding. The web-based monitoring interface, presented in Fig. 1, contains a map for localization of sensors, environmental context, and incorporates analysis icons as well as access to live video feeds of the monitoring area. The single display is capable of monitoring a wide area in a compact workspace.

The block diagram in Fig. 2 depicts the major components of CANVAS. There are three separate design layers; the Sensor Layer, a Hidden Layer, and the User Layer. The Sensor Layer provides the interface to the physical environment by taking measurements with a number of sensors. The Hidden Layer is the processing backbone of the system and is transparent to the end user. In this layer, the raw sensor measurements which describe the current state of the scene are archived in the system database. In addition, computational models are trained to understand the environment (*e.g.* distinguish pedestrians from vehicles or model highway traffic flow) in real-time for live analysis. The User Layer, provides the web monitoring interface for video contextualization and environmental and situational awareness. A user is able to query the database for pertinent information and have the display updated in real-time.

III. SENSOR LAYER

Environment perception is handled by the Sensor Layer where the Data Collection block delivers the meaningful signals for CANVAS. Low level data extraction occurs through sensor specific filters which are designed to transform raw sensor output into informative features, *e.g.* tracking for motion description and measurements of object size and shape.

The main sensing modality for CANVAS are video cameras. Fig. 3 shows a map of UCSD along with images of the many camera nodes situated around campus. A variety of environments, both indoor and outdoor, with different coverage area, scale, and objects of interest are present. Both pan-tilt-zoom (PTZ) and wide are covering omni-directional cameras [2] are utilized to monitor highway traffic along Interstate 5, human/vehicle interactions on campus roads, and people indoors. Most video processing is performed remotely by transmitting video data across the network. Non-streaming cameras with a local capture machine can be used to limit the bandwidth requirements necessary from very-large scale video networks by transmitting just archival data.

In addition to video, GPS enabled devices provide a secondary sensor. The popularity of smart phones can provide

978-1-4244-7658-9/10/$26.00 ©2010 IEEE

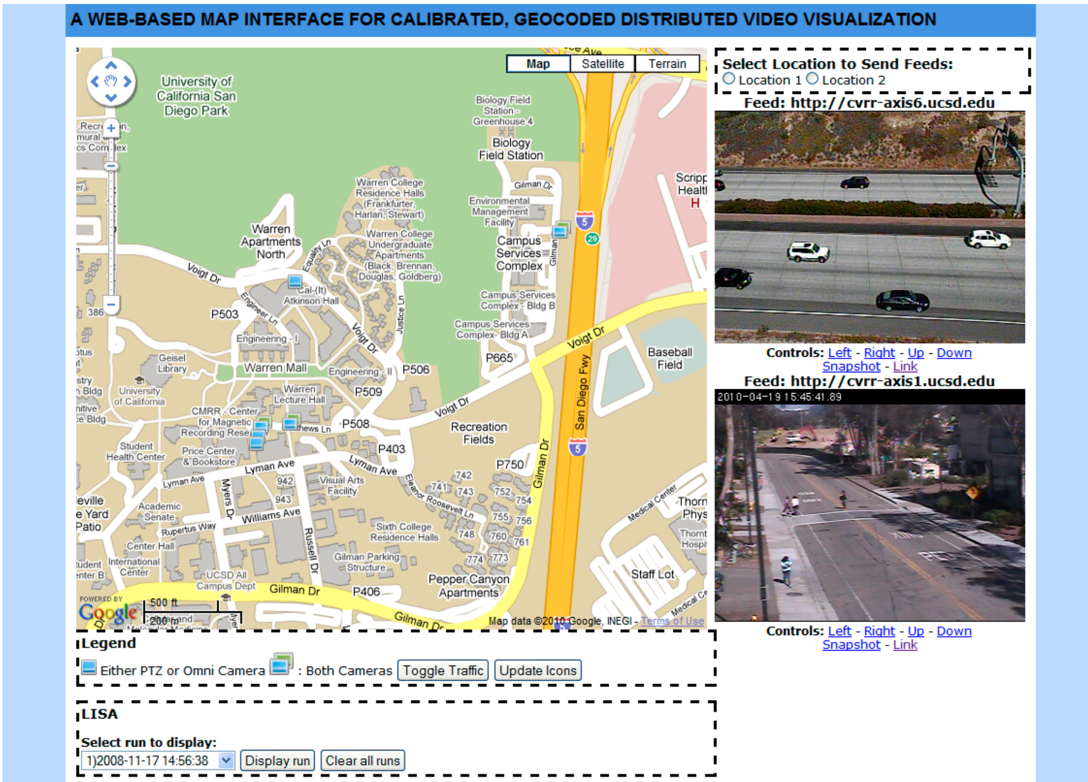


Fig. 1: CANVAS provides a web-based user interface for a user to contextualize the spatial proximity of sensors, view live video streams, and compile processing and analysis results. A map shows the location of sensors, provides information about the coverage area, and contains iconic display of events and activities. Live video provides raw, unprocessed, visual information.

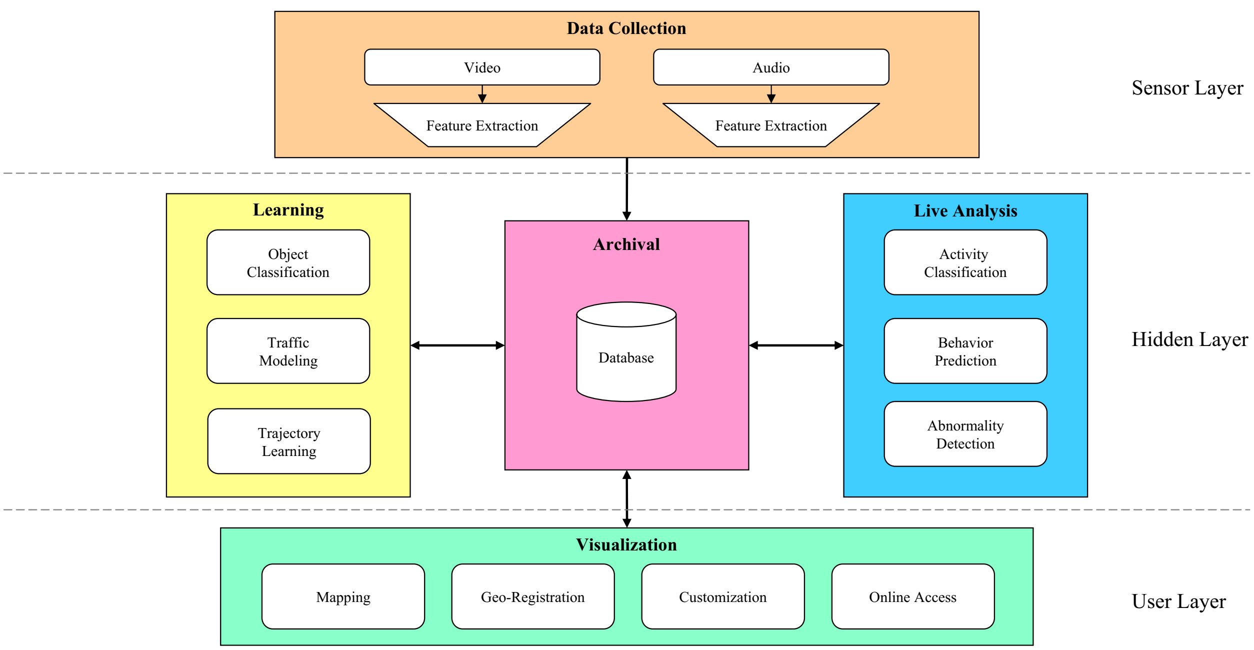


Fig. 2: CANVAS Monitoring Diagram: The monitoring framework is composed of a Sensor Layer which provides an interface to the physical environment, a Hidden Layer which houses a measurement database used to learn and infer the current activity, and the User Layer which provides contextual visualization in real-time.

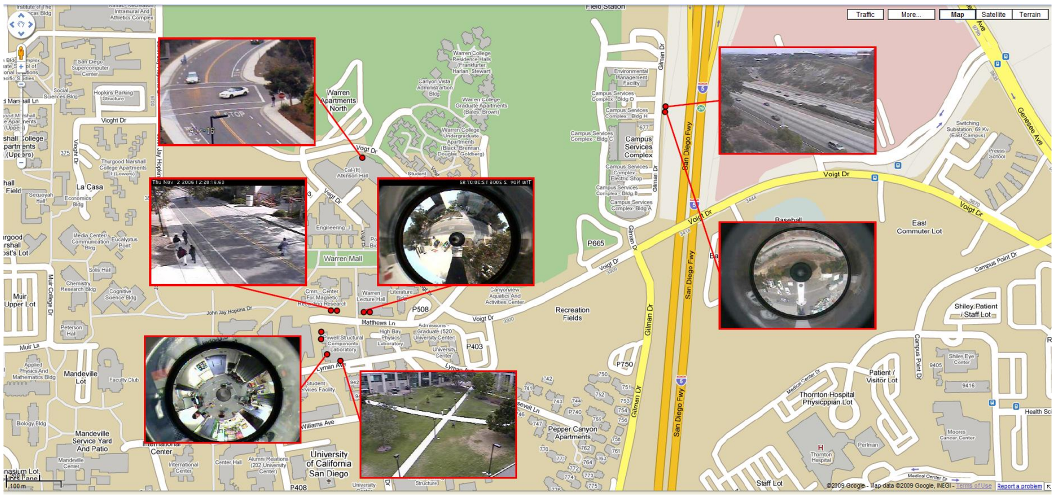


Fig. 3: UCSD video network. A network of video cameras is situated around campus to provide coverage of different environments. Both rectilinear PTZ as well as omni-directional cameras are used to monitor highway vehicle traffic and the close interactions of people and vehicles on campus.

location information from a number of users. The LISA automotive testbeds (driving capture and analysis testbeds) [3] are equipped with GPS receivers to provide tagged driving parameters such as speed and steering. With mobile internet connectivity, these measurements could be streamed in real-time.

Together the positions from infrastructure mounted cameras and mobile devices provide the raw data for situational awareness and activity understanding.

IV. Hidden Layer

The hidden CANVAS layer provides the underlying data analysis and environmental perception for activity understanding. The monitoring tasks require the storage of sensor data in order to learn methods for describing and understanding the scene in real-time.

A. Information Archival

At the heart of CANVAS is the database archival system which is implemented as a MySQL relational database. Sensor data, which provides measurements on the state of the monitored world, is timestamped and stored. Over time, a historical context emerges which enables accurate scene understanding based on real observations.

The database is split into three main partitions; data, models, and live information. The data partition holds sensor data as it is extracted. The models partition maintains the results obtained from the Learning modules. This information is used during Live Analysis to process new data. The analysis output is automatically entered into the live database partition to provide the information necessary for visualization.

B. Learning and Analysis

The Learning module develops models which can interpret sensor data through offline training. These models can then be used during Live Analysis to understand the current state of the monitoring scene. The Analysis modules are essential for effective monitoring because it eases the cognitive load of a human observer. In addition, multiple analysis tasks can be run in parallel on multiple video feeds which is something quite difficult for a human.

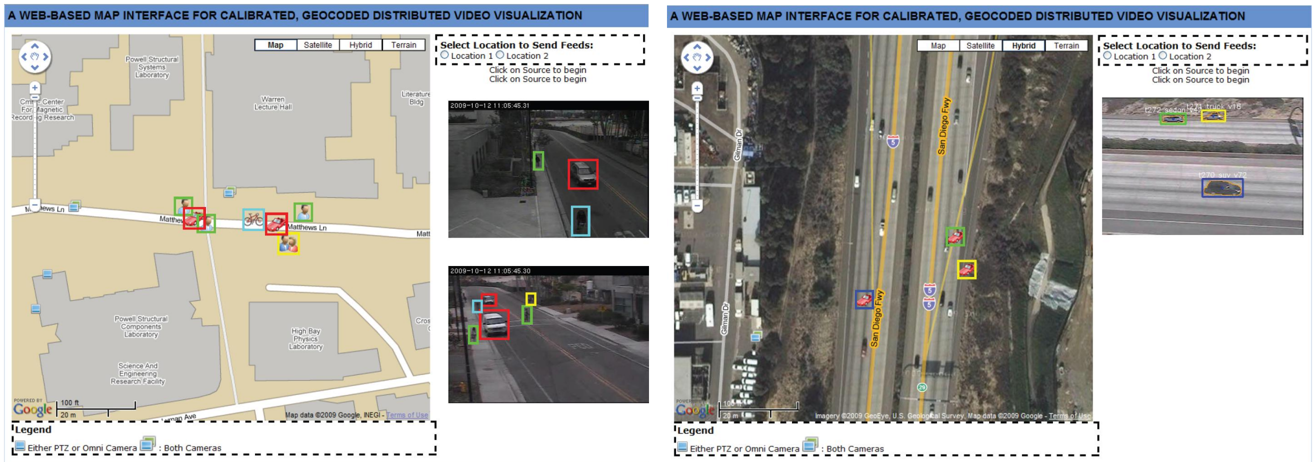
1) Object Classification: Automatically detected objects can have their type identified based on their visual signatures [4]. The 7 most often occurring vehicle types {Sedan, Pickup, SUV, Van, Semi, Truck, Bike} are identified in highway streams. This detailed real-time fleet composition is a missing management component essential for estimating emissions or infrastructure load assessment [5]. On campus, detected objects are marked as either {car, pedestrian, biker, skateboarder, or a group of people}. This classification helps with criticality assessment of situations when vehicles and people interact in close proximity.

2) Traffic Modeling: Intelligent traffic management relies on up-to-date measurements of the transportation network. A single infrastructure camera can effectively monitor a highway link [4] to extract the essential lane level measures of flow ($\frac{\#vehicles}{time}$), density ($\frac{\#vehicles}{distance}$), and speed (MPH). These traffic parameters are stored in the database where they can be aggregated over time to build the daily speed profiles which are used to detect abnormal driving.

3) Trajectory Learning: Recently, one of the most popular techniques for automated surveillance and monitoring is trajectory learning [6]. This technique makes it easier to monitor larger video networks because activity models are learned automatically without need for manual specification. Object trajectories, consisting of location and speed, are compared and clustered to build probabilistic models of typical activity [7]. These models are utilized during live analysis to describe, predict, and detect abnormalities, all critical for scene and situation understanding.

V. User Layer

The User Layer provides a common visualization environment for the display or real-time information and live



(a) (b)

Fig. 4: CANVAS Visualization Page (with processed output video for clarity rather than raw streams). (a) A campus street is monitored using two overlapping cameras. The output of object classification and tracking is marked using icons which are geo-registered on the map. (b) Environmental context is encoded using an aerial image of the highway where detected vehicles are placed in the appropriate lane.

analysis. Situational awareness is realized through functional display layers built for each of the Analysis modules. Each additional visualization layer provides a more detailed picture of the monitoring state while preserving surrounding environmental context.

Instead of overloading the display with large amounts of annotations, information is distilled and visualized through the use of icons and avatars (examples in Fig. 4). The filtered view of information limits cognitive load and helps focus attention on the locations most likely to be interesting through automatic highlighting [8].

CANVAS' web based visualization indicates the location of sensors with respect to one another, gives access to raw video feeds, presents pertinent analysis results, and provides a user interface to navigate, query, and customize the display.

A. *Mapping*

The monitoring environment is encoded in a 2D map because it increases situational awareness by providing surrounding environmental context which assists comprehension of spatial relationships between objects [9]. The user display is built using the Google Maps API because it is a familiar interface (often used for directions) and its wide coverage makes it applicable to most outdoor locations. Environmental context is presented through different modalities such as aerial imagery or geographical information system (GIS) type layers depicting structures and areas of interest. The map lets the user know where the monitoring occurs.

B. *Geo-Registration*

Visualization of sensor readings and analysis requires proper alignment with the map coordinates. Sensor coordinates must be transformed into GPS latitude and longitude coordinates in a process called geo-registration. Geo-registration requires calibration between the sensor space and the map space. Simple spot sensors, such as inductive loops, only acquire measurements from a single location which makes the calibration straight forward; the sensor output can be overlayed on the GPS coordinate of the sensor location. It is more difficult to calibrate spatial sensors because of their coverage area. In this case, it is necessary to transform points in the sensor FOV into a corresponding map location.

In order to geo-register a camera, the locations of objects in the image plane and the corresponding latitudes and longitudes on the map need to be known. This is a multi view registration problem. One view of the scene is generated by the camera and the second view is the map (satellite image). Typically, the epipolar constraint can be used to determine the relative pose between the two cameras and solve for the transformation between views. But, since the map is only a 2D representation of the world, full three dimensional mappings are not required. The transformation between the map coordinates and image coordinates reduces to a mapping between 2D planes. This calibration is learned as a homography transformation, H, mapping the image pixel locations on the ground plane (*e.g.* the road) $x_{im} = [x, y]^T$ to its corresponding latitude and longitude coordinates on the map $X_{gps} = [X, Y]^T$

$$X_{gps} = Hx_{im} = Rx_{im} + T. \tag{1}$$

The homography matrix H explains the rotation R and translation T relating the camera and satellite map image and can be found by using a GPS receiver to collect the latitudinal and longitudinal coordinates of specific image locations.

Corresponding points between the map and video were obtained by walking on the street and using an iPhone as a GPS receiver while being recorded by the camera. GPS coordinates were extracted at specific points by remaining still until the GPS reading stabilized. The corresponding

(a) (b)

Fig. 5: Geo-registration calibration with GPS coordinates obtained using an iPhone. (a) Image locations of ground plane calibration points. (b) Google maps satellite image with GPS location of calibration points.

(a) (b)

(c)

Fig. 6: (a) A driver's awareness is limited to what can be seen by the driver. (b) Using infrastructure, situational awareness can be transferred to the driver. The car is warned of the occluded pedestrian tying his shoe on the left side of the road (c) A GPS enabled mobile device can be detected even through visual occlusion in order to relay appropriate safety messages to both the vehicle and pedestrian of the impending crosswalk situation.

image point was manually marked at the point of contact between road and feet. Fig. 5a shows the camera view of Matthews Lane on campus. The aerial image with corresponding GPS points marked is shown in Fig. 5b. Given at least 4 corresponding points, the homography matrix H can be estimated in a least squares by solving the system of equations

$$X^j_{gps} \times Hx^j_{im} = 0 \qquad j = 1, 2, \ldots, n \tag{2}$$

by singular value decomposition using the four-point algorithm [10] ($\times$ denotes vector cross product).

Due to the quality of the GPS receiver, the coordinates coordinates obtained by the iPhone do not fall exactly where expected on the Google road map. The coarse resolution and the narrow strip of road covered by the camera some numerical instability during the mapping from image to map coordinates but will improve with newer GPS sensors.

C. Customization

The Visualization block only presents information to the user when it is needed because complex environments are filled with distracting activities and events. Only those of interest are displayed to minimize visual clutter.

Clickable controls are used to select camera feeds, change environmental context, and display analysis results. Two live feeds may be initialized to view raw video (right side of Fig. 1). The map provides the common visualization space for video analysis and its scale, navigation, and image selection (map layer in Fig. 4a or aerial imagery in Fig. 4b) is controlled by the Google Map API. Toggle buttons overlay Analysis results onto the map and enable information display customization through layer selection. These buttons generate the appropriate SQL commands which removes the need for user training. Figure 4 shows two different classification layers; a classification layer denoting pedestrians and vehicles on campus is shown in Fig. 4a while Fig. 4b shows vehicle tracking.

VI. Wide Area Activity Analysis

By exploring the environment with the map-based representation, activities can be understood within a larger spatial context. The relationships between cameras and monitored objects are contained in a single view to abstract the particulars of a specific location. In Fig. 6a, a campus road is shown as seen from inside a vehicle. The driver's view is limited through the front windshield but with help from infrastructure cameras, the pedestrian behind the vehicle is detected and a warning (yellow bounding box) could be relayed to the driver upon approach (Fig. 6b).

The integration of GPS into mobile devices provides a broader medium for understanding behavior. Using GPS enabled phones, a new stream of trajectory information can be acquired which supplements infrastructure sensing. A pedestrian is tracked through occlusion in Fig. 6c and an alert is sent to the phone warning of the oncoming vehicle. The mobile devices provide a level coverage not feasible using infrastructure. Fig. 7 shows the route of a probe vehicle. The vehicle enables coverage well beyond the extent of the campus network, yet still can be seamlessly integrated in the map interface. The trajectories obtained from mobile devices and automobiles help complete the environment behavior and activity picture [11].

VII. Conclusions

The CANVAS monitoring system provides a unified interface for monitoring of large areas. Live video streams can be selected and viewed by an operator but the focus is on delivering clean computational output that abstracts underlying analysis. The user interface localizes events on a map, which most people are familiar with, for spatial context using simple icons. The icons highlight regions of interest, enabling wider coverage and ultimately ultimately improves the effectiveness of the monitoring by focusing attention through the presentation of only the most relevant information. CANVAS was designed to be scalable in order to accommodate new sensors, analysis processes, and information visualization. With future advances in wireless communication, rather than just providing a webpage, services can be run to provide real-time alerts.

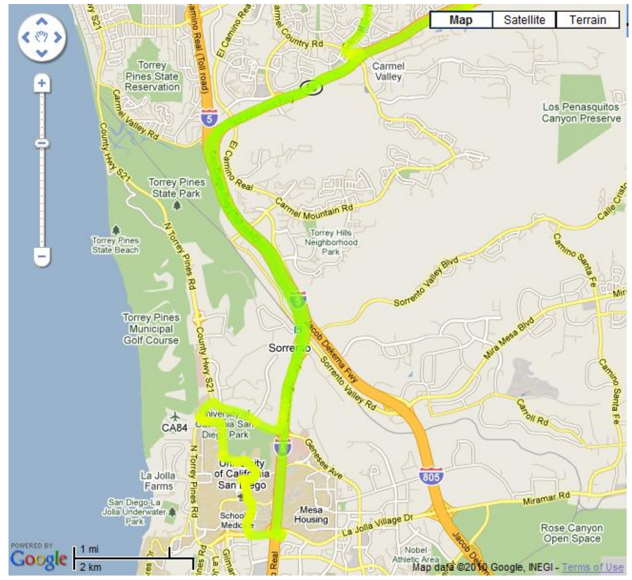


Fig. 7: GPS enabled vehicles and devices are seamlessly integrated into the map. A recorded route taken by a GPS equipped vehicle is overlayed on the map. The route is color coded based on the speed of the automobile with respect to speed limits.

References

[1] H. M. Dee and S. A. Velastin, "How close are we to solving the problem of automated visual surveillance?" *Machine Vision and Applications*, vol. 19, no. 5, pp. 329–343, Oct. 2008.

[2] M. M. Trivedi, T. L. Gandhi, and K. S. Huang, "Distributed interactive video arrays for event capture and enhanced situational awareness," *IEEE Intell. Syst.*, vol. 20, no. 5, pp. 58–66, Sep. 2005.

[3] J. McCall, O. Achler, M. Trivedi, P. F. Jean-Baptiste Haué, D. Forster, J. Hollan, and E. Boer, "A collaborative approach for human-centered driver assistance systems," in *Proc. IEEE Conf. Intell. Transport. Syst.*, Oct. 2004, pp. 663–667.

[4] B. T. Morris and M. M. Trivedi, "Learning, modeling, and classification of vehicle track patterns from live video," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 425–437, Sep. 2008.

[5] (2008) Traffic monitoring guide. U.S. Department of Transportation - Office of Highway Policy Information. [Online]. Available: http://www.fhwa.dot.gov/ohim/tmguide/

[6] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1114–1127, Aug. 2008, Special Issue on Video Surveillance.

[7] B. Morris and M. Trivedi, "Learning and classification of trajectories in dynamic scenes: A general framework for live video analysis," in *Proc. IEEE International Conference on Advanced Video and Signal based Surveillance*, Santa Fe, New Mexico, Sep. 2008, pp. 154–161.

[8] M. A. Goodrich, B. S. Morse, D. Gerhardt, J. L. Cooper, M. Quigley, J. A. Adams, and C. Humphrey, "Supporting wilderness search and rescue using a camera-equipped mini uav," *Journal of Field Robotics*, vol. 25, no. 1-2, pp. 89–110, Jan. 2008.

[9] J. L. Drury, J. Richer, N. Rackliffe, and M. A. Goodrich, "Comparing situation awareness for two unmanned aerial vehicle human interface approaches," The MITRE Corporation, Tech. Rep. 06-0692, Jun. 2006.

[10] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer, 2005.

[11] A. T. Ali and M. M. Venigalla, "Global positioning systems data for performance evaluation of hov and gp lanes on i-66 and i-395/i-95," in *Proc. IEEE Conf. Intell. Transport. Syst.*, 2006, pp. 915 –920.