Vision Modules for a Multi-Sensory Bridge Monitoring Approach*

Remy Chang, Tarak Gandhi, and Mohan M. Trivedi
Computer Vision and Robotics Research Laboratory
University of California at San Diego
{rems,tgandhi,mtrivedi}@ucsd.edu
http://cvrr.ucsd.edu

Abstract— **This paper presents the use of multiple sensor modalities in order to perform traffic analysis for health monitoring of transportation infrastructure. In particular, testbeds containing video and seismic sensors giving complementary information about vehicles are described. Computer vision algorithms are used to detect and track the vehicles and extract their properties. This information is combined with the data from seismic sensors for robust classification of vehicles. Experimental results obtained with our testbeds are described.**

I. INTRODUCTION AND SYSTEM OVERVIEW

The deterioration of the civil infrastructure in North America, Europe and Japan has been well documented and publicized. In the United States, 50 percent of all bridges were built before the 1940's and approximately 42 percent of these structures are structurally deficient [1][2][3]. Since the earthquakes in the Northridge, California (1994), and the Kobe, Japan (1995), there has been a quantum jump in the number of civil structures that have been instrumented for monitoring purposes. Federal Highway Administration has been actively involved in projects concerning bridge and infrastructure monitoring using various types of sensors.

Contact sensors such as strain gauges and accelerometers are being widely used for structural health monitoring. These sensors can provide a temporal signature of the vehicles passing over them that could be used to extract the weight and the effect of the vehicles on the structure with good accuracy. In addition to these sensors, the use of video sensors can give rich information about the shape, size, color, velocity, and track history of vehicles. In fact, the information provided by the contact sensors and video sensors could be combined to improve the reliability of the overall system.

This paper deals with the use of multiple modalities including vision and contact sensors for performing vehicle detection, tracking, quantitative property extraction, and classification. Figure 1 shows the block diagram of the overall system under development. Input from multiple cameras as well as contact sensors can be obtained. To correlate information from various sensors, synchronization issues are addressed using timestamping. Video streams are processed by vision module to perform detection and tracking of vehicles. Properties such as size, speed, and track history are extracted. These are used in conjunction with the information obtained from contact sensors in order to get reliable classification. Since the video data is too large for complete archival, snapshots of vehicles detected by vision module can be stored. The extracted properties of the vehicles can also be stored in the database.

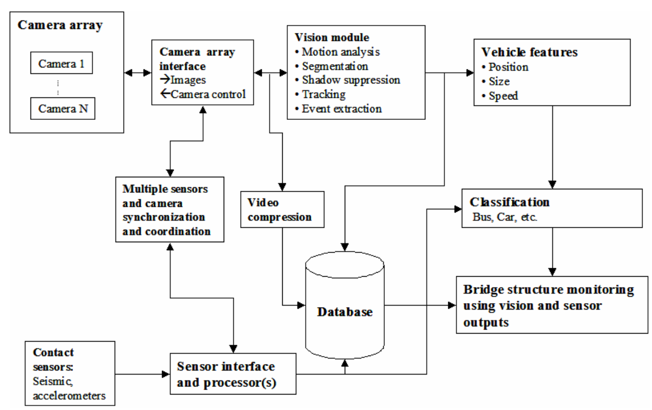


Figure 1 Block diagram for structural health monitoring of bridges

II. VEHICLE DETECTION AND TRACKING

Sensors such as strain gauge sensors and accelerometers typically output 16 bit data with a sampling rate of 200 Hz each. For the testbed used, the total data rate of the sensors is 51.2 kilobits per second. This relatively low data rate allows the strain gauge sensors to be operational at all times, transmitting sensor data to the data storage server. Though the video capture system is capable of recording video at 640x480 at 30 frames per second in MJPEG compressed format, the data rate for this system is 45 megabits a second. This renders long-term data acquisition

* This work was sponsored in part by the NSF-ITR grant.

and transmission over remote links impractical.

As the goal of the project is to study the effect of traffic on the health of bridges and other structures, storage of full motion video is not critical to the analysis of the state of the bridge. It is more useful to process the videos in real-time to provide a source of traffic data in addition to those provided by the strain gauges and accelerometers. In addition, it is useful to store snapshot sequences around interesting events detected by vision system. The algorithm used to accomplish traffic statistics extraction is capable of processing the video stream captured at 30 Hz and can be accelerated with the use relatively simple DSP or FPGA hardware as well as leverage the use of SIMD vector processing available in many modern day general purpose processors.

A. Motion Analysis and Video Segmentation

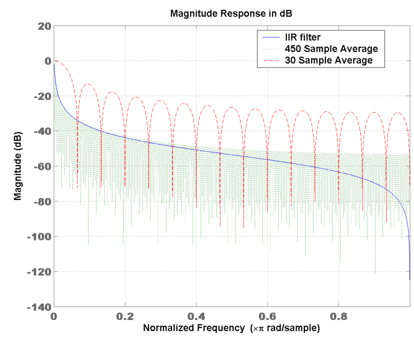
To identify moving objects in a video sequence, it is necessary to identify locations where changes occur in the video scene. A computationally inexpensive method of accomplishing this is through background subtraction for segmentation, where a background image is subtracted from the current video image.

The background is generated using an IIR low-pass filter. The use of an IIR low-pass filter allows for the background to be slowly updated for changes in lighting and scenery while utilizing a small memory footprint as well as having a low computational complexity. Other methods [4][5] using running average and median filtering require substantially more frames to be kept in memory to provide a sufficient estimate of the background. On the other hand, Gaussian mixture models [6] and optical flow analysis [7] are capable of detecting moving objects more precisely, they are computationally intensive in contrast to the method proposed.

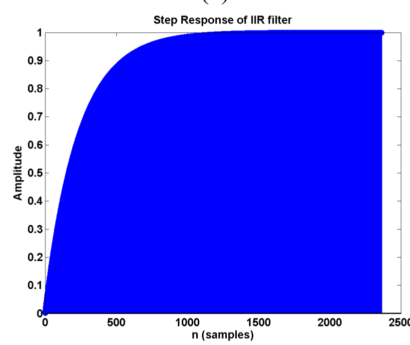
The IIR low pass filter utilized is a 1^{st} order Chebyshev type I low pass filter, implemented in Direct Form II, with a pass-band frequency of 0.06 Hz and ripple of 10 for a sampling rate of 30 frames per second to produce a sharp cutoff at DC that is representative of the mean over the past several hundred frames. Figure 2 (a) illustrates the effectiveness of the filter in isolating the static background in a video sequence as compared to background averages using 30 frames and 450 frames. The IIR filter compares favorably in that the response closely matches that of a 450-frame average. The transfer function for the filter is shown below:

$$H(z) = \frac{a - az^{-1}}{1 - bz^{-1}}$$

The step response of the filter, as shown in Figure 2 (b), indicates that the output will reach unity after 1250 frames. As this will require approximately 40 seconds for the background to fully initialize, the background model has an initialization flag that initializes the delay value to a quarter of the feed forward value a so that the response begins from half the input value instead of a full step. The flag also increases the feedback value. By increasing the feedback b to a value slightly greater than unity and reducing the magnitude of the step, the step response will reach unity within 30 frames. This provides a rough estimate of the background and allows the background to be quickly initialized when quick, drastic changes within the scene can occur, such as with pan, tilt, and zoom cameras. The drastic changes can be detected if there is a large discrepancy between the background estimate and the current frame that persists for several frames throughout the entire image.



(a)



(b)

Figure 2 (a) Magnitude response of IIR low-pass filter in frequency domain (b) Step response of IIR low-pass filter in time domain.

An advantage of utilizing an IIR filter is the relatively inexpensive computational requirements. Only 2 frames are kept in memory to represent the filter taps of a 1^{st} order low-pass filter. The calculations for generating the background are deterministic and computationally inexpensive; requiring only 3 multiplies and 4 accumulates per pixel channel. In the interest of preserving low computational complexity, only the intensity channel is utilized to generate the background image.

B. Segmentation Analysis

The background is subtracted from the incoming frame to generate a difference image. At this point, some shadow suppression is accomplished by introducing an offset to the difference image before taking the absolute value for the difference in intensity. This effectively sets two different levels for the hysteresis thresholds that are utilized to form binary connected blobs. Since shadows represent a negative shift in intensity, the offset will compensate for the shadows cast by objects within the scene.

After the shadow suppressed difference image is acquired, a hysteresis threshold on the magnitude is employed to produce full blobs in the difference image. The threshold levels are set based on observations of the magnitude of the difference image.

After the binary image of blobs has been obtained, morphological closing is used to eliminate spurs. The following constraints are placed on the blobs to select the blobs of interest: (1) Minimum dimension requirement. Since the view of the camera is a road surface, where the vehicles travel vertically in the field of view of the camera, a minimum width constraint decreases the likelihood of pedestrians being mistaken as a vehicle. (2) Minimum size requirement. To make sure that the blob is sufficiently large to be an object of interest. (3) Blob density requirement. Requires the bounding box of the blob to be of specific density in order to eliminate cases of loosely connected blobs. The result of these various thresholds produces an image containing only relatively large connected blobs.

The bounding box for each blob is used as an initial estimate of the location of the vehicle that utilize the mean-shift algorithm [8] to produce a better estimate of the vehicle's size and location. This is based on the assumption that the vehicles are more textured compared to the roadway, thus the bounding box should increase or decrease to fit the contour of the vehicle.

C. Vehicle Tracking

After the bounding box is extracted for each vehicle, pixel locations are remapped using planar homography so that the coordinates are in terms of the ground plane to correct for non-linear movements due to the camera angle. Kalman filtering is used to predict the positions of the remapped centroid and bounding box locations of each vehicle as well as estimate the velocity. Frame-to-frame association is performed by first using the nearest neighbor centroid locations for several frames to allow for the Kalman filter to converge. The closest centroid to the predicted centroid and closest instantaneous velocity to the estimated velocity is then used for association. In the case where a vehicle is not detected in the next frame, the Kalman prediction of the bounding box location and centroid are used in conjunction with the mean-shift algorithm to estimate the next location of the vehicle.

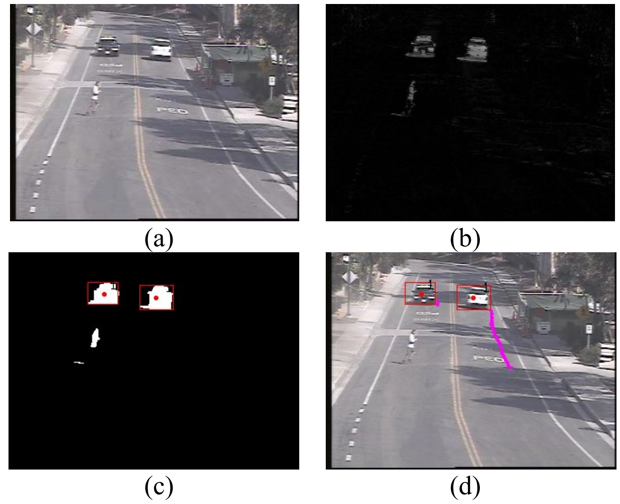


Figure 3 Steps in the implementation of vehicle tracking algorithm: (a) Original frame from the video sequence (b) Output after background subtraction (c) Identification of vehicle regions (d) Tracking over frames using Kalman filter.

D. Application

Figure 4 shows the application with the detection and tracking of vehicles. The snapshots on the side of the vehicles passing the seismic sensors are recorded. The bottom figures show the output from the strain gauge sensors. Note that the current implementation has a delay in getting sensor output; hence the blips due to the vehicle are not visible in the image.

The algorithm for detection and tracking of vehicles was also used on another camera in our testbed that overlooks a freeway. The results are shown in Figure 5.

Figure 4 Snapshot of the vehicle detection and tracking application. The detected vehicles in left and right lanes are identified and snapshots of them shown on the side are archived. The bottom figures show the output from the strain gauge sensors. Note that the current implementation

has a delay in getting sensor output, hence the blip due to the vehicle is not visible in the image.

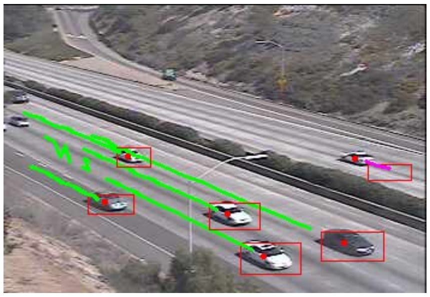


Figure 5 Results of vehicle detection and tracking algorithm on freeway traffic.

III. Multimodal Sensors and Testbeds

To develop and test algorithms for this and various other projects, we are using the testbeds of video as well as seismic sensors installed on our campus. Our laboratory has designed the distributed array of video sensors, whereas the Department of Structural Engineering designed the seismic sensor array.

A. *Distributed Video Array*

An array of video sensors shown in Figure 6 has been installed near the CVRR laboratory [9][10] and is used for a wide variety of computer vision based projects in addition to this project. Several sensor clusters are configured, with two of them located on streetlights near a busy intersection on campus where buses, cars, bikes, and people are regularly in motion. Other sensors are mounted on a building with coverage of nearby roads and courtyards. One of the clusters has a direct view of a freeway 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.

B. *Seismic Sensor Array*

To monitor the long-term performance under usual traffic loading conditions, the Department of Structural Engineering at UCSD has installed three fiber-reinforced polymer (FRP) bridge-deck panels along the same roadway [11][12] as shown in Figure 7. The composite deck is mounted with 16 strain gauge sensors. In addition to this, mounting sensors on a campus bridge is proposed.

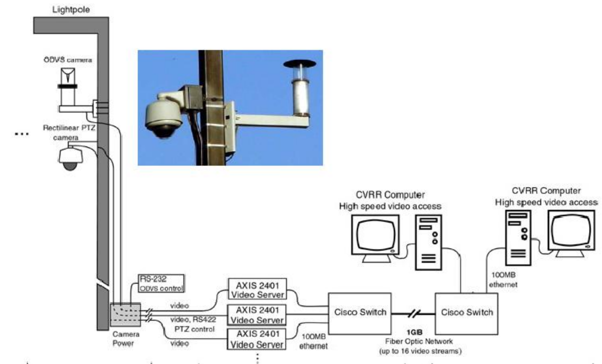


Figure 6: Outdoor Video Clusters provide continuous streaming video of roads and other areas over standard or high-speed networks.

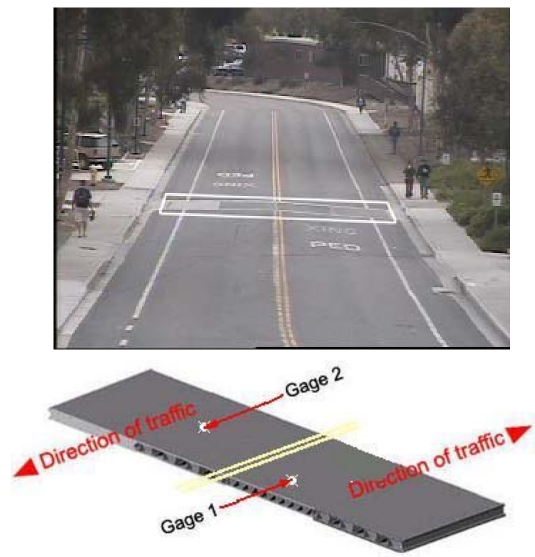


Figure 7: The composite deck panel testbed [12].

C. *Capture and Synchronization Issues*

The block diagram of the video and seismic event capture system is shown in Figure 8. The video camera that provides the view of the testbed in Figure 7 is used to capture video in a time synchronized manner with the strain gauge capture system by synchronizing the system clocks of the separate capture systems that control the camera and the strain gauge sensors and adding a timestamp to the data as soon as samples have been acquired by each system. The system is modularized into data capture components and the data processing component. This allows incremental improvements to be made to each component without directly affecting other parts of the system. The components include strain gauge data capture, video data capture, data processing, and data archival.

Both of the data capture systems reside on a local network and transmits the data to be processed over a network link while maintaining a local data buffer of the past 15 seconds of data that will be transmitted to the data archival system when an event of interest occurs. The strain gauge data capturing component is able to capture and transmit time-stamped data samples with 100 microsecond precision over the local network with no issues due to the low data rate, where as the video capture system down-samples and compresses the video data to maintain a consistent frame rate. The compression method chosen was lossy JPEG compression due to its frame-to-frame independence and relatively low computational complexity. The frame-to-frame independence is necessary to allow for an accurate history of the past 15 seconds worth of video, but also allows for more straightforward access of the archived data when the data is to be reviewed. A simple bi-directional protocol is implemented between the data analyzer and the video server to transmit the timestamp along with every captured frame and receive triggers that will send the data to be archived to the storage server when necessary.

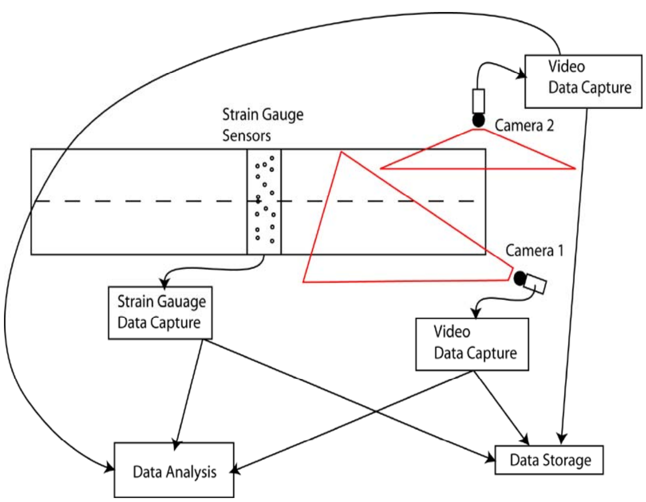


Figure 8 Block diagram of seismic and video event capture system

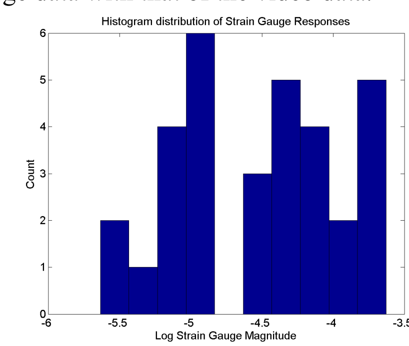
IV. Sensor Fusion and Classification

The classification of vehicles moving over civil infrastructures is a useful feature for studying the effect of vehicles in deteriorating roadways. This can be accomplished with both strain gauge data as well as blob size of objects in the image.

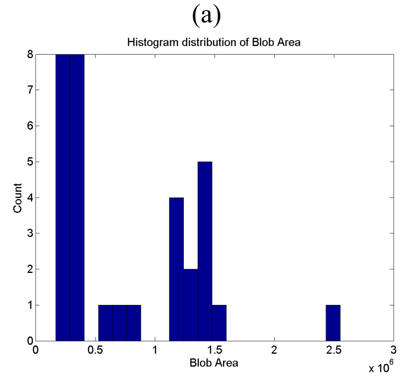
Setting a threshold on the peak strain collected by the sensor can accomplish classification of vehicles. The strain gauge response is higher for heavier vehicles, and relatively low for passenger cars. However, the magnitude response of the strain gauge sensors can vary greatly depending on the load the vehicles are carrying as shown in the histogram distribution in Figure 9a.

For video based classification, the blob area is useful in that it directly measures the size of the vehicle, distinguishing classes based on physical size. The histogram distribution for the sizes of vehicles measured is shown in Figure 9b. However, the accuracy of blob area deteriorates in presence of shadows. Thus, a combination of strain gauge response and blob area is likely to be more robust for classification.

Another issue with video based classification is that camera placement can play a crucial role in the accuracy of classifications of vehicles. As shown in Figure 8, there are two cameras overlooking the roadway that the composite deck lies on. Camera 2, which has a side view, shown in Figure 10 that overlooks the road from a high perspective, is used for vehicle classification. This camera was chosen over the frontal view of Camera 1, shown in Figure 3, since there is less perspective distortion to influence vehicle area measurements. This camera view, however, does not overlook the strain gauge sensors on the composite deck, and currently requires manual registration to correlate the strain gauge data with that of the video data.



(a)



(b)

Figure 9 Histogram of (a) strain gauge responses (b) blob area corresponding to vehicles.

The conditions in which the video data were acquired for two of the three sequences, however, were close to ideal with very little pedestrian traffic as well as little shadows to affect the area of the blobs. In the third sequence, there were some shadows that significantly influenced some of the area measurements as shown in Figure 10. To improve vehicle classification, the strain gauge data can be used in conjunction with the extracted video data. Independently, each vehicle classification metric will fail depending on changing conditions.

(a) (b)

Figure 10 Effect of shadow on vehicle detection and measurement

Analysis of both metrics is first done by correlating the area of the object blob in the image with the peak response of the strain gauge channels by generating a scatter plot of the data associated with each vehicle as shown in Figure 11. The distribution of small vehicle blob area and peak strain gauge magnitude form a tight cluster while large vehicles appear to have large variances in both peak strain gauge response and blob area.

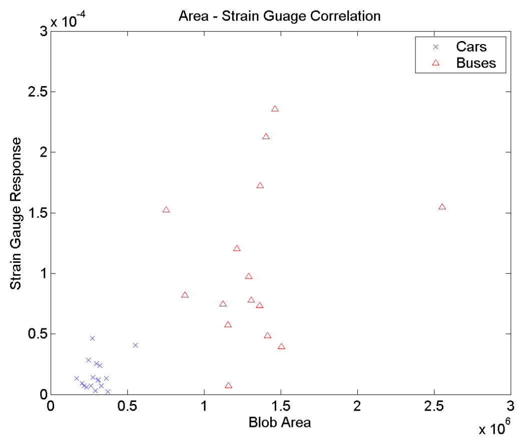


Figure 11 Scatter plot of data associated with each vehicle. X-axis corresponds to the blob area, and Y-axis corresponds to the strain gauge response. Two clusters corresponding to cars and buses (verified ground truth) can be distinguished.

V. SUMMARY

This paper described the use of multi-modal sensors to detect, track, and classify vehicles in the context of structural health monitoring application. Detection and tracking of vehicles using computer vision algorithms was described. The testbeds used for the application were described along with issues such as synchronization. Data from seismic sensors was used to complement the results of detection using vision in order to obtain robust classification of vehicles.

ACKNOWLEDGMENT

We are thankful for the grant awarded by the NSF, which provided the primary sponsorship of the reported research. We would like to thank our collaborators Prof. Ahmed Elgamal and Prof. Joel Conte along with their research groups from the Department of Structural Engineering, UCSD. In particular we thank Mike Fraser for his valuable assistance on data collection. In addition, we thank our colleagues from the CVRR laboratory for their contributions and support.

REFERENCES

[1] A. Elgamal, J. P. Conte, S. Masri, M. Fraser, T. Fountain, A. Gupta, M. M. Trivedi, M. El Zarki, "Health Monitoring Framework for Bridges and Civil Infrastructure," *4th International Workshop on Structural Health Monitoring*, Stanford University, Stanford, CA, September 15 - 17, 2003

[2] Aktan, E., Chase, S., Inman, D., and Pines, D. Monitoring and Managing the Health of Infrastructure Systems. *Proc. 2001 SPIE Conference on Health Monitoring of Highway Transportation Infrastructure*, March 6-8, 2001.

[3] Catbas, N., Ciloglu, K., Celebioglu, A., Popovics, J., and Aktan. E. 2001. Fleet Health Monitoring of Large Populations: Aged Concrete T-Beam Bridges in Pennsylvania. *6th Annual Int'l Symposium on NDE for Health Monitoring and Diagnostics*, SPIE, Newport Beach, CA USA, March 4-8, 2001.

[4] K. Karmann, A.Brandt, R. Gerl, "Moving object segmentation based on adaptive reference images". *Signal Processing* Vol.5 1990, pp951-954.

[5] A. Lai, N. Yung, "A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence". *Proc. IEEE International Symposium on Circuits and Systems*, Vol. 4 1998, pp241-244.

[6] C. Stauffer, W.E.L Grimson, "Adaptive background mixture models for real-time tracking". *IEEE Conf. Computer Vision and Pattern Recognition*, Fort Colins, CO, June 1999.

[7] B.D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision", *DARPA Image Understanding Workshop*, pp.121-130, 1981.

[8] D. Comaniciu, V. Ramesh, P. Meer, "Real-Time Tracking of Non-Rigid Objects using Mean Shift, ICPR00, Hilton Head Island, South Carolina, 2000.

[9] G. Kogut, M. M. Trivedi, "Real-time Wide Area Tracking: Hardware and Software Infrastructure," 5th International IEEE Conference on Intelligent Transportation Systems, Singapore, September 3-6, 2002, pp. 587–593.

[10] B. Hall, M. M. Trivedi, "A Novel Graphical Interface and Context Aware Map for Incident Detection and Monitoring," *9th World Congress on Intelligent Transport Systems*, Chicago, Illinois October, 2002.

[11] Fraser, M., Elgamal, A., Oliver, K., Conte, J. P., "Data Fusion Application For Health Monitoring." *Proc. of the 1st International Workshop on Advanced Smart Materials and Smart Structures Technology*, Honolulu, HI, January 12-14, 2004.

[12] http://healthmonitoring.ucsd.edu