# A Real-Time Computer Vision System for Vehicle Tracking and Traffic Surveillance

Benjamin Coifman (<u>corresponding author</u>)

(zephyr@eecs.berkeley.edu)

Institute of Transportation Studies

University of California

Berkeley, California, 94720

http://www.cs.berkeley.edu/~zephyr

voice: (510) 848-5121

Fax: (510) 642-1246


David Beymer[1], (Beymer@ai.sri.com)

Philip McLauchlan[2], (P.McLauchlan@ee.surrey.ac.uk)

and Jitendra Malik, (Malik@cs.berkeley.edu)

Dept. of Electrical Engineering and Computer Sciences

University of California

Berkeley, California, 94720-1776

http://www.cs.berkeley.edu/~pm/RoadWatch

---

[1]  D. Beymer is now with SRI International, 333 Ravenswood Ave, Menlo Park, CA 94025
[2]  P. McLauchlan is now with the University of Surrey, School of EE, IT and Math., Guildford, Surrey GU2 5XH, UK

**ABSTRACT:**

Increasing congestion on freeways and problems associated with existing detectors have spawned an interest in new vehicle detection technologies such as video image processing.  Existing commercial image processing systems work well in free-flowing traffic, but the systems have difficulties with congestion, shadows and lighting transitions.  These problems stem from vehicles partially occluding one another and the fact that vehicles appear differently under various lighting conditions.

  We are developing a feature-based tracking system for detecting vehicles under these challenging conditions.  Instead of tracking entire vehicles, vehicle features are tracked to make the system robust to partial occlusion.  The system is fully functional under changing lighting conditions because the most salient features at the given moment are tracked.  After the features exit the tracking region, they are grouped into discrete vehicles using a common motion constraint. The groups represent individual vehicle trajectories which can be used to measure traditional traffic parameters as well as new metrics suitable for improved automated surveillance.  This paper describes the issues associated with feature based tracking, presents the real-time implementation of a prototype system, and the performance of the system on a large data set.

**KEY  WORDS:**

Traffic Surveillance, Wide-Area Detection, Vehicle Tracking, Video Image Processing, Machine Vision

**INTRODUCTION**

In recent years, traffic congestion has become a significant problem. Early solutions attempted to lay more pavement to avoid congestion, but adding more lanes is becoming less and less feasible. Contemporary solutions emphasize better information and control to use the existing infrastructure more efficiently.

The quest for better traffic information, and thus, an increasing reliance on traffic surveillance, has resulted in a need for better vehicle detection such as wide-area detectors; while the high costs and safety risks associated with lane closures has directed the search towards non-invasive detectors mounted beyond the edge of pavement. One promising approach is vehicle tracking *via* video image processing, which can yield traditional traffic parameters such as flow and velocity, as well as new parameters such as lane changes and vehicle trajectories.

Because the vehicle tracks, or trajectories, are measured over a length of roadway, rather than at a single point, it is possible to measure true density instead of simply recording detector occupancy. In fact, by averaging trajectories over space and time, the traditional traffic parameters are more stable than corresponding measurements from point detectors, which can only average over time. The additional information from the vehicle trajectories could lead to improved incident detection, both by detecting stopped vehicles within the camera's field of view and by identifying lane change maneuvers or acceleration/deceleration patterns that are indicative of incidents beyond the camera's field of view. The trajectory data could also be used to automate previously labor intensive traffic studies, such as examining vehicle maneuvers in weaving sections or bottlenecks. The vehicle tracking system can produce individual vehicle data (e.g., spacing, headway, velocity, acceleration), which could lead to better traffic flow modeling and an improved understanding of driver behavior. Finally, our group has demonstrated that the system can extract vehicle signatures and match observations of the same vehicle at multiple detector stations (Huang and Russell, 1998). This signature matching can be used to measure true link travel time and thus, quantify

conditions between widely spaced detectors rather than assuming that local conditions are representative of the entire link.

To be an effective traffic surveillance tool, whether by mimicking loop detectors or actually tracking vehicles, a video image processing system (VIPS) should meet several stringent requirements:

1) Automatic segmentation of each vehicle from the background and from other vehicles so that all vehicles are detected.

2) Correctly detect all types of road vehicles - motorcycles, passenger cars, buses, construction equipment, trucks, etc.

3) Function under a wide range of traffic conditions - light traffic, congestion, varying speeds in different lanes.

4) Function under a wide variety of lighting conditions - sunny, overcast, twilight, night, rainy, etc.

5) Operate in real-time.

Even though a number of commercial VIPS for monitoring traffic have been introduced to the market, many of these criteria still cannot be met.

**State of the Practice**

Most of the commercial VIPS available today are *tripwire* systems which mimic the operation of loop detectors, but they do not track vehicles. That is, they do not identify individual vehicles as unique targets and follow their movements in time distinct from other vehicles. The following detectors are examples of commercial tripwire systems: AUTOSCOPE, CCATS, TAS, IMPACTS and TraffiCam (Hockaday, 1991, Chatziioanou, et al, 1994, Klein & Kelley, 1996, MNDOT, 1997, and Hoose, 1992). The systems typically allow a user to specify several detection regions

in the video image and then the given system looks for image intensity changes in the detection regions to indicate vehicle presence/passage. The comparisons are not computationally intensive and can be implemented on Intel 386 based PC's. The primary advantage of these systems is the ease of placing/replacing detector zones and the fact that there is no need to cut the pavement. Some of these systems use a large number of detection zones to follow successive detector actuations through the image, (e.g., IMPACTS), but they do not *track* vehicles.

Some commercial systems do track vehicles, the so-called "third generation" VIPS, e.g., CMS Mobilizer, Eliop EVA, PEEK VideoTrak, Nestor TrafficVision, and Sumitomo IDET (Chatziioanou, et al, 1994, Klein & Kelley, 1996, MNDOT, 1997, and Nihan, et al, 1995). Generally, these systems use region based tracking, i.e., vehicles are segmented based on movement. Unfortunately, if one moving target (including its shadow) occludes another, the two targets may become *merged* together by the tracking software.

Recent evaluations of commercial VIPS found the systems had problems with congestion, high flow, occlusion, camera vibration due to wind, lighting transitions between night/day and day/night, and long shadows linking vehicles together (Hockaday, 1991, Chatziioanou, et al, 1994, Klein & Kelley, 1996, MNDOT, 1997, and Nihan, et al, 1995). The need for traffic surveillance under ALL conditions has led to research in more advanced video-based vehicle detection. For example, Chao, et al, (1996) have developed an algorithm to differentiate vehicles from shadows. On a larger scale, the FHWA has sponsored a major research effort administered by the Jet Propulsion Laboratory (JPL) to advance wide-area traffic detector technology (JPL, 1997; Condos, 1996). Five VIPS were funded by the JPL project, of which, three were existing commercial products (AUTOSCOPE, CMS Mobilizer, and Nestor TrafficVision). The two remaining systems were produced in university laboratories: Autocolor (Chachich, et al, 1996; Zeng & Crisman, 1996), which uses color features to identify vehicles, segment them from the background image and track them through the camera's field of view; and Roadwatch, the subject of this report.

For some years, our group has been developing a vision-based vehicle tracking system: Koller, et al, (1994a); Koller, et al, (1994b); Beymer, et al, (1997). The system uses video from wayside cameras and processes it curbside, then, the data is transmitted in summary form to a central location such as a Traffic Management Center (TMC) for collation and computation of multi-site parameters (e.g., link travel time). Processing occurs in three stages:

1) Segmentation of the scene into individual vehicles and tracking each vehicle inside a tracking zone to refine and update its position and velocity in 3D world coordinates. The time track of these position estimates yield a vehicle trajectory.

2) Processing the trajectory data to compute local traffic parameters (e.g., flow and lane change frequency). These parameters, together with vehicle signature information (e.g., time stamp, vehicle type, color, shape, position), are communicated to the TMC at regular intervals.

3) At the TMC, local traffic parameters from each site are collated and signature information from neighboring camera sites are processed to compute section parameters such as link travel time and origin-destination counts. The data is then passed on to automated and operator assisted applications.

In this paper, we focus on the first two stages: vehicle segmentation and tracking, and the computation of traffic parameters from the tracking data. First, we present different vehicle tracking strategies from the computer vision literature. Then we focus on feature based tracking as a means to improve detector performance in congestion and difficult lighting conditions. The true wide-area detection yields vehicle trajectories and facilitates new and improved traffic parameters. The remainder of the paper presents the testing results from a real-time prototype on 44 lane-hours of data.

## VEHICLE TRACKING STRATEGIES

Multi-object tracking and data association have received considerable attention in the computer vision field and much of the background work has been in non-transportation applications. From the computer vision literature, the different tracking approaches for video data can be classified as follows:

### 3D Model Based Tracking

Three-dimensional model-based vehicle tracking systems have previously been investigated by several research groups, the most prominent being the groups at Karlsruhe (Koller, et al, 1993) and at the University of Reading (Baker and Sullivan, 1992; Sullivan, 1992). The emphasis is on recovering trajectories and models with high accuracy for a small number of vehicles. The most serious weakness of this approach is the reliance on detailed geometric object models. It is unrealistic to expect to be able to have detailed models for all vehicles that could be found on the roadway.

### Region Based Tracking

In this approach, the VIPS identifies a connected region in the image, a 'blob', associated with each vehicle and then tracks it over time using a cross-correlation measure. Typically, the process is initialized by the background subtraction technique. A Kalman filter-based adaptive background model (Karmann and Brandt, 1990; Kilger, 1992) allows the background estimate to evolve as the weather and time of day affect lighting conditions. Foreground objects (vehicles) are detected by subtracting the incoming image from the current background estimate, looking for pixels where this difference image is above some threshold and then finding connected components.

This approach works fairly well in free-flowing traffic. However, under congested traffic conditions, vehicles partially occlude one another instead of being spatially isolated, which makes

the task of segmenting individual vehicles difficult. Such vehicles will become grouped together as one large blob in the foreground image. Figure 1A illustrates this phenomena on a hypothetical one dimensional roadway (i.e., no width) viewed from the camera's perspective, henceforth referred to as the image plane. The vehicles have finite length, hence the trajectories are shown as 'thick' bands in the time space plane. By time $t_3$, vehicle 2 partially occludes vehicle 1. Region based tracking would erroneously merge the two 'blobs' together at this point.

## Active Contour Based Tracking

A dual to the region based approach is tracking based on active contour models, or *snakes*. The basic idea is to have a representation of the bounding contour of the object and keep dynamically updating it. The previous system for vehicle tracking developed in our group, Koller, et al, (1994a); Koller, et al (1994b), was based on this approach. The advantage of having a contour based representation instead of a region based representation is reduced computational complexity.

However, the inability to segment vehicles that are partially occluded remains. If one could initialize a separate contour for each vehicle, then one could track even in the presence of partial occlusion (Koller, et al, 1994a). However, initialization is the difficult part of the problem! Consider the example in Figure 1A: if the vehicles enter the detection region partially occluded, the system will group two vehicles into a single object and this will result in significant measurement errors.

## Feature Based Tracking

An alternative approach to tracking abandons the idea of tracking objects as a whole and instead, tracks sub-features such as distinguishable points or lines on the object. The advantage of this approach is that even in the presence of partial occlusion, some of the features of the moving object remain visible. Furthermore, the same algorithm can be used for tracking in daylight, twilight or

night-time conditions; it is self-regulating because it selects the most salient features under the given conditions[1].

Figure 1B shows feature tracking for the same two vehicles in the earlier example. Individual features are highlighted at three instants in time and the lines indicate their respective tracks[2]. For illustration, the features from different vehicles are shown with different symbols, but, in practice, the features would be indistinguishable at the tracking level. By $t_3$, some of the features from the first vehicle are occluded and lost, however, other features from this vehicle remain and continue to be tracked.

While detecting and tracking vehicle features makes the system more robust to partial occlusion, a vehicle will have multiple features. This introduces a new problem of grouping, i.e., what set of features belong to the same vehicle? To address this problem we use a common motion constraint; features that are seen rigidly moving together are grouped together.

Returning to the simple example once more, applying a common motion constraint to the features in Figure 1B and collecting the feature tracks into discrete vehicles yields Figure 1C. The open circles in this figure denote features that were lost to occlusion at some point in their track, and thus, not included in the final grouping.

**FEATURE BASED TRACKING ALGORITHM**

This section presents our vehicle tracking system, which includes: camera calibration, feature detection, feature tracking, and feature grouping modules. First, the camera calibration is conducted once, off-line, for a given location and then, the other modules are run continuously on-line in real-time.

---

[1] E.g., window corners, bumper edges, etc. during the day and tail lights at night.
[2] To avoid confusion, "trajectory" will be used when referring to entire vehicles and "track" will be used when referring to vehicle features.

**Off-Line Camera Definition**

Before running the tracking and grouping system, the user specifies camera-specific parameters

off-line. These parameters include:

1) line correspondences for a projective mapping, or homography, as explained below,

2) a detection region near the image bottom and an exit region near the image top, and

3) multiple fiducial points for camera stabilization.

Since most road surfaces are flat, the grouper exploits an assumption that vehicle motion is

parallel to the road plane. To describe the road plane, the user simply specifies four or more line or

point correspondences between the video image of the road (i.e., the image plane) and a separate

'world' road plane, as shown in Figure 2. In other words, the user must know the relative

distance in world coordinates between four points visible in the image plane. Ideally, this step

involves a field survey; however, it is possible to approximate the calculations using a video tape

recorder, known lane widths and one or more vehicles traveling at a constant velocity. The vehicle

velocity can be used to measure relative distance along the road at different times and the lane

widths yield relative distance between two points on the edge of the road, coincident with the

vehicle's position.

Based on this off-line step, our system computes a projective transform, or homography,

*H*, between the image coordinates *(x,y)* and world coordinates *(X,Y)*, (Figure 2). This

transformation is necessary for two reasons. First, features are tracked in world coordinates to

exploit known physical constraints on vehicle motion (e.g., finite acceleration). Second, the

transformation is used to calculate distance based measures such as position, velocity and density.

Once the homography has been computed, the user can specify the detection region, exit region

and fiducial points in the image plane.

**On-Line Tracking and Grouping**

A block diagram for our vehicle tracking and grouping system is shown in Figure 3.  First, the raw

camera video is stabilized by tracking manually chosen fiducial points to subpixel accuracy and

subtracting their motion from the entire image.  Second, the stabilized video is sent to a detection

module, which locates corner features in a detection zone at the bottom of the image.  In our

detection module, "corner" features are defined as regions in the gray level intensity image where

brightness varies in more than one direction.  This detection is operationalized by looking for

points in the image, $I$, where the rank of the windowed second moment matrix, $\nabla I \cdot \nabla I^{T}$, is two

(see Beymer, et al, 1997).  Figure 4A shows some example corners detected by the system.

Next, these corner features are tracked over time in the tracking module.  The tracking

module uses Kalman filtering (Gelb, 1974) to predict a given corner's location and velocity in the

next frame, $\left(X, Y, \dot{X}, \dot{Y}\right)$, using world coordinates.  Normalized correlation[3] is used to search a

small region of the image around the estimate for the corner location.  If the corner is found, the

state of the Kalman filter is updated; otherwise, the feature track is dropped.  Figure 4B shows the

temporal progression of several corner features in the image plane.

Vehicle corner features will eventually reach a user defined exit region that crosses the

entire road near the top of the image (or multiple exit regions if there is an off ramp).  Once corner

features reach the exit region, they are grouped into vehicle hypotheses by the grouping module,

e.g., Figure 4C.  The grouper uses a common motion constraint to collect features into a vehicle:

corner features that are seen as moving rigidly together probably belong to the same object.  In

other words, features from the same vehicle will follow similar tracks and two such features will

be offset by the same spatial translation in every frame.  Two features from different vehicles, on

the other hand, will have distinctly different tracks and their spatial offset will change from frame

to frame.  A slight acceleration or lane drift is sufficient to differentiate features between most

vehicles; note that both lateral and longitudinal motion are used to segment vehicles.  Thus, in

---

[3]  For normalized correlation, a 9x9 template of each corner is extracted when the corner is first detected.

order to fool the grouper, two vehicles would have to have identical motions during the entire time they were being tracked.  Typically, the tracking region is on the order of 100 m along the road.  In congested traffic, vehicles are constantly changing their velocity to adjust to nearby traffic and remain in the field of view for a long period of time, giving the grouper the information it needs to perform the segmentation.  In free flowing traffic, vehicles are more likely to maintain constant spatial headways, or spacings, over the short period of observation, making the common motion constraint less effective.  Fortunately, under free flow conditions, drivers take larger spacings (in excess of 30 m), so a spatial proximity cue is added to aid the grouping/segmentation process.

The grouper considers corner features in pairs.  Initially points A and B that are less than a prespecified distance, $r_o$, apart will be hypothesized to belong to the same vehicle.  By monitoring the distance, $d(t) = p_A(t) - p_B(t)$, between the points, this hypothesis can be dismissed as soon as the points are found to move relative to each other.  The distance, $d$, is measured in the world coordinates by multiplying the image distance with a depth scaling factor computed from the homography.  More details can be found in Beymer, et al (1997).

Because features must share a common motion to be grouped into a vehicle, one feature track from each group is selected as being representative of the vehicle trajectory.  In particular, the grouper selects the feature point closest to the camera because it is likely to be near the ground plane and thus, is less likely to suffer from distortions due to the viewing angle.  Finally, traffic parameters such as flow, average speed, and density are computed from the vehicle trajectories.

**REAL-TIME SYSTEM**

We have implemented the vehicle tracker on a network of 13 Texas Instruments C40 digital signal processing (DSP) chips.  The computationally heavy operations in the tracking algorithm, convolution in the feature detector and correlation in feature tracker, are placed on the C40 network, while the grouper is run on the host PC.  Running the grouper on the PC is necessitated

by memory requirements; the grouper needs to store feature tracks, which would quickly exhaust the limited memory available on the C40 modules.  Keeping the grouper on the PC is also beneficial from a load balancing perspective, as the PC is a 150 MHz Pentium and thus, equivalent to 3 to 4 C40's.

The performance of the vehicle tracker is 7.5 Hz in uncongested traffic, dropping to 2 Hz in congested traffic (i.e., near jam density), where many more feature tracks are in progress at any given time.  This reduction in computational speed does not of itself lead to a reduction in performance of the vehicle tracker.  Vehicle speeds in congested traffic are slower, and so the required tracking rate is reduced.

## MEASURING TRAFFIC PARAMETERS USING AN WIDE-AREA DETECTOR

Traditional traffic parameters such as flow, occupancy and velocity are usually defined with respect to point detectors and they are only averaged over time.  The vehicle tracker extracts vehicle trajectories over a significant distance and is a true wide-area detector.  Using the traditional point based definitions would mean discarding information available from the detector.  Instead, it is possible to measure the generalized traffic parameters over the entire tracking region and thus, average over time and space.

Consider a region, A, in the time-space plane with $m$ vehicle trajectories passing through it, as shown in Figure 5.  Following Edie (1963), let $d(A)$ be the sum of the distance traveled by all vehicles in region A and $t(A)$ be the sum of the time spent by all vehicles in region A (i.e., the vertical and horizontal components of the trajectories within A, respectively).  Finally, let $|A|$ denote the 'area' of region A, $L \times T$, then, the generalized flow, $q(A)$, density, $k(A)$, and velocity, $v(A)$, are defined as:

$$q(A) = \frac{d(A)}{|A|}; \quad k(A) = \frac{t(A)}{|A|}; \quad v(A) = \frac{q(A)}{k(A)} = \frac{d(A)}{t(A)} \tag{1}$$

For the real-time system, the generalized parameters are measured for each lane using the entire tracking region, *L*, and user defined sample period, *T*. The definitions are robust to lane changes[4], e.g., vehicle 3 leaving the lane and vehicle 5 entering the lane as shown in Figure 5.

Refining traditional metrics is only one aspect of the vehicle tracker. The vehicle trajectories allow for new approaches to measuring traffic and ultimately, better automated surveillance. Figure 6 shows the time space diagram for vehicle trajectories from a single lane as a shock wave passes through the detection region (solid lines indicate the vehicle tracker output and dashed lines show manually generated ground truth (Coifman, 1997)). Note that horizontal trajectories correspond to stopped vehicles and the tracker did not lose them; also notice that the tracker followed almost all of the vehicles even though the traffic momentarily reached jam density and most of the vehicles were partially occluded. This figure only shows the longitudinal position along the roadway, the vehicle tracker also extracts the lateral position. Thus, it is a trivial matter to detect lane changes within the surveillance region.

After the tracker has extracted the vehicle trajectories, it is possible to send this data over a low bandwidth communications link for scene reconstruction and automated surveillance at a remote location. It is also possible to extract a vehicle signature for vehicle reidentification at a downstream site and measure section data such as travel time and O/D patterns (see Huang and Russell, 1998); thus, making it possible to quantify conditions between detectors even if there are no observable effects at the individual detector stations.

---

[4] Note, the precise lane change location depends on which feature's track is selected to represent the vehicle trajectory. The important property of these measures is that at any instant, a given vehicle will only be counted in a single lane.

**TESTING  RESULTS**

The tracking and grouping system has gone through two major phases of testing.  First, the system was tested off-line, using pre-digitized video sequences, largely for development purposes.  This testing gave us a 'microscopic' view of the system, allowing us to analyze errors such as false detections, false negatives, and overgroupings.  For space considerations, detailed results are not presented here, the interested reader is referred to the full report, Malik, et al (1997).  Second, the real-time system was tested on a large data set to see if the system could accurately measure aggregate traffic parameters.

**On-Line  Testing  of  Traffic  Parameters**

The second phase of testing evaluated the on-line system's ability to measure aggregate traffic parameters.  In particular, we tested flow, average velocity and density.  These parameters are computed separately for each lane of traffic and are sampled over a user specified sample period (taken to be 5 minutes in the included examples).

Generating manual ground truth data is very time consuming and labor intensive.  It was decided to use existing loop detectors to verify the vehicle tracker over a large data set.  Ground truth data came from inductive loop speed trap data collected concurrently with the video data. Each lane of traffic has two loops separated by 6.1 m (20 ft) and vehicle arrivals were recorded every 1/60 seconds.  Although a paired loop speed trap is usually operated as a single point detector, the testing methodology used the vehicle arrival data to apply Edie's method to the speed trap data.  The vehicle tracker followed vehicles for approximately 100 m, but, to ensure that video and loop data are compatible, traffic parameters were only measured from the video data for the 6.1 m over the speed trap.  Thus, referring to Figure 5, $L$=6.1 m and $T$=5 minutes.  It was recognized that loop detectors can produce several errors, therefore, a random sample of data[5] was manually

---

[5]  The total manual sample size was 1.75 lane-hours

calibrated to verify the loops were in good operating condition: yielding accurate velocity measurements without over-counting or under-counting vehicles.

The vehicle tracker was tested on approximately 44 lane-hours of video from the Florin Road interchange along State Highway 99 in Sacramento, Calif. (see Figures 4B-4C for sample frames). The test set includes all observed operating conditions: day, night, twilight, long shadows and rain; congestion and free flow. Lane 1, on the left, is a carpool (HOV) lane and exhibited little if any congestion. Lane 3, on the right, exhibited some degree of congestion for approximately 20% of the tests. Finally, the loops in the center lane were bad and the lane was excluded from the final analysis. The final test set included approximately 40,000 vehicle observations which were distributed over 514 samples.

Comparing the two detection systems, Figure 7 shows scatter plots of the tracker output versus concurrent loop data for velocity and flow. Table 1 summarizes the error distribution for velocity, flow and density.

As one would expect from a feature based tracker, the measured velocity is very accurate. Even if the vehicle tracker overgroups or oversegments vehicles, the erroneous blobs still move at the prevailing speed. The errors in flow and density are primarily due to missed or oversegmented vehicles. Often, an error of two or three vehicles in one sample can be very significant. For example, one missed vehicle in a five minute sample at 1,000 veh/hr results in a 2% error. At the mean flow for the data, 910 veh/hr, the error per missed vehicle is slightly higher, at 2.2%.

The greatest source of error appears to be from the frame grabber adding noise to the frames. As a result, the vehicle tracker drops feature points because the Kalman filter loses noisy points or the grouper oversegments vehicles because relative distance is not preserved with the noisy feature tracks. Off-line testing with pre-digitized sequences showed significant performance improvements. The next generation of the vehicle tracking system will use an improved frame grabbing routine and should show significant improvement over these results.

The results are promising nonetheless and the error distribution in Table 1 should be taken in context. The tests emphasized difficult conditions where earlier image processing systems break

down: congestion, long shadows linking vehicles together, and the transition between night and day (Chatziioanou, et al, 1994). The tests consisted of 2 hour samples and half of the test segments included a night-day (or day-night) transition. All of the night-day transition sequences included long shadows that spread across one or more lanes, both from vehicles and from wayside objects. Approximately 20 percent of the samples included rain and 20 percent of the segments included severe camera movement due to high winds.

Performance did not show any significant changes under different conditions. Space constraints prevent examining all conditions here (see Malik, et al (1997) for a detailed study); instead we will highlight the vehicle tracker performance under two of the more challenging conditions: the night to day transition with long shadows during the daytime portion. Consider the two hour sequence starting at night (5:30 AM, Figure 8A), progressing through sunrise and long shadows, and ending with daylight conditions (7:30 AM, Figure 8B). Figure 9 shows measured and ground truth $q$ and $v$ for the sequence. There are 48 samples of 5 minutes each, with a total of roughly 4,600 vehicles. Note that the morning peak starts during the sequence and approximately 30 minutes of data from lane 3 are under light congestion, and thus, frequent occlusions.

## CONCLUSIONS

Recent evaluations of commercial VIPS found the existing systems have problems with congestion, occlusion, lighting transitions between night/day and day/night, camera vibration due to wind, and long shadows linking vehicles together. We have presented a vehicle detection and tracking system that is designed to operate under these challenging conditions. Instead of tracking entire vehicles, vehicle features are tracked, which makes the system less sensitive to the problem of partial occlusion. The same algorithm is used for tracking in daylight, twilight and nighttime conditions, it is self-regulating by selecting the most salient features for the given conditions. Common motion over entire feature tracks is used to group features from individual vehicles and

reduce the probability that long shadows will link vehicles together. Finally, camera motion during high wind is accounted for by tracking a small number of fiducial points.

The resulting vehicle trajectories can be used to provide traditional traffic parameters as well as new metrics such as lane changes. The trajectories can be used as input to more sophisticated, automated surveillance applications, e.g., incident detection based on acceleration/deceleration and lane change maneuvers. The vehicle tracker is well suited both for permanent surveillance installations and for short term traffic studies such as examining vehicle movements in weaving sections. The vehicle tracking system can also extract vehicle signatures to match observations between detector stations and quantify conditions over extended links.

A real-time version of the system has been implemented using a network of DSP chips. The system has been tested on approximately 44 lane-hours of data and has demonstrated good performance under the challenging conditions that have limited earlier VIPS.

## ACKNOWLEDGMENTS

## REFERENCES

Baker, K, Sullivan, G (1992) "Performance assessment of model-based tracking", *Proc. of the IEEE Workshop on Applications of Computer Vision*, Palm Springs, CA, pp. 28-35.

Beymer, D., McLauchlan, P., Coifman, B., and Malik, J. (1997) "A Real-Time Computer Vision System for Measuring Traffic Parameters", *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp 495-501.

Chachich, A, Pau, A, Barber, A, Kennedy, K, Olejniczak, E, Hackney, J, Sun, Q, Mireles, E, (1996) "Traffic Sensor Using a Color Vision Method", *Transportation Sensors and Controls: Collision Avoidance, Traffic Management*, and ITS, SPIE Proc. Vol 2902, pp 156-165.

Chao, T, Lau, B, Park, Y (1996) "Vehicle Detection and Classification in Shadowy Traffic Images Using Wavelets and Neural Networks", *Transportation Sensors and Controls: Collision Avoidance, Traffic Management*, and ITS, SPIE Proc. Vol 2902, pp 136-147.

Chatziioanou, A, Hockaday, S, Ponce, L, Kaighn, S, Staley, C (1994) *Video Image Processing Systems Applications in Transportation, Phase II. Final Report*, California Polytechnic State Univ., San Luis Obispo, CA.

Coifman, B (1997) *Time Space Diagrams for Thirteen Shock Waves*, PATH Report No. UCB-ITS-PWP-97-1, University of California, Berkeley, CA.

Condos, F (1996) "Traffic Surveillance and Detection Systems for Intelligent Transportation Systems Applications", *Proc. 1996 ITE Annual Conference*, Minneapolis, MN, pp 354-356.

Edie, L (1963) "Discussion of traffic stream measurements and definitions", *Proc. Second International Symposium on the Theory of Traffic Flow*, OECD, Paris, France, pp 139-154.

Gelb, A [editor] (1974) *Applied Optimal Estimation*, MIT Press, Cambridge, MA.

Hockaday, S (1991)  *Evaluation of Image Processing Technology for Applications in Highway Operations - Final Report*, California Department of Transportation Technical Report 91-2. California Polytechnic State University.

Hoose, N (1992)  "IMPACTS: an Image Analysis Tool for Motorway Surveillance", *Traffic Engineering and Control*, Vol 33, No 3, March 1992, pp 140-147.

Huang, T, Russell, S (1998)  "Object Identification: A Bayesian Analysis with Application to Traffic Surveillance.", *Artificial Intelligence*, 103, pp 1-17.

JPL (1997)  *Traffic Surveillance and Detection Technology Development, Sensor Development Final Report*,  Jet Propulsion Laboratory Publication No. 97-10.

Karmann, K, Brandt, A, (1990)  "Moving object recognition using an adaptive background memory",  *Time-Varying Image Processing and Moving Object Recognition, 2*, V Cappellini, editor, Elsevier, Amsterdam, The Netherlands.

Kilger, M (1992)  "A shadow handler in a video-based real-time traffic monitoring system", *IEEE Workshop on Applications of Computer Vision*, Palm Springs, CA, pp 1060-1066.

Klein, L, Kelley, M [Hughes Aircraft Company], (1996)  *Detection Technology for IVHS: Final Report*, FHWA Report No. FHWA-RD-95-100.

Koller, D,  Daniilidis, K, Nagel, H, (1993)  "Model-based Object Tracking in Monocular Image Sequences of Road Traffic Scenes",  *International Journal of Computer Vision*, 10: 257-281.

Koller, D, Weber, J, Malik, J (1994a)  "Robust multiple car tracking with occlusion reasoning",
*ECCV*, Stockholm, Sweden, pp 189-196.

Koller, D, Weber, J, Huang, T, Malik, J, Ogasawara, G, Rao, B, Russell, S (1994b)  "Towards
robust automatic traffic scene analysis in real-time.",  *ICPR*, Israel, Vol 1, pp 126-131.

Malik, J., Beymer, D., Coifman, B., Huang, T., Lyddy, D., McLauchlan, P., Russell, S, and
Weber, J. (1997)  *Traffic Surveillance and Detection Technology Development: New Traffic
Sensor Technology Final Report*, PATH Report No. UCB-ITS-PRR-97-6, University of
California, Berkeley, CA.

MNDOT (1997)  *Field Test of Monitoring of Urban Vehicle Operations Using Non-Intrusive
Technologies*, FHWA Report No. FHWA-PL-97-018

Nihan, N., Leth, M., Wong, A. (1995)  Video Image Processing for Freeway Monitoring and
Control: Evaluation of the Mobilizer, Washington State Department of Transportation Report No.
WA-RD 398.1/TNW 95-03

Sullivan, G, (1992)  "Visual interpretation of known objects in constrained scenes.",  *Phil. Trans.
Roy. Soc (B)*, 337: 361-370.

Zeng, N, Crisman, J (1996)  "Evaluation of Color Categorization for Representing Vehicle
Colors", *Transportation Sensors and Controls: Collision Avoidance, Traffic Management*, and
ITS, SPIE Proc. Vol 2902, pp 148-155.

Coifman, Beymer, McLauchlan, Malik

FIGURE 1: A) An example of region based tracking. By time $t_3$, vehicle 2 has partially occluded vehicle 1, resulting in potential segmentation problems. (B) An example of feature based tracking with features shown at three times. Note that individual features are indistinguishable to the tracker, but are shown using different symbols for illustration. (C) Features are grouped together based on common motion, avoiding segmentation problems due to occlusion. Because the features move together, one feature track is selected as being representative of the vehicle trajectory.

Coifman, Beymer, McLauchlan, Malik
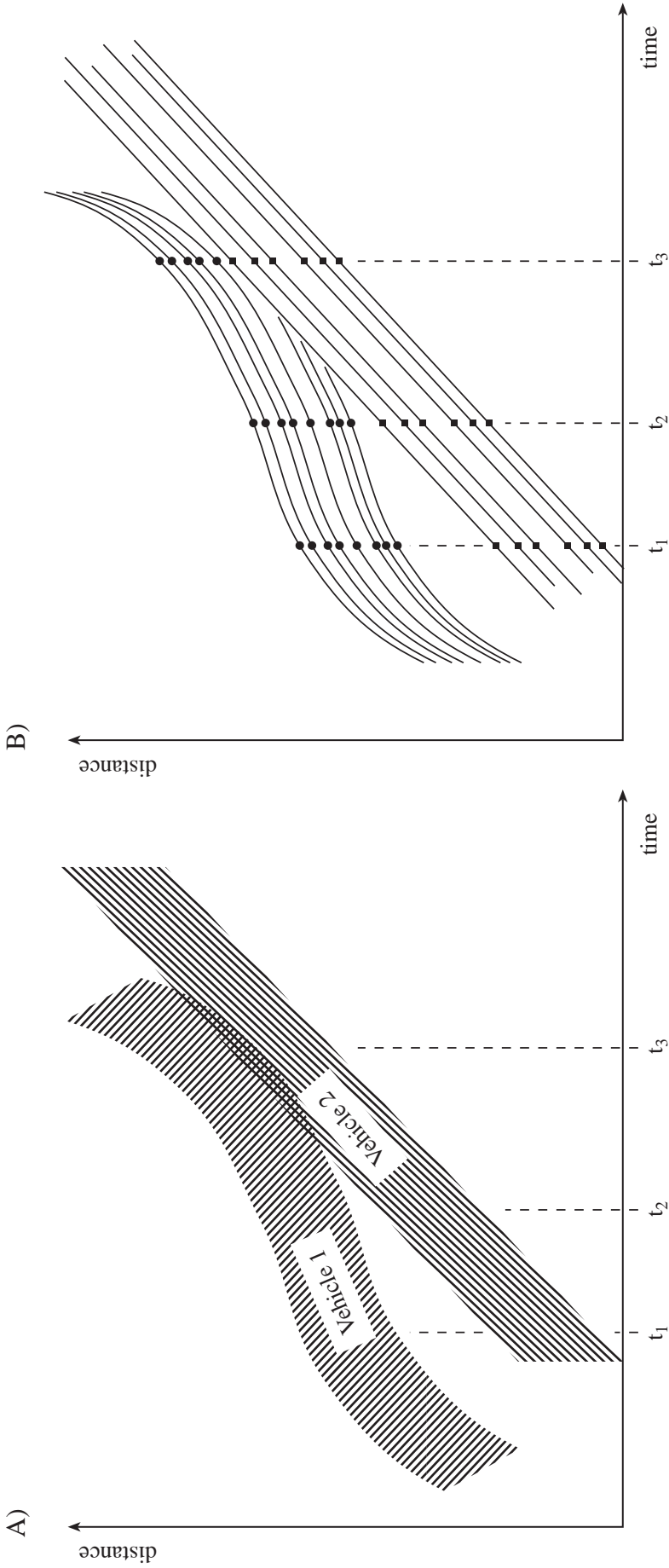
FIGURE 1: Continued...

C)
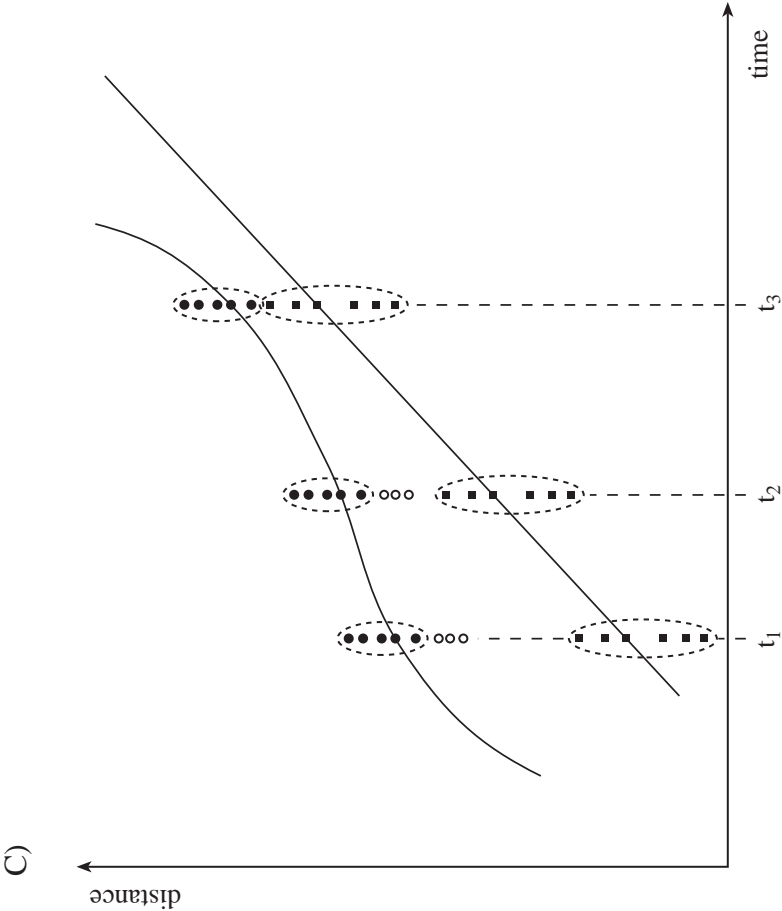
FIGURE 2:    A projective transform, $H$, or homography is used to map from image coordinates, $(x,y)$, to world coordinates, $(X,Y)$.
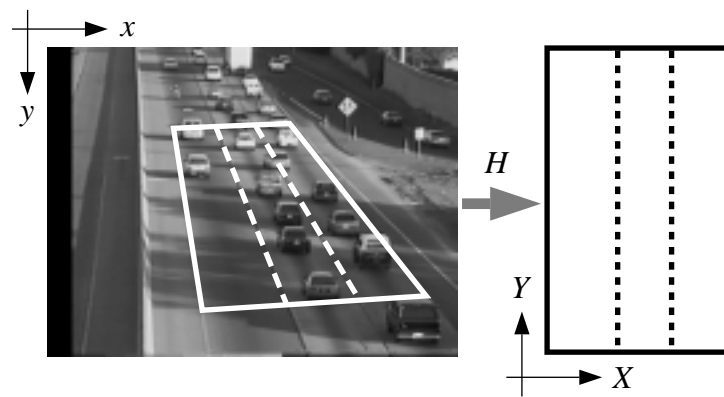
FIGURE 3: Block diagram of the vehicle tracking system. In the future, we plan to add a vehicle classification module, as indicated by the dashed lines.
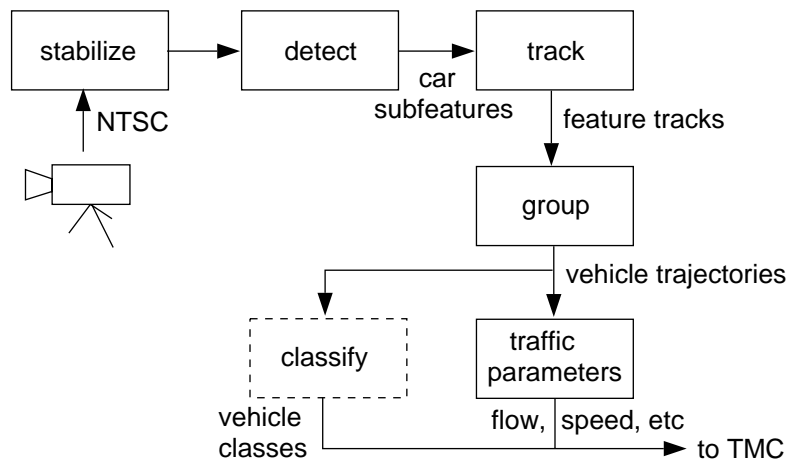
Coifman, Beymer, McLauchlan, Malik

FIGURE 4:    (A) Sample corner features identified by the tracker (B) Sample feature tracks from
            the tracker (C) Sample feature groups from the grouper.

A)



B)

Coifman, Beymer, McLauchlan, Malik

FIGURE 4:     continued...

C)

FIGURE 5: Region, *A*, in the time-space plane, bounded by the length of the detection region, *L*, and user defined sample duration, *T*. This region is used to measure the generalized traffic parameters for a single lane.
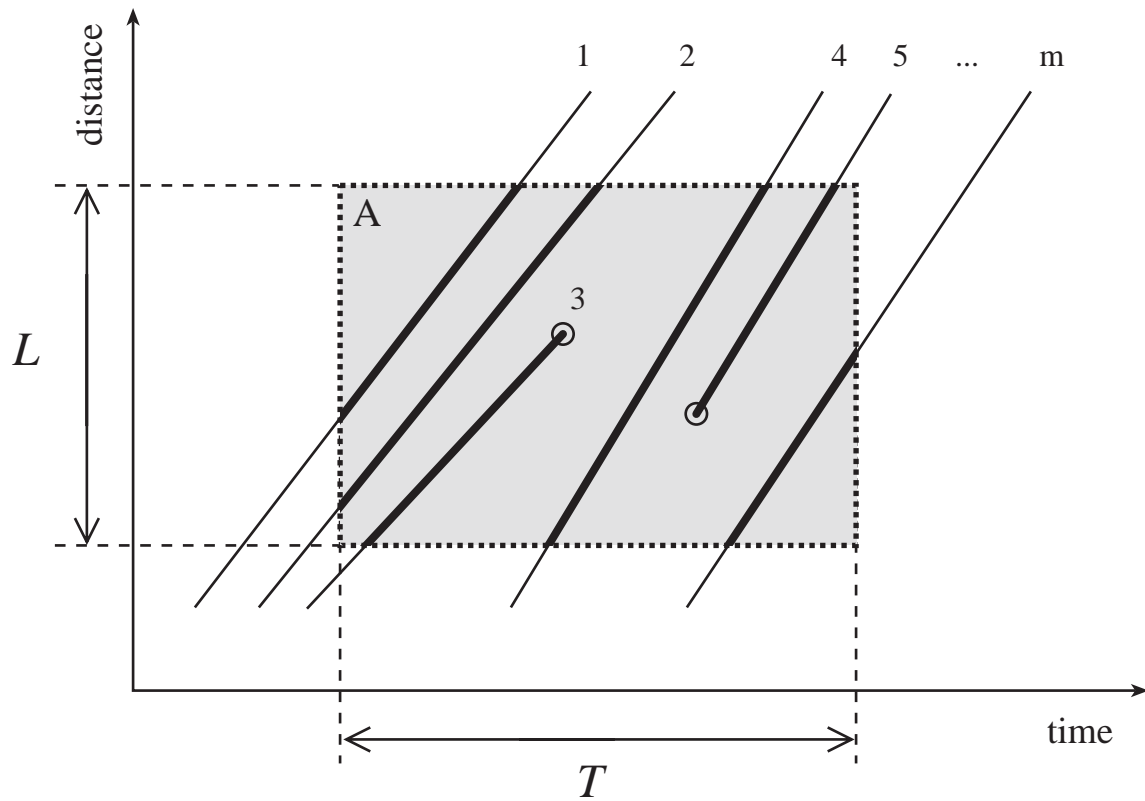
FIGURE 6:    Vehicle tracker output and ground truth as a shock wave passes the surveillance site.
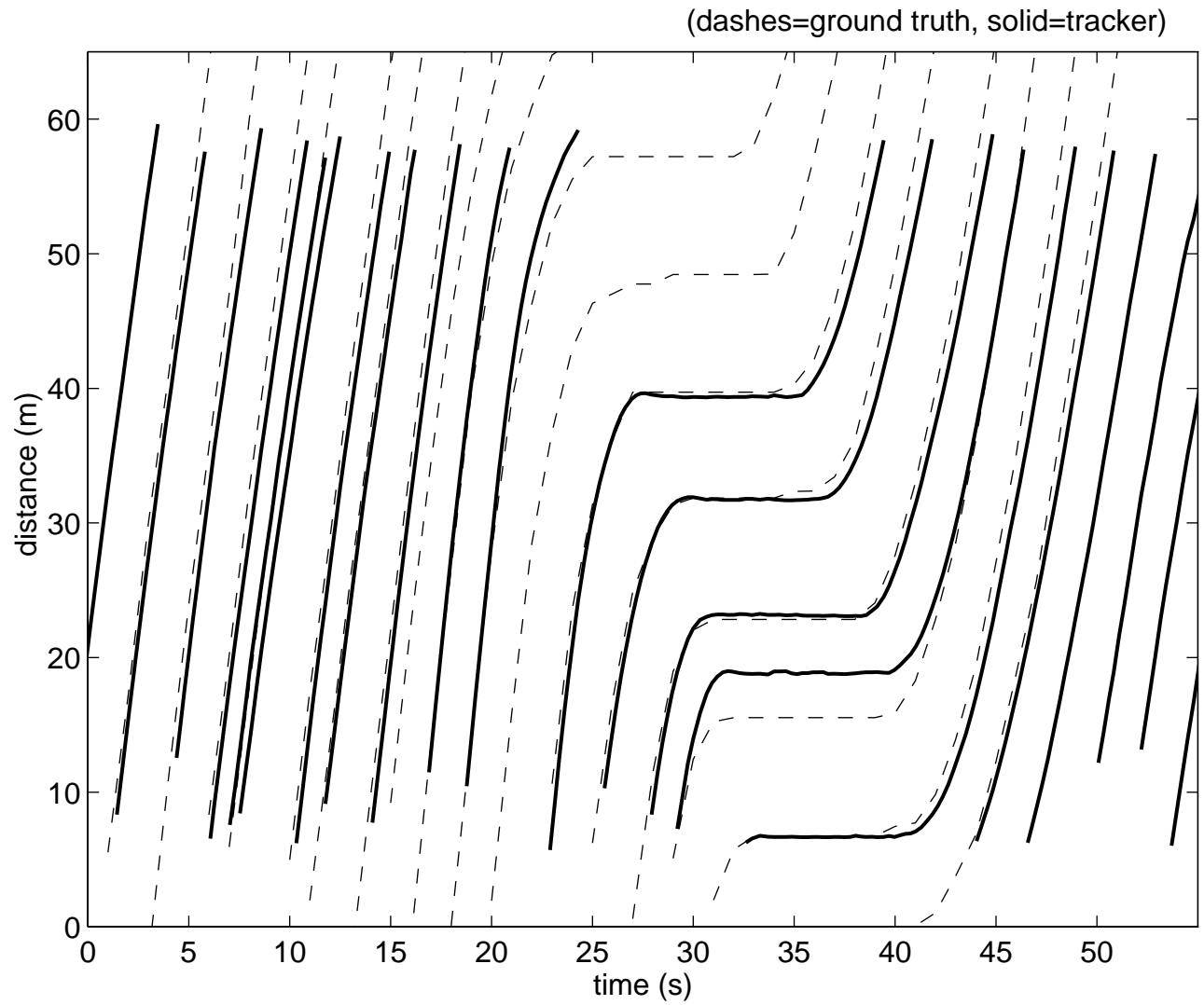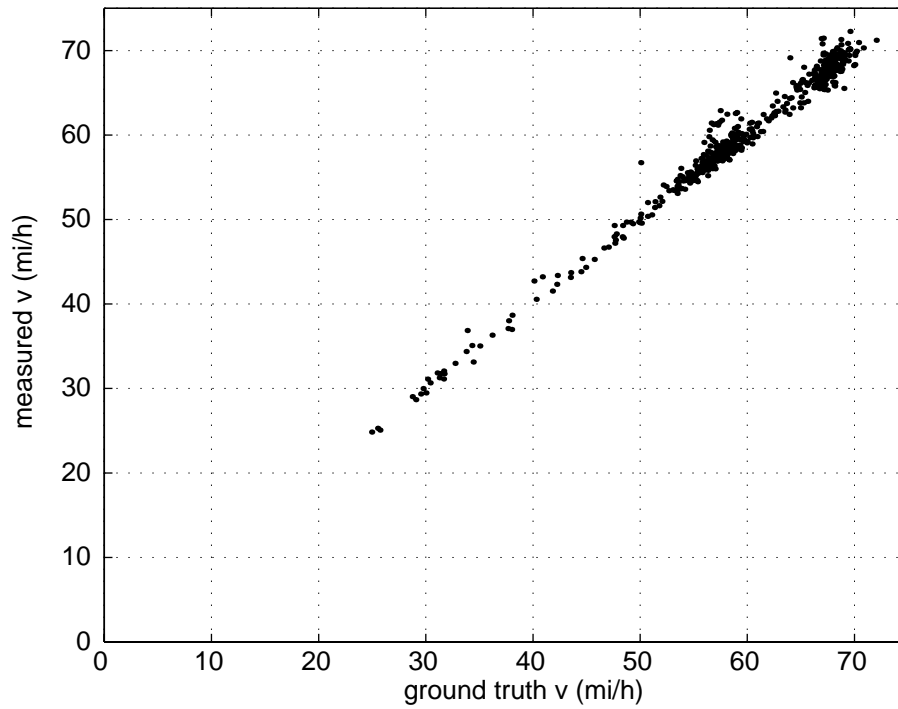
(dashes=ground truth, solid=tracker)

FIGURE 7:    Scatter plots comparing tracker performance to loop detector ground truth, 44
lane-hours of data aggregated into 5 minute samples:  (A) velocity, (B) flow.
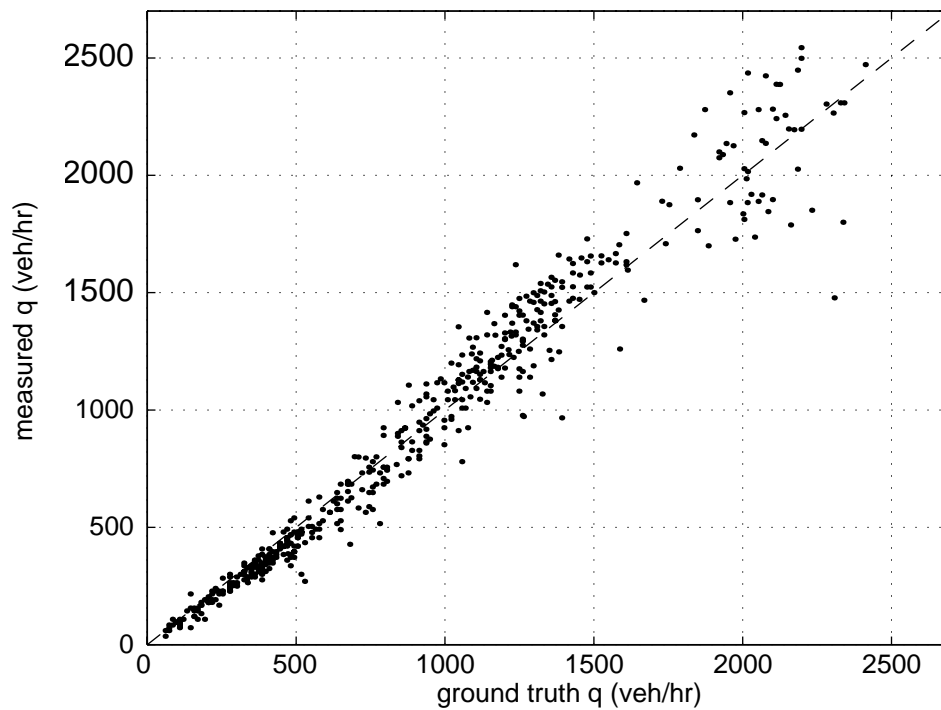
(A)



(B)

FIGURE 8:     Two frames from the start and end of a continuous two hour sequence used to test the vehicle tracker, (A) 5:30 AM, (B) 7:30 AM.
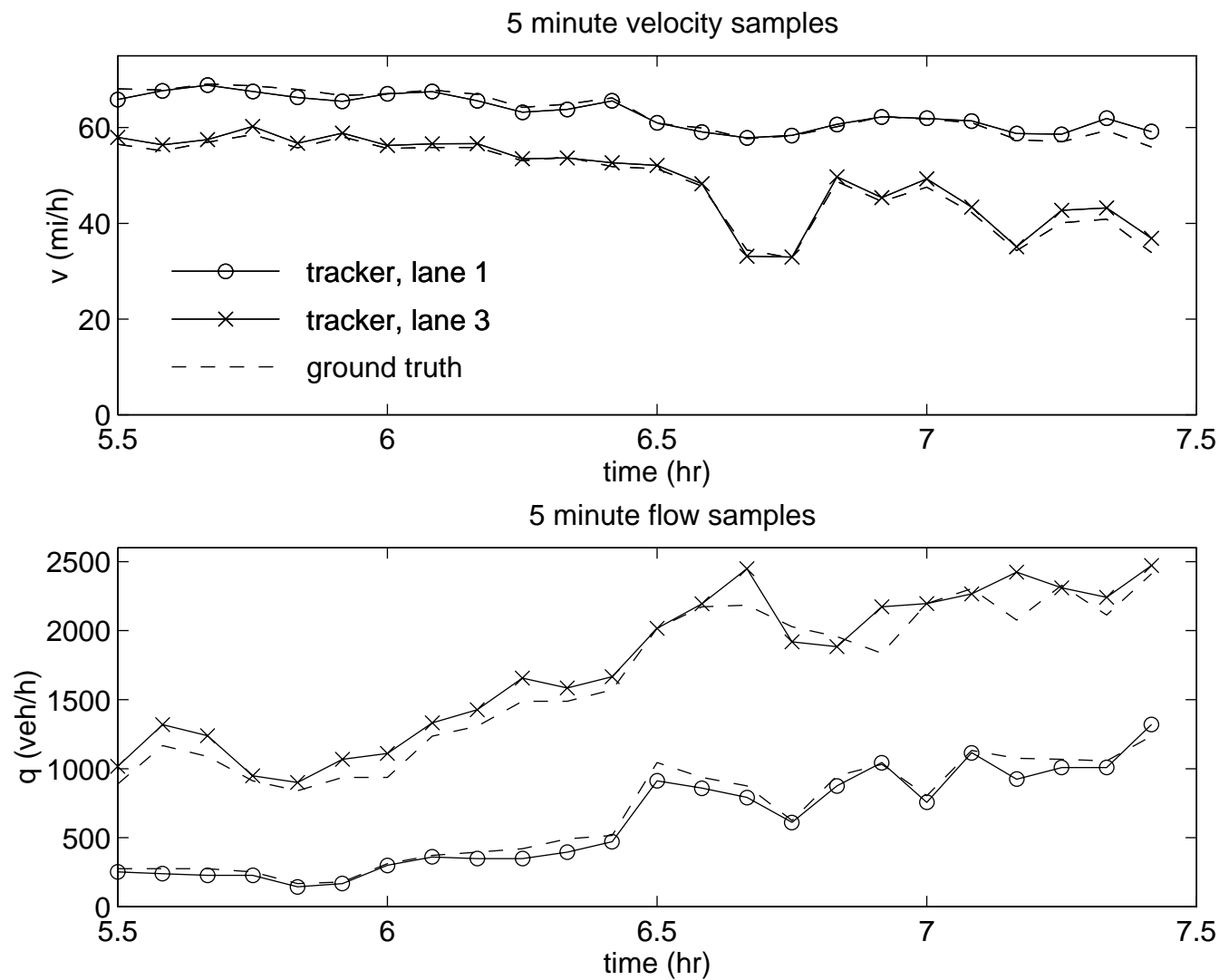
A)

B)

Coifman, Beymer, McLauchlan, Malik

FIGURE 9:   Time series flow and velocity for the two hour night to day transition as shown in Figure 8.

5 minute velocity samples



5 minute flow samples

**TABLE 1: Error distribution for velocity, flow and density over 44 lane-hours of data.**

| % error less than | % velocity samples | % flow samples | % density samples |
|---|---|---|---|
| 2.5% | 86% | 18% | 19% |
| 5% | 95% | 31% | 33% |
| 10% | 100% | 60% | 59% |
| 15% | 100% | 79% | 79% |
| 20% | 100% | 91% | 90% |
| 25% | 100% | 96% | 96% |