A Computer Vision Detection System for Network Model Validation

C. A. MacCarley
Computer Engineering Program
California Polytechnic State University
San Luis Obispo, California, 93407 USA
Phone: 805 7562317
FAX: 805 7561458
Email: cam@ee.calpoly.edu

Brian Hemme
Loragen Corporation, Transportation Electronics Group
3576 Empleo St., Ste 1
San Luis Obispo California, 93401 USA
Phone: 805 7818461
FAX: 805 781 88464
Email: brian@loragen.com

ABSTRACT
This paper describes the development and testing of a computer-vision-based traffic detection system capable of uniquely identifying vehicles in a traffic network from video images, and later re-identifying them at subsequent detection sites in the network. This capability may serve as the basis for deterministic validation of microscopic network flow models, origin-destination tables, and travel-time tables. The system is referred to as the Video-based Vehicle Signature Analysis and Tracking (V²SAT) System.

Using video cameras as primary sensors, detection modules generate a numeric Video Signature Vector (VSV) for each vehicle, and transmit these to a central Internet-connected correlation computer via a public low-power wireless network. The correlation computer attempts to match vehicles, as represented by their VSVs, generated at successive detection sites to enable a sampled real-time microscopic flow representation of the freeway network. Test results indicate a 93.6% accurate ability to correctly re-identify vehicles at successive sites. The tendency of the system to incorrectly match different vehicles at successive sites was measured at 0.0116%.

Potential advantages are low cost in widespread deployment, simplicity and reliability of detection, minimal bandwidth and storage requirements for transmission of the signature vector, and reasonable identification ability without violation of privacy rights. The primary disadvantages are the requirement for the placement of video cameras above each traffic lane, and video imaging limitations related to reliance upon ambient illumination.
BACKGROUND

The use of video sensing and real-time computer vision processing for detection of vehicles on highways is a well-established area of investigation and product development, a few recent examples being (1,2,3,4,5,6). Most detection means are intended to provide either a vehicle presence indication for signal actuation, or traffic flow metrics such as speed, volume, level of service, queue length, or possibly inference of traffic incidents. Areas of current study include the use of neural nets and improved training algorithms to improve the robustness of detection and metric estimation (7,8,9,10). Video imagery is rich with information, which if appropriately processed, has the potential to provide data beyond that derived from simple recognition of a vehicle. In particular, these visual characteristics make possible the estimation of the vehicle class, and possibly some distinguishing features of individual vehicles. Ideally, vehicles might be recognized and characterized to a degree sufficient to permit robust re-identification at subsequent detection sites. Such a capability would admit the possibility of automated collection of travel data on an individual-vehicle basis, which may be of potential value for support of transportation resource planning, roadway engineering, and integrated network-wide traffic management strategies.

In addition to and in support of these strategies, many computer simulation programs have been developed for modeling and prediction of traffic flow patterns. These models generally fall into two classes: macroscopic models, in which vehicle flow is treated as a continuum much like compressible fluid flow, and microscopic models, in which individual vehicle behavior is simulated. This latter class of models, while more sophisticated and more useful in transportation engineering, is much more difficult to validate since individual vehicle lane and turning movements and traveler origin-destination data must be recorded over extended time periods. A vast amount of work has been done on traffic simulation models, some of which is summarized in (11).

Existing data collection techniques for model validation are typically manual or semi-automated in nature, e.g., extrapolation from loop detector data, human observation, floating car studies, and traveler surveys. Accuracy and adequate sample size are known weaknesses, and cost per data unit is often an obstacle. Data is very limited if individual tracking is required for model validation. Analog inductive loop signatures are known to provide crude signatures of vehicles, which are variable between detection sites. Computer-vision-based license plate (character recognition) readers have been generally
unsuccessful when deployed on freeways, and if effective, involve legal risks such as privacy rights issues. Intrusive monitoring means, such as vehicle tags or markers, have been considered unsuitable for large-scale data collection since they require the consent and cooperation of a large number of routine travelers in the test network area (12,13,14).

This paper describes the development and testing of an experimental means for video-based recognition and re-identification of individual vehicles as they progress through a roadway network. The system and underlying computer vision technique are referred to as the Video-based Vehicle Signature Analysis and Tracking (V^2SAT) System. Work on this method and the V^2SAT system has been in progress since 1995 at Loragen Corp. of San Luis Obispo, California. This work builds upon a large and well-established body of existing research, including published algorithms and techniques for target recognition, facial feature characterization for security, and optical character recognition (15,16,17,18). The extension and application of these methods to roadway monitoring was supported by the California Department of Transportation (Caltrans) via the Partners for Advanced Transit and Highways (PATH) Program.

The detection concept involves real-time optical measurements from digitized video images of vehicles, acquired from an overhead (downlooking) camera placement above individual freeway lanes. A conventional color video camera serves as the primary sensor. Detection modules are placed directly above traffic lanes on an overcrossing or similar support structure, with one detector for each lane.

Any detection system capable of recognizing and re-identifying individual vehicles must generate a tag or vector for each vehicle, which uniquely identifies it with an adequate amount of high-quality information. For the V2SAT system, the information packet is derived from computer-vision measurements of selected optical features, and is referred to as a Vehicle Signature Vector (VSV). In view of the possibly large number of targets in track at any time in a network-based system, minimization of the binary size of the VSV is critical to avoid saturation of available network bandwidth. The uniqueness and redetectability of the VSV vector must be balanced against cost, practicality and deployability factors for the overall system. Additional traffic flow metrics of possible value in traffic management may also be generated and collected by the system: individual or traffic-averaged vehicle speed (both instantaneous or segment averaged), accumulated or time-averaged traffic counts, traffic density, and estimated vehicle classification (passenger auto, light truck, heavy truck).
For each passing vehicle, the VSV is locally generated and transmitted by a detection module to a central correlation computer, via a public low-power wireless Internet service provider. A network-connected correlation computer continuously receives VSVs in near-real-time (network latency only) via the Internet and attempts to match feasible pairings of VSVs transmitted from all detection sites. Successful vehicle matches are added to a database from which they may be used to drive a graphically represented microscopic traffic network flow model. Individual and average traffic flow metrics are optionally accumulated, including vehicle speed at each site and interval count. From these raw metrics may be derived traffic volume, occupancy, level of service, and indications of traffic disruptions.

As a prerequisite to field deployment or testing, we were encouraged by the New Technology and Research Division of the California Department of Transportation to comply with as many as possible of the requirements stated below. While the experimental version of the V²SAT system does not comply with requirements 2 and 6 below, it is anticipated that a product-prototype version will comply.

1. Detectors shall be mounted on overcrossings or rigid structures above or to the side of traffic lanes.
2. Detectors shall be capable of internal battery powered operation for 24 hour period.
3. System shall have the ability to uniquely identify each vehicle passing beneath detector, and re-identify detected vehicles at successive sites along freeway network.
4. Detection means shall be safe, non-intrusive, and low cost on per-site basis.
5. Detection elements shall be installable without disruption to traffic flow.
6. Integrate with existing network-wide traffic data collection systems.

**DESCRIPTION OF DETECTION METHOD**

The system utilizes a low-cost EIA-RS170/NTSC video camera as a sensor to provide optical information adequate for the development of a unique VSV for each vehicle. One camera per lane is used, mounted overhead on the side of an overcrossing or sign standard, as illustrated in Figure 1. A video digitizing subsystem and Pentium-II class processor is integrated with each camera, either as a detached unit (experimental system) or integrated in the same enclosure (prototype system). Collectively, these elements are referred to as a detection module.

The VSV is generated locally in real time by analysis of the stream of video fields containing vehicle images. The VSV-generation method is driven by several elemental image processing operations:

- Accumulation of a time-averaged background image using an IIR filter operating on individual pixels.
• Subtraction of the object (vehicle) image from the accumulated background along a set of selected scan lines to identify object boundaries.
• Field-to-field ensemble averaging of centerline traces from successive images to distinguish true object features from image artifacts and transient shadows.
• Intensity profile measurements along the true vehicle centerline to extract points of optical contrast, and extremal measurement of the vehicle width.
• Trigonometric correction of image coordinates to scene coordinates, including effects of camera height and view angle, to yield normalized length measurements.
• Average primary color hue and saturation measurement from parsed RGB values along the vehicle centerline.

If the camera is oriented perpendicular to the road surface, directly above the detected lane, only the height of the detector above the traffic lane is required in the system setup for geometric correction purposes. A tall vehicle may appear longer than a low one, but site-to-site differences are normalized with respect to camera height. The algorithm determines the true (image) center line of the vehicle, even if it is significantly off-axis with the lane. The algorithm then extracts metrics from the optical signature of the

FIGURE 1  V^2SAT Detector Camera Positioning above traffic lane.
vehicle, to the maximum extent possible for each vehicle: physical lengths between key points of abrupt intensity and chromatic change along the vehicle centerline, which for a typical sedan might correspond to the distance from bumper to windshield (L1), distance from bumper to rear windshield (L3) and at least two other distance metrics (L2) and (L4). These measurements are illustrated for a sample vehicle image in Figure 2.

Background subtraction and rejection of shadow artifacts is accomplished by using a combination of the color hue (H) and intensity (I) components extracted from the composite NTSC video signal by transformation of 24-bit RGB pixel values\(^1\) produced by a color frame grabber. In NTSC composite encoding of color video, hue is measured as the angle of the color “vector” in degrees. The inclusion of color measurements in the VSV are treated as optional, since they are not available for many vehicles, or under low-light conditions. The basic components of the VSV are summarized below. Each is encoded as an 8-bit integer, with the exception of the site code S and time code t, which occupy two bytes each. Not all components may be known for a particular vehicle; the lack of a component in the vector is encoded as a zero value, and ignored in correlation processing.

\[
\chi_{i,t} = (L_0, L_1, L_2, L_3, L_4, W, C_1, C_2, V, N, S, t) .
\]

where:

- \(L_0\) = Total vehicle length along vehicle centerline.
- \(L_1\) = Length from front of vehicle to first optical feature, typically the bottom edge of the windshield.
- \(L_2\) = Length from front of vehicle to second optical feature, typically the top edge of the windshield.
- \(L_3\) = Length from front of vehicle to third optical feature, typically the top of the rear window of a conventional sedan.
- \(L_4\) = Length from front of vehicle to fourth optical feature, typically the bottom edge of the rear window of a conventional sedan.
- \(W\) = Vehicle body extremal width, exclusive of mirrors or other small side projections.
- \(C_1\) = Primary color intensity component, measured as a normalized magnitude.
- \(C_2\) = Primary color hue component, measured in degrees.
- \(V\) = Vehicle velocity.
- \(N\) = Lane number at site.
- \(S\) = Site code number.
- \(t\) = Absolute time code, resolution to one video field, 1/60 second.

\(^1\) RGB = Red-Green-Blue color signal decomposition. HSI = Hue-Saturation-Intensity color signal decomposition. The HSI decomposition is derived from RGB by simple trigonometric calculations.
The velocity (V) component of the VSV is measured by time-of-flight. Since the dimensions and angle of the detection zone are known, this measurement is easily found from frame-to-frame geometric measurements. As the vehicle moves through the detection zone, the optical flow front is detected as a change in H (hue) and I (intensity) compared with the accumulated background. Along the true vehicle centerline, H and I level inflection points form the basis for the length metrics (L0 - L4).

FIGURE 2  VSV Component Measurements for Typical Raw Vehicle Image.
Figure 3 illustrates a set of typical centerline intensity profiles aligned and processed via this method. The two more similar traces (squares and diamonds) represent vectors from the same vehicle, generated at different sites. For comparison, the trace with triangle data points was generated by a different vehicle.

The complete algorithm and vector generation process is described in (19). A block diagram of the detection algorithm appears in Figure 4. A screen capture of the Windows 95 user interface of the demonstration version of the detection module program appears in Figure 5.
FIGURE 4  Detection Algorithm Block Diagram
THE VECTOR CORRELATION PROGRAM (CORRELATION ENGINE)

The data correlation program takes sets of vehicle data vectors from pairs of adjacent sites and attempts to match vehicles via their VSVs. Comparisons are limited to vectors generated within a reasonable time window, calculated for each vector pair by dividing the site separation distance by a maximum and a minimum speed admissible for the time-averaged traffic conditions at both sites.

A correlation calculation is then performed on every pairing of admissible vectors. The pairing with the lowest correlation error (best correlation) is declared the matching vehicle, subject to an absolute maximum
error limit, which allows "no match" conclusions to be made when all pairings are invalid. This is necessary since not all vehicles are present at both sites (vehicles can enter or exit the freeway between sites, or stop in the network section). The setting of this limit involves a trade-off: higher values increase tolerance for vague matches, increasing both the number of correct matches and the number of false matches, especially in cases of vehicles present at only one site. Correlation error is calculated as:

$$e = 100\% \cdot \frac{\sum_{j=1}^{n} W_j \cdot C_j \cdot |X_j|}{\sum_{j=1}^{n} (W_j \cdot C_j)}$$

Where the summations are indexed by $j$ over all mutually present vector components.

$X_j =$ normalized individual component error (0.0 to 1.0)

$W_j =$ component weight in final decision (0.0 to 1.0)

$C_j =$ confidence factor for component (0.0 to 1.0)

$n =$ total number of mutually present vector components

A screen capture of the user interface for the correlation engine is shown in Figure 6. A simplified flow chart for the pairing and correlation process is given in Figure 7.
FIGURE 6  Server / Correlation Engine User Interface.
FIGURE 7  Flow chart for Correlation Processing.
SYSTEM TESTS
Four field detection units and one central correlation server were fabricated and field tested with live traffic video, using video cameras mounted approximately 0.5 meters out from the side of the overcrossing deck, facing downward. The field test configuration is shown in Figure 8.

Tests were performed at pairs of consecutive overcrossings on US Hwy 101 along the central California coast. Lighting conditions and traffic conditions varied to the normal extent over the course of a typical commuting day, and night conditions were also acquired to allow assessment of the detection method under low-illumination conditions.

FIGURE 8  V2SAT Field Test Configuration.
The system was tested in two phases over a three year period (1997-99) by the Transportation Electronics Laboratory of the California Polytechnic State University. Initial tests, designated Phase 1, were intended to study the characteristics of video images of vehicles, to determine if adequate measurable information was present to permit unique and reliable re-identification of vehicles. This data assisted in the refinement of the detection method and the composition and weighting of components of the VSV. Field tests of the computer-vision system using live traffic were conducted under Phase 2.

SUMMARY OF TEST RESULTS

Field tests were performed on a four-lane section of US Highway 101 in San Luis Obispo County, California, using two detection sites separated by 0.34 mile (0.55 km). Results generated by the system in real time were compared off-line against manually verified results from video tapes. Three daylight illumination conditions were examined: morning, noon, and afternoon. Night conditions were not tested, since Phase 1 observations had confirmed that the VSV could not be generated without illumination adequate for imaging of the complete vehicle (not just headlights and tail lights). Test results for the noon illumination condition are illustrated graphically in Figure 9.

Over all conditions, a total of 4,243 individual vehicles were examined. Self-correlation accuracy, or the ability to correctly re-identify vehicles at successive sites, was found to average 93.6%. False correlation errors, due to incorrect matches of different vehicles at successive sites, occurred for 0.0116% of all time-admissible pairings. Finally, it was determined that the detection modules generated valid vectors for 97.0% of all actual vehicles, including some for which a reasonable VSV cannot be generated such as motorcycles or vehicles in transition between lanes.
Actual:

57 minute data run
1009 vehicles at site 1
893 vehicles at site 2
694 vehicles from site 1 arrived at site 2 (total possible matches)
315 vehicles from site 1 did not arrive at site 2
901,037 possible pairings
900,343 total possible incorrect matches

\( V^2 \)SAT Results:

743 matches reported: 642 correct + 101 incorrect\(^*\)
52 failures to match

Normalized Accuracy:

Percent vehicles correctly matched out of total possible matches: \( 92.5\% \)
Percent incorrect matches out of total possible incorrect matches: \( 0.0112\% \)

\(^*\) Of the 101 incorrect matches, 87 were pairings consisting of one or both vehicles that appeared at only one site.

**FIGURE 9** Sample Actual Traffic Flow Elements and \( V^2 \)SAT Results, Field Test 9-28-00, Noon.
OBSERVATIONS RELATED TO SYSTEM ERRORS

Observed reasons for false matches:

- Vehicle detected at first site does not arrive at second site, but $V^2$SAT finds another vehicle that is very similar (almost all error cases).
- Different vehicles have very similar top views.
- Vehicles have little or no chromatic information. Only 32.9% of vehicles observed had any usable chromatic information. White, silver, gray and black were observed to be the most common colors for cars and trucks.

Observed reasons for failures to match:

- Vehicle not detected, therefore no VSV generated. 3.0% of actual vehicles are missed, mostly due to poor lane alignment or lane change in progress.
- Video artifacts change the image of a vehicle sufficiently to make it appear different at another site.
- Vehicles have little or no chromatic information (same as above).
- Vehicle changes speed radically between sites.

SYSTEM LIMITATIONS

Among the observations from Phase 2 tests are the following intrinsic limitations of the $V^2$SAT system:

- $V^2$SAT, like all computer vision systems, requires adequate, even illumination of targets. This generally limits it to daylight operation, or night operation with artificial illumination.
- At least four meters of the detection area must be at the same illumination level; it cannot contain both very bright and very dark areas – e.g., bright daylight and a dark shadow such as that cast by an overcrossings. This contrast exceeds the dynamic range of CCD video cameras (20,21).
- Individual video cameras are required for each lane in the current version of system. This limits the number of detection sites to those with appropriate overhead structures (usually overcrossing bridges).

GENERAL CONCLUSIONS AND FUTURE DIRECTION

Computer vision algorithms and hardware were developed to mechanize the automated generation of Video Signature Vectors for every vehicle passing beneath video detection modules on a freeway. Field tests were conducted using an experimental version of the system at two sites on US 101 in San Luis Obispo, California. Over a range of normal daylight illumination conditions and 4,243 vehicles:

1. Under daylight conditions, the system correctly re-identified vehicles at successive sites for 93.6% of vehicles that were actually present at both sites, and falsely matched vehicles for 0.0116% of all time-admissible incorrect pairings.
2. Chromatic (color) information is of limited value for vehicle correlation. Only 32.9% of vehicles observed had any usable chromatic information. Also, CCD color video cameras are subject to the loss of chromatic information under low-light conditions and at very high shutter speeds.

3. Poor vehicle alignment in lane (changing lanes) contributes to reduced vector accuracy, since only part of the vehicle may be in the detection zone.

4. Scene illumination is a critical factor for detection accuracy. V²SAT requires adequate and even illumination of targets, limiting it to daylight operation, or night operation with a supplemental light source.

5. Video artifacts such as harsh shadows at one site can change a vehicle image sufficiently to make it appear different at another site. This appears to be a key source of error. The allowable illumination contrast range in the detection zone is limited by the dynamic range of the video camera.

6. For current vector generation algorithms, the camera must be able to clearly image at least the first four meters of each vehicle in order to generate a complete vector.

7. The need for individual video cameras above each lane, for the current version of the system, is probably the most significant deployment-limiting factor, since it restricts usage to roadways for which an adequate number of overhead structures (usually overcrossing bridges) are available to permit cohesive sampling of vehicle progression.

8. On the basis of these observations, we conclude that, subject to limitations associated with detector placement and illumination, the vehicle re-identification ability of the V²SAT system has the potential to serve as the basis for non-intrusively tracking the progress of individual vehicles along a roadway network.

9. In future work, we hope to expand testing to cover multiple detection sites, and evaluate the ability of the V²SAT system to provide data necessary to support the generation of origin-destination tables, travel time estimates, and the validation of microscopic traffic flow models.

REFERENCES


