Title
Section Related Measures of Traffic System Performance: Final Report

Authors
Ritchie, Stephen
Sun, Carlos

Publication Date
1998-11-01

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Section Related Measures of Traffic System Performance: Final Report

Stephen G. Ritchie, Carlos Sun
University of California, Irvine

California PATH Research Report
UCB-ITS-PRR-98-33

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

Final Report for MOU 224

November 1998

ISSN 1055-1425

SECTION RELATED MEASURES OF TRAFFIC SYSTEM PERFORMANCE

FINAL REPORT

Stephen G. Ritchie

Carlos Sun

Department of Civil and Environmental Engineering

University of California, Irvine

prepared for

Partners for Advanced Transit and Highways

PATH MOU 224

November 1998

ACKNOWLEDGEMENT AND DISCLAIMER

The research reported in this paper was supported by the California Department of Transportation PATH (Partners for Advanced Transit and Highways). The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This paper does not constitute a standard, specification, or regulation.

The authors gratefully acknowledge the assistance provided by our research partners Mr. Joe Palen of the California Department of Transportation, Mr. Craig Gardner of Gardner Transportation Systems, Dr. Reinhart Kühne from Steierwald Schönharting und Partner, and Dr. Kevin Tsai from the University of California, Irvine. The authors would also like to thank 3M Intelligent Transportation Systems, Peek Traffic Sarasota, Mr. Rick Nelson from the City of Irvine, and Dr. R. Jayakrishnan from the University of California, Irvine, for their assistance.

ABSTRACT

This project developed an important component of intelligent surveillance systems by investigating a new set of advanced traffic surveillance techniques based on pattern recognition technology. The approach uses standard inductive loops and transformed them into intelligent sensors that have the capability of reidentifying individual vehicles from one loop detector station to the next. The exercise in reidentifying vehicles from station to station is called the vehicle reidentification problem. One of the components that made this intelligent surveillance system possible was a high-speed (1-12 millisecond scan rate) scanning inductive loop detector card that has the capability of outputing inductance counts. The other critical component was the vehicle signature processing and correlation algorithm capable of solving the vehicle reidentification problem. The correlation algorithm was formulated and solved as a lexicographic optimization problem. The lexicographic optimization formulation is a preemptive multi-objective formulation that combined goal programming, classification, and Bayesian analysis techniques. The details of field implementation and data collection design are also presented. The solution of the vehicle reidentification problem has the potential to yield reliable section measures such as travel times and densities, and enables the measurement of specific dynamic origin/destination demands as well as the development of new algorithms for ATMIS (Advanced Transportation Management and Information Systems) implementations of the approach using conventional surveillance infrastructure. Freeway inductive loop data from SR-24 in Lafayette, California, demonstrates that robust results can be obtained under different traffic flow conditions. A discussion is also presented of several applications using section related measures of traffic system performance. The use of existing surveillance infrastructure coupled with this approach could allow development of widespread applications in Intelligent Transportation Systems (ITS).

Keywords: surveillance, section measures, vehicle reidentification, travel time

EXECUTIVE SUMMARY

This project presents a new set of advanced traffic surveillance techniques based on inductive vehicle waveforms and pattern recognition technology. The focus is on demonstrating and evaluating a new method for obtaining true section related (as opposed to point) performance measures for freeways such as section travel time and section density. In addition to section measures, this project also enables the measurement of lane-by-lane traffic movement and specific origin/destination demands. The results of this investigation can be applied to various Intelligent Transportation System components, and specifically to Advanced Transportation Management and Information Systems (ATMIS). For example, the proposed techniques can improve traveler information message content, better automated incident detection and congestion monitoring algorithms, and advance the use of section measures in dynamic formulations. This project addressed PATH's research needs under the ATMIS program by developing and testing a new mechanism for accurate and reliable measurement of traffic system performance. This unique effort complements other PATH research underway in the ATMIS area.

The vehicle reidentification algorithms are developed using classification and signal processing theory. The complex inductive waveforms are first reduced to useful feature vectors, and then reidentified based on a lexicographic optimization technique. This technique allows multiple objectives to be considered at the same time and allows the incorporation of new objectives in the future.

The vehicle reidentification algorithms for deriving section related measures of traffic systems performance performed well with freeway test data from SR-24 when compared with the video ground truthing data. The algorithms were able to perform well under both moderate traffic flow conditions and congested traffic flow conditions. Also, the section related measures derived by using the vehicle reidentification algorithms such as travel time and density were shown to be much better than the traditional way of extrapolating from "point" measures.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 OVERVIEW

Because of the widespread use of inductive loop detectors (ILD), Intelligent Transportation Systems (ITS) have a constant source of information on traffic system conditions. If these detectors could be used in a "smarter" way, more useful information or section measures of traffic system performance such as travel time and density could be obtained. This in turn translates into better traffic management and information systems via the use of accurate section measures. One critical application using section measures is dynamic origin/destination demand estimation. This application is a vital component of other ATMIS strategies such as traveller information, traffic assignment, and route guidance. The way to "squeeze out" more information from loop detectors is by using the vehicle waveforms that are produced when a vehicle passes over the loop detector. In order to produce meaningful traffic information such as travel time and density, the signature of a vehicle needs to be reidentified between different sites.

A general transportation network can be modeled using individual links where each link has only one ingress point and one egress point. Some examples of network links include multi-lane freeway sections, arterial sections, and ramps. If both the beginning and the end of each link is appropriately instrumented with a detector station, say a loop station with vehicle signature output, then vehicle waveforms can be obtained from the beginning (upstream) and ending (downstream) detectors. A vehicle waveform pair can then be formed using one downstream waveform and one upstream waveform. The set of vehicle waveform pairs is increasing over time as more vehicles enter and exit the link. If approached in a sequential fashion, the vehicle reidentification problem is to find the matching upstream vehicle waveform given a downstream vehicle waveform. Stated more formally, the vehicle reidentification problem is as follows:

given a set of vehicle waveform pairs $\boldsymbol{x}_i$, where $\boldsymbol{x}_i \in S,\ i = 1, \ldots, N_c,$ S is the set of vehicle waveform pairs, and N_c is the number of waveform pair

combinations. Find the waveform pair which is produced by the same vehicle.

The single solution of the reidentification problem produces the origin/destination tracking of a single vehicle, while the sequential solution of a stream of vehicles will produce measures such as link density and group travel time. Both kinds of solutions produce valuable data that are of interest to operating agencies, researchers, and individual travelers.

In a sense, the problem of vehicle reidentification is the same as the system identification problem. Only in this case, each vehicle is representative of its own system. Traditionally, system identification is split into two separate components: feature extraction and classification (pattern matching). Within feature extraction, there are so-called direct methods and also parametric methods. Since this system needs to be field implementable, there are many "real-life" constraints that limit the resources of waveform analysis. The various trade-offs between accuracy, computational intensity, and information bandwidth should be carefully weighed and considered.

This report describes the process of developing feature extraction and vehicle pattern matching algorithms and the subsequent derivation of section density and travel time based on conventional inductive loops. The vehicle reidentification problem is formulated as a lexicographic optimization problem because it is a multi-objective approach, and it allows the objectives to be incorporated sequentially. The multi-objective approach allows the incorporation of multiple feature vectors as well as Bayesian discrimination via probability functions estimated from historical data. The sequential solution of the objectives avoids the problem of finding weights that are meaningful among all the objectives, and also derives extremal solutions that can not be achieved with weighted averaging schemes. If vehicle waveform information is available from other kinds of detectors in the future, say laser detectors, then these waveforms can also be used with the reidentification algorithms if appropriate modifications are made to the algorithms. Freeway field data collected from State Route-24 in Lafayette, California, were used for algorithm development and the testing of algorithm performance. The performance of the section measures produced by the vehicle

reidentification algorithm is shown to be much better than the current method of extrapolation from local point measures.

This report also discusses applications of the vehicle reidentification system. One application is a new methodology for the estimation of dynamic origin/destination demand for use in Advanced Transportation Management and Information Systems (ATMIS) as well as other traditional transportation systems. The focus of this application will be to explore the dynamic considerations in an origin/destination estimation framework so that accurate values can be obtained for use in real time traffic management and control.

The work on vehicle reidentification compliments other research projects in Advanced Transportation Management and Information Systems. For example, this report can benefit dynamic traffic assignment, dynamic traffic control, traffic prediction, congestion monitoring, and driver information system.

1.2 MOTIVATION FOR SECTION RELATED VARIABLES

Occupancy is often used by traffic engineers as a surrogate for density. This is due to the fact that occupancy is easily computed using standard "presence" type detectors, but actual density requires "spatial" detection systems that are less accessible. Consequently, occupancy or even flow is used in place of density as an independent traffic measure. Occupancy and/or flow is also specified as the input variable for many traffic applications, including dynamic traffic assignment, incident detection, and dynamic traffic control.

Among the three basic traffic flow parameters (speed, flow, density), density is a direct measure of traffic demand. Density is not only a quantitative measure, but it is also a qualitative measure of the "closeness" of vehicles, which can reveal even the psychological comfort of drivers. For example, the stress level associated with driving in free flow traffic is much less than that associated with "stop-and-go" traffic.

The common practice is to use traffic flow as a measure of traffic demand. However, demand does not occur as a rate of flow. Traffic demand is generated (or absorbed) at sources and sinks according to land use. The origin/destination demand results in the network loading of individual links. The vehicles on these links is what produces the rate of flow and the speed.

Another problem with using flow is the fact that the relationship between flow and speed is not monotonic, thus the relationship between flow and travel time (inverse of speed) is also not monotonic. This behavior of the flow-speed curve causes many problems in algorithm formulations that use flow as a variable and require the computation of travel time cost. On the other hand, density and speed (or travel time) do possess a monotonic relationship. Travel time is important, since it is a measure of the traffic system cost. Figure 1.1 illustrates this point by the two figures based on the Greenshields hypothesis.

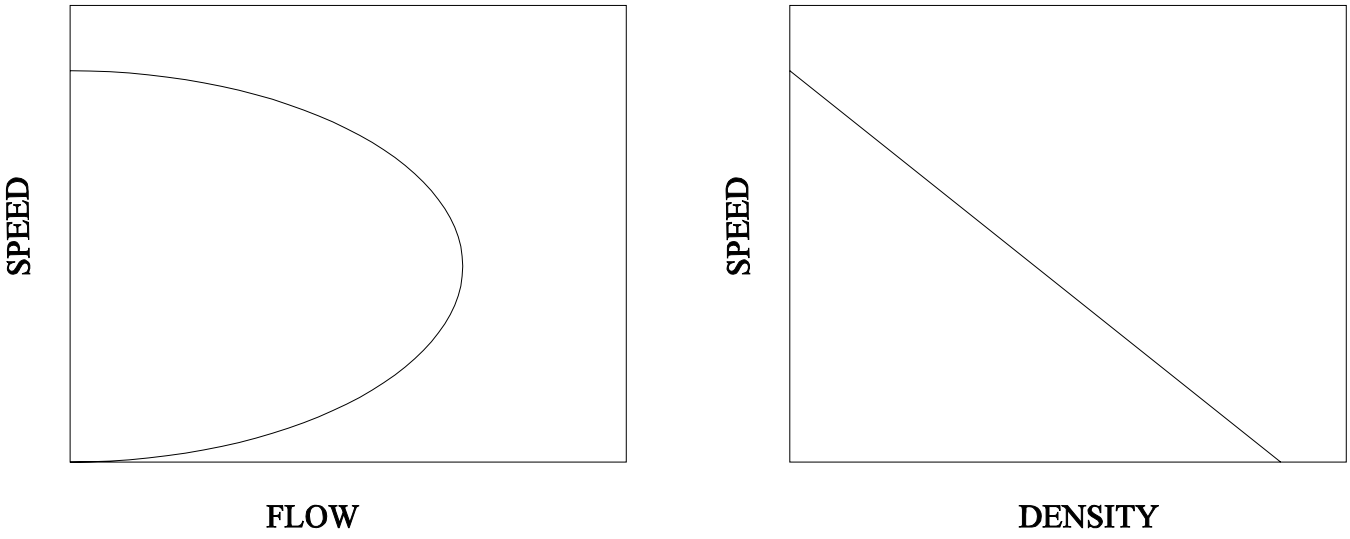


Figure 1.1. Speed-Flow and Speed-Density Plots Based on Greenshields Hypothesis

The fundamental difference between density and flow is that density is measured over a length of space at a particular instant in time, while flow is measured over a period of time at a particular point in space. In dynamic frameworks, density should clearly be the traffic variable of choice and not

flow.

Figure 1.2 displays a sample of the section density plots compared with the estimated density for a section of the SR-24 freeway in Lafayette, California. More results and plots will be presented in Chapter 5. The estimated densities were computed using point measures such as speed and flow. Average ten second estimated densities were computed while the section densities were taken at the ten second time intervals labelled in the x-axis. These data results confirm the intuitive view that spatial and temporal measures are fundamentally quite different. Significant errors can be introduced by the use of the temporal measure, occupancy, to estimate the spatial measure, density.

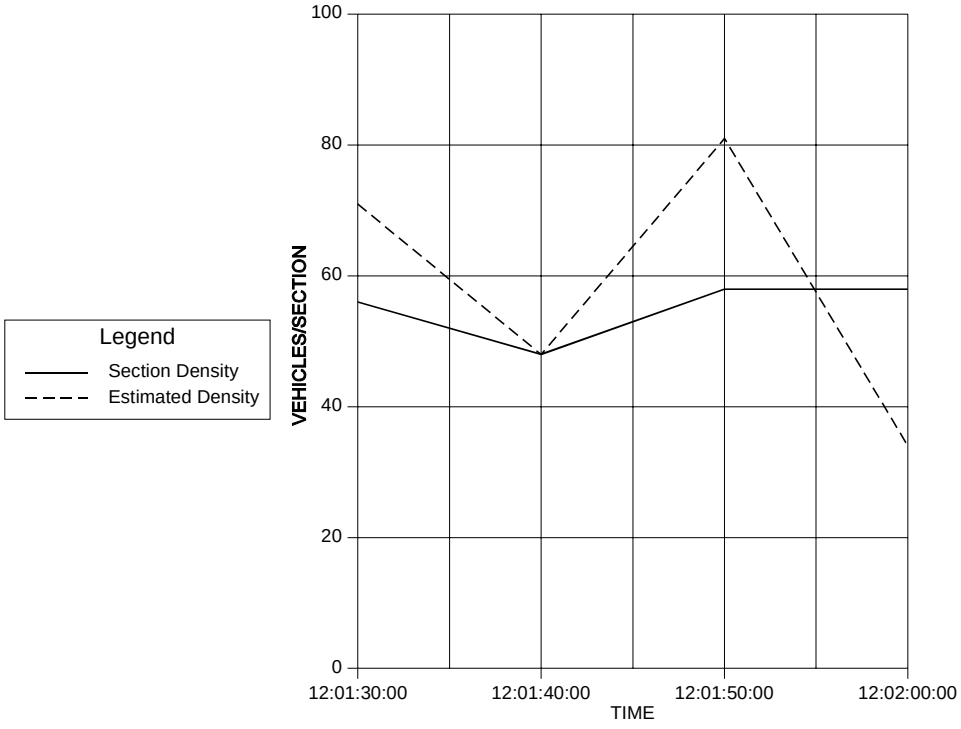


Figure 1.2. Comparison of True Section Density with Estimated Density

1.3 MOTIVATION FOR LEXICOGRAPHIC APPROACH

Lexicographic optimization is a preemptive multi-objective optimization method. Preemptive refers to the hierarchical structure of the various optimization objectives. The existence of multi-objectives is common in many real world problems. In lexicographic optimization, the various objectives need to be traded off.

Another multi-objective optimization method is weighted averaging. The lexicographical method has two main advantages over weighted averaging schemes. First, many objectives are specified in different units. With weighted averaging schemes, a common unit is needed to compare the different objective values. However, with lexicographic optimization, there is no such requirement. Second, only with lexicographic optimization can extremal Pareto solutions be achieved (Rentmeesters, 1998). When a weighted combination of all the objectives is used, the extreme points are compromised.

One useful characteristic of lexicographic optimization is the fact that the previous level optimization yields the feasible set of the next level. This allows constraints to be incorporated elegantly as goal programs. Goal program allows the establishment of a level of achievement for each criterion.

Lexicographic optimization is often used when real cost functions are not easily specified. In the case of the vehicle reidentification problem, the real functional form of the cost function is not known due to system uncertainties. Only selected values of the real cost function are derived from the limited ground truth database. The lexicographic optimization problem then becomes a surrogate for the unknown real cost function.

1.4. STRUCTURE OF REPORT

This report is intended to explore all aspects of the vehicle signature analysis and vehicle reidentification framework. This includes detailed discussions on field hardware and software design, definition of the vehicle reidentification problem, explanation of the lexicographical

optimization formulation, and suggested applications of the framework. The organization of the report is as follows. Chapter 1 is an introduction to this report. Chapter 2 discusses the background of vehicle signature analysis and reviews existing implementations of similar systems. Chapter 3 is a comprehensive discussion of all field implementation issues including data collection design, field equipment, database format, and existing field database. Chapter 4 explains the lexicographical optimization approach for solving the vehicle reidentification problem. Chapter 5 presents various results of the vehicle reidentification, section measures derivation, and other related efforts of this research. Chapter 6 discusses several applications of the vehicle reidentification and section measures derivation system. Chapter 7 presents the conclusions of this research, summarizes the original research contributions, and outlines directions for future research.

CHAPTER 2. BACKGROUND AND LITERATURE REVIEW

2.1 LITERATURE REVIEW OF ADVANCED INDUCTIVE LOOP SURVEILLANCE

Increasingly, in both the United States and Europe, and particularly in California, there is interest in investigating methods for obtaining true section-related traffic performance measures. Some of these measures include travel time, space mean speed, density (as opposed to occupancy), and even origin/destination information. By using non-obtrusive and anonymous tracking methods, individual vehicles could be identified and correlated over numerous identification stations. If pattern matching algorithms prove to be proficient, then very specific real time data could be obtained for any vehicle.

The actual physical sensor for these section-related traffic measure systems can be from different sensor technologies. It can also be combinations of technologies. Some extensive evaluations of newer detections technologies were sponsored by the FHWA (Federal Highway Administration) and is summarized in Klein (1995). More recently, a report of non-intrusive traffic detection technologies was presented by Bahler et al. (1998). Nine different technologies were evaluated which include inductive loop, passive infrared, active infrared, magnetic, radar, doppler microwave, pulse ultrasonic, passive acoustic, and video. However, all of these evaluations mainly involved the use of these technologies in a traditional sense, i.e. to replicate the functionality of standard inductive loop detectors. In a few instances some newer functionalities were tested, but they were not discussed in detail. The ability to output vehicle signatures for the formulation of the vehicle reidentification problem was not included in these evaluations.

This report only discusses the use of inductive loop detectors, although many aspects of this research can also be applied with the use of other detector technologies. Some manufacturers of detectors that output vehicle signatures include 3M Intelligent Transportation Systems (1997), Peek Traffic Sarasota (1992), and Intersection Development Corporation (IDC, 1995) in the United States, and ave Verkehrs- und Informationstechnik (AVE) in Germany (Böhnke, 1995). IDC and AVE have commercial systems that will also produce section related measures. But, as commercial products,

the details of these algorithms are not available in the public domain for evaluation.

The published literature on the use of inductive vehicle waveform analysis and reidentification is scarce. Böhnke and Pfannerstill (1986) discussed the use of inductive waveforms and the Karhunen-Lowe transformation in reidentifying vehicle sequences. Even though the use of the Karhunen-Lowe transformation reduces the data requirements, it also simplifies the feature vectors to render individual vehicle reidentification an impossibility. Kühne in cooperation with other researchers has published several reports of section measure instrumentation for obtaining inductive loop vehicle signatures. A freeway control system using a dynamic traffic flow model and vehicle reidentification technique was discussed by Kühne (1991). Another system discussed by Kühne and Immes (1993) detailed a freeway control system using section-related traffic variable detection. Both implementations also rely on the matching of vehicle sequences and assume that the vehicle sequence will be more or less consecutive from one station to the next. More recently, Kühne et. al. (1997) presented an overview of European projects and initial ideas for a study approach in an early phase of this project.

Laser sensor systems in the future will also be able to capture vehicle signatures. These laser sensors are complex units composed of several detectors that monitor contour and traffic parameter information independently. As a vehicle moves across the detector, abrupt changes in reflections are detected, especially at transition points such as the transition from the bumper to the hood. This signature would be different from the inductance signatures obtained from loop detectors.

2.2 BACKGROUND ON VEHICLE SIGNATURE ANALYSIS AND REIDENTIFICATION

The inductive loop detector (ILD) operates on the principle of mutual inductance between the loop and the equivalent conducting plate of a vehicle as seen in Figure 2.1.

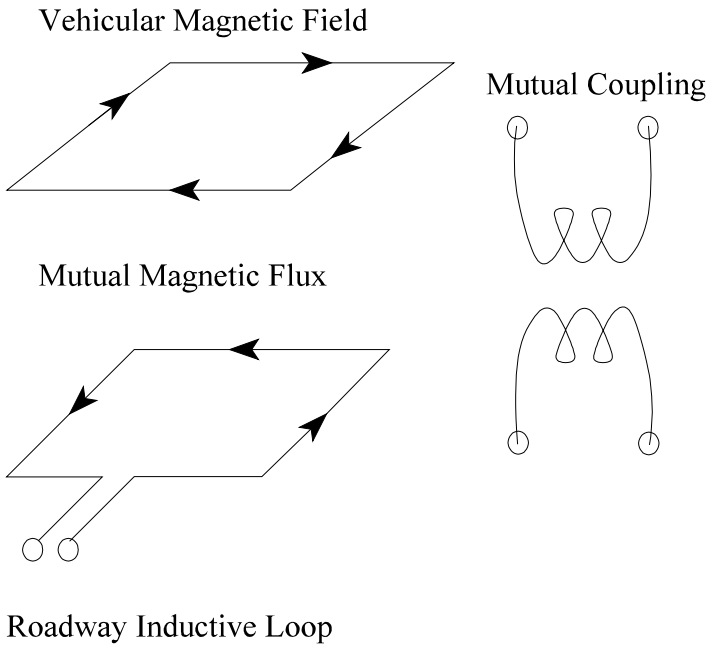


Figure 2.1. Vehicle/Loop Mutual Inductance

The inductance of an ILD is (ITE, 1990):

$$L_{12} = \frac{N_2 \Phi_{12}}{I_1}$$

where N_2 is the number of turns (usually 1 for an equivalent shorted turn), I_1 is the current flow, and Φ_{12} is the magnetic flux normal to the vehicle expressed as (Cheng, 1985):

$$\Phi_{12} = \int_{S_2} \boldsymbol{B_1} \cdot d\boldsymbol{s_2}$$

where $\boldsymbol{B_1}$ is the magnetic field, and $\boldsymbol{s_2}$ is the bounding surface of the vehicle.

This electromagnetic resonance circuit operates as follows. When a metallic mass passes through the magnetic field generated by the inductive loop, the disturbance produces a net reduction in the

loop inductance or frequency, and the resonance circuit properties are altered. A motorcycle, for example, could produce a frequency shift of up to 0.08% (80nH), while an automobile could cause a shift of up to 3% (3500nH) (Detector Systems). The metallic component of the vehicle is what disturbs the loop inductance. As a result, double-axle trucks produce a twin-peaked vehicle signature when the resolution of the detector is adequate. Thus, in concept this method can easily be used for vehicle-type identification purposes. Figure 2.2 shows some sample vehicle signatures of different types of vehicles.

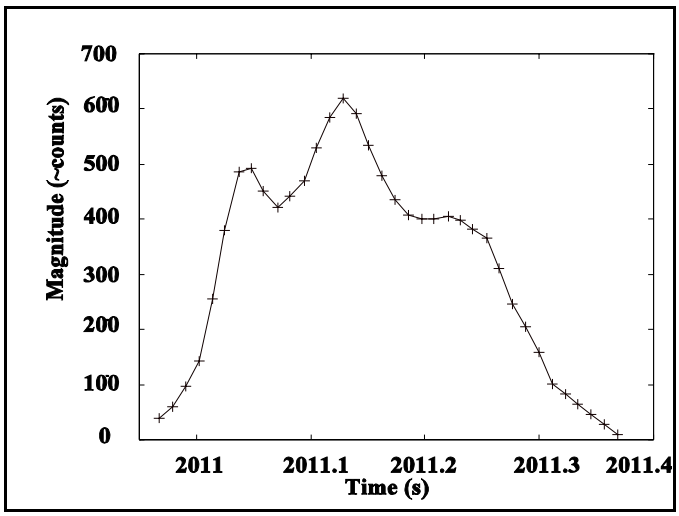


Figure 2.2. a) Semi-trailer Truck Signature

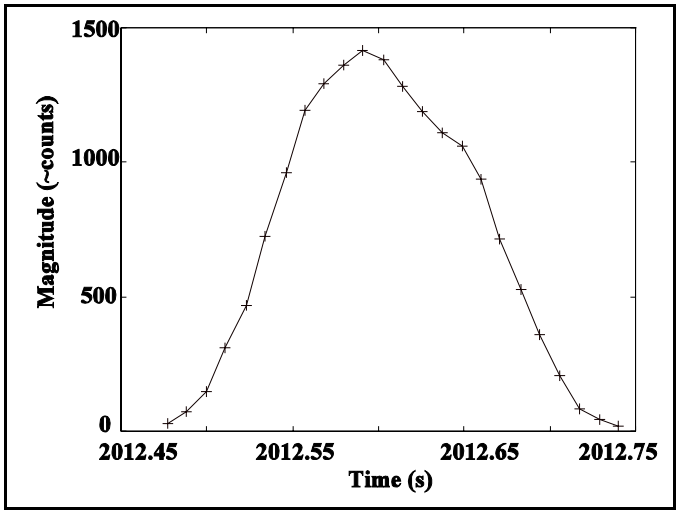


Figure 2.2. b) Minivan Signature

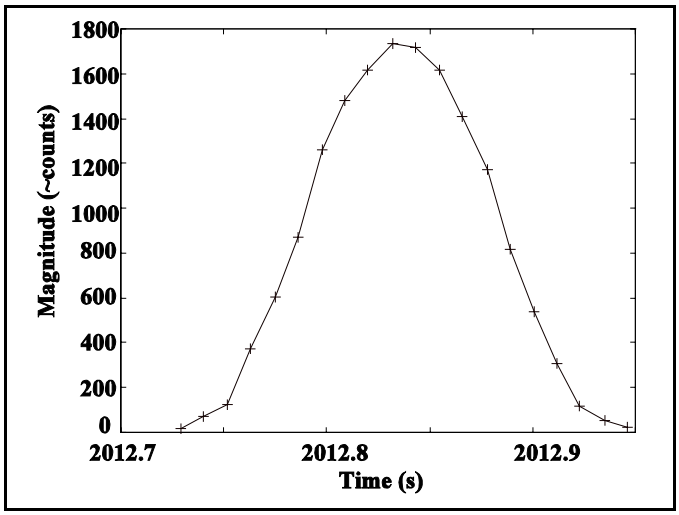


Figure 2.2. c) Passenger Car Signature

Figure 2.2. Sample Inductive Vehicle Signatures

Figure 2.3 is a conceptual model of the measurement system, where u(t) represents the input to the system, y(t) is the output of the system and e(t) represents disturbances or noise. Note that e(t) does not have to refer to a physical error source. It could include imperfections in the input or the output of the system. In the vehicle re-identification context, u(t) would be an idealized vehicle. In other

words, a vehicle with "perfect" vertical and horizontal alignment with respect to the loop detector. In the ideal case, the same vehicle u(t), will produce the same exact y(t) every time it crosses the loop. Since each vehicle traverses a loop each time at a different orientation with differing dynamic vertical clearance, there is a deviation from the ideal vehicle signature. This alignment error source can be lumped together with other noise as e(t). Sometimes the input disturbance is called the "uncontrollable input" (Ljung, 1987). Another error source includes the varying inductance of the loops caused by external environmental conditions, changing loop characteristics, or variations in the signal energizing circuitry. Even though sensitivity analysis of the system along with pinpointing of error sources is desirable, utilization of current infrastructure restricts the focus of this research to the use of current inductive loop technology and not the design of an ideal detector system.

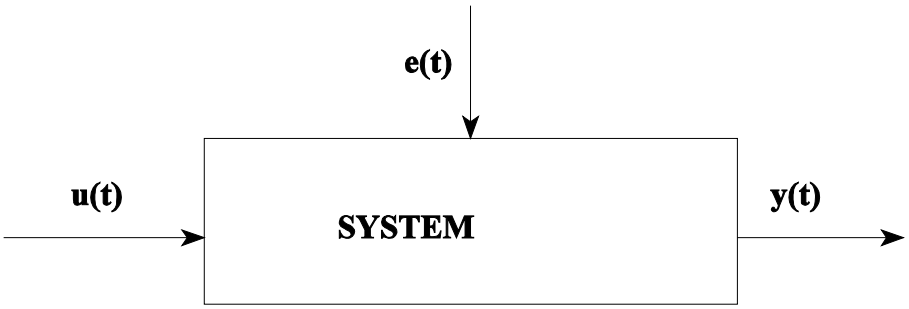


Figure 2.3. System Model

Vehicle identification is based on a set of feature vectors extracted from the raw vehicle waveforms, and other information. Such additional information can include vehicle speeds computed from speed traps, known lane numbers, and location geometry. The main assumption in this approach is that the feature vectors would be less sensitive to the error disturbances than to the input, otherwise noise would dominate the system and classification would be impossible.

Figure 2.4 illustrates the two phases of the vehicle re-identification problem. The feature extraction portion seeks to extract the salient components of the waveform that would sufficiently differentiate

vehicles. In order to avoid redundancy, each vector would contain different information. This is similar to the process of deriving a basis in linear algebra or finding the principal components in data analysis. In a likewise fashion, the goal here is to find an orthogonal set of vectors that would span the space of possible vehicles signatures. The practical constraints of the system, such as the cost of extraction and storage requirements, restrict the implementation to an incomplete basis set. However, this is not a serious concern since adequate accuracy is theoretically achievable with a subset of basis vectors.

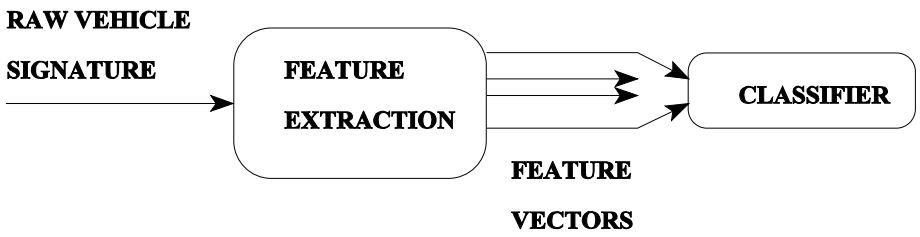


Figure 2.4. Vehicle Reidentification System

The two types of feature extraction methods are direct estimation and parametric methods. Direct estimation includes extracting time domain components, estimating the impulse response, and estimating the frequency response. Parametric estimation involves the assumption of a specific model structure, and the resulting estimation of the model parameters.

Direct estimation includes the use of the raw signals and other features derived from the raw signals. For example, in vehicular feature extraction, the raw vehicle signature is a unique feature vector. From the vehicle signature, other features such as maximum inductive amplitude and inductive loop occupancy time can be derived. Additional instrumentation provides even more feature vectors. For example, a speed trap gives an estimate of the vehicle speed, and the lane number is also output by the loop detector.

The impulse response is a convenient representation of a system that possesses the qualities of

linearity and time-invariance. Linear systems are defined as those systems that satisfy the following equation:

$$T(\omega x_1[n] \quad \omega x_2[n]) = \omega T(x_1) \quad \omega T(x_2) = y_1[n] \quad y_2[n]$$

where $y_1[n]$ and $y_2[n]$ are outputs of the system, $x_1[n]$ and $x_2[n]$ are inputs, ω and ωare scalar coefficients, and $T(.)$ is the system operator. The principle of superposition can thus be applied to linear systems. Time invariant (or shift invariant) systems will produce the output sequence $y_1=y[n-n_0]$ given an input of $x_1=x[n-n_0]$ for all time shifts, n_0.

In a linear time-invariant system (LTI), let $h_k[n]$ be the response of the system to an impulse $\omega[n-k]$ at time $n=k$. The output of the system for an input $x[k]$ would then be:

$$y[n] = T\left(\sum_{k=-\infty}^{\infty} x[k]\omega[n \quad k]\right) = \sum_{k=-\infty}^{\infty} x[k]T\omega[n \quad k] = \sum_{k=-\infty}^{\infty} x[k]h[n \quad k] = x[n] \quad h[n]$$

Consequently, a LTI system is completely characterized by its impulse response, since the output due to any input $x[n]$ is simply the convolution with the impulse response (Oppenheim and Schaffer, 1989). In the case of vehicle re-identification, the pure input (vehicle metal mass) is not measured so the impulse response can not be easily derived and used as a feature vector.

A Fourier transform pair is given by:

$$x[n] = \frac{1}{2\omega}\int_{\omega}^{\omega} X(e^{j\omega})e^{j\omega}d\omega, \qquad X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{\ j\omega n}$$

A discrete sequence, $x[n]$, is thus represented by its Fourier integral. The Fourier transform determines how much frequency component is contained at each frequency harmonic. The complex Fourier signal can be expressed as the magnitude $|X(e^{j\omega})|$ and the phase $X(e^{j\omega})$. For a LTI system, the frequency response of a system is simply the Fourier transform of the impulse response. Figures

2.5 and 2.6 show the magnitude spectrum pairs for a truck and a minivan. Since the critical frequency content of both signatures is close to the origin, the higher frequency components can be filtered out without significant loss of information. The frequency domain figures show only the fundamental frequency, therefore the Fourier transform was not included for use as a feature vector.

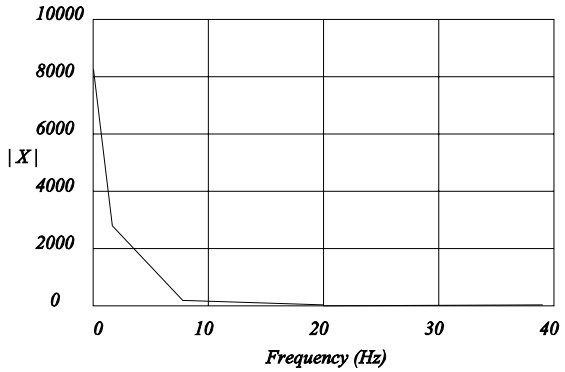


Figure 2.5 a) Fourier Analysis of Upstream Truck Signature

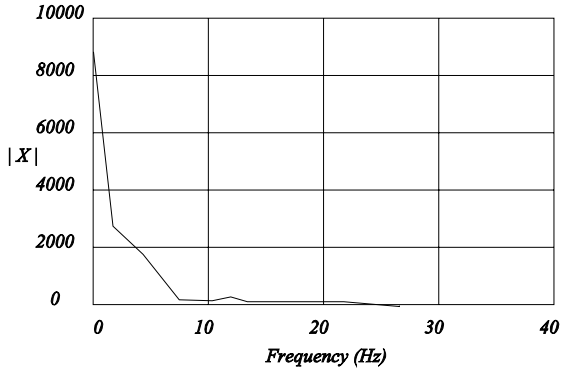


Figure 2.5 b) Fourier Analysis of Downstream Truck Signature

Figure 2.5 Fourier Analysis of Truck Signature

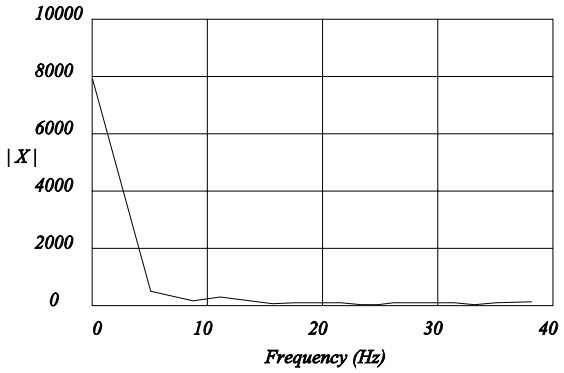


Figure 2.6 a) Fourier Analysis of Upstream Minivan Signature

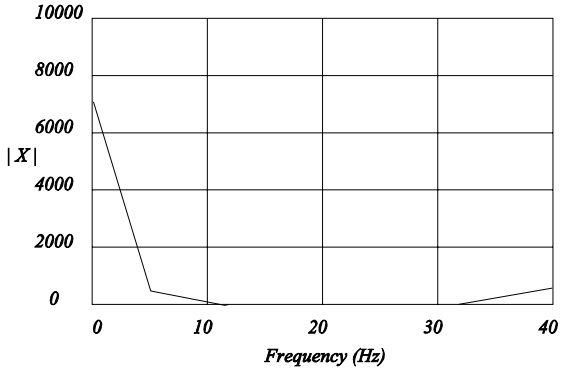


Figure 2.6 b) Fourier Analysis of Downstream Minivan Signature

Figure 2.6 Fourier Analysis of Minivan Signature

The following is a list of commonly used parametric models:ARX - autoregressive

- ARMAX - autoregressive/moving average
- Output-Error
- Box-Jenkins
- State-Space

The general structure of ARX, ARMAX, Output-Error, and Box-Jenkins is

$$A(q)y(t) = \sum_{i=1}^{\omega} [B_i(q)/F_i(q)]u_i(t \quad nk_i) \quad [C(q)/D(q)]e(t)$$

The state-space model structure is defined by the following set of equations:

$$x(t\ \ 1) = Ax(t)\ \ Bu(t)\ \ Ke(t)$$

$$y(t) = Cx(t)\ \ Du(t)\ \ e(t)$$

The standard way of estimating the coefficients in these models is by using the least squares method to minimize error, or Maximum likelihood (Bow, 1992). Since direct methods are less computationally intensive and less complicated than parametric modelling, the re-identification algorithms only use feature vectors that are derived from direct methods.

Given that a feature vector $\boldsymbol{X}$ consisting of N features is measured from each vehicle pattern, this vector belongs to the N-dimensional feature space, ω_x. The classification problem is to partition the feature space in such a way so that each vector is assigned to the proper pattern class (i.e. vehicle).

In classification, discriminant functions corresponding to the possible pattern classes ω_1, ω_2, ω_3, ... ,ω_m are used. If the discriminant function $D_i(X)$ is associated with the pattern class ω_i and the feature vector X is in class ω_i, then the classification rule would be $D_i(X) > D_j(X)$ for all j .

Normally in pattern recognition, the goal is to determine whether a given pattern belongs to some prescribed class of patterns. In the case of vehicle matching, there are two samples of the same vehicle, say an upstream and downstream sample, that need to be matched. The upstream sample (or downstream) can then be treated as the class or prototype, and the downstream samples are then searched for an adequate match.

In nonparametric classification, each signature can be treated as a point in the pattern space (Bow, 1992). Different classes will pertain to different regions in the pattern space. The definition of a class can be as restrictive as a single vehicle type (i.e. specific make and model) or more general, such as sports utility vehicles. As a side note, in regards to grouping vehicles, the critical element

is the electrical signature and not other descriptions such as coupe or sedan that may or may not differentiate the vehicle's inductance characteristic. Discriminant functions are used to separate classes which categorize patterns to their appropriate regions. In an *n*-dimensional surface, there are (*n*-1) hyperplanes or decision surfaces. Therefore, the decision surface is defined by a discriminant function, $d(\boldsymbol{x})$.

Figure 2.7 shows a simple example of a one-dimensional pattern space. The two points represent the decision "surfaces" separating three pattern classes. Figure 2.8 shows the two-dimensional case. The two dimensional decision surface is the following line:

$$w_1x_1 + w_2x_2 + w_3 = 0$$

If $d_1(\boldsymbol{x})$ is the discriminant function for class 1 and $d_2(\boldsymbol{x})$ for class 2, then the separating surface would be defined by:

$$d_1(\boldsymbol{x}) - d_2(\boldsymbol{x}) = 0$$

so if $d_1(\boldsymbol{x}) > d_2(\boldsymbol{x})$ then the pattern is classified to be in class 1, ω_1.

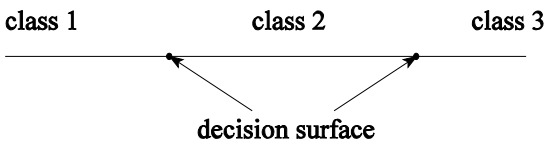


Figure 2.7. One-Dimensional Pattern Space

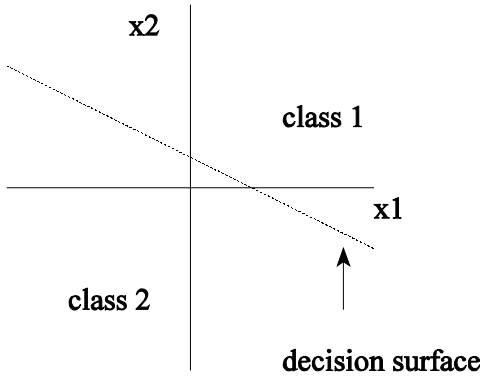


Figure 2.8. Two-Dimensional Pattern Space

In general, for an *n*-dimensional pattern space, the equation of the decision surface becomes

$$w_1x_1 + w_2x_2 + w_3x_3 + ... + w_nx_n + w_{n+1} = 0$$

or the dot product of the weight vector and the feature vector. The term w_{n+1} takes care of coordinate translation.

$$\boldsymbol{w}^T \cdot \boldsymbol{x} = 0$$

where

$$\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \cdot \\ \cdot \\ \cdot \\ w_n \\ w_{n\ 1} \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_n \\ 1 \end{bmatrix}$$

The general linear discriminant function has the following form:

$$d_k(\boldsymbol{x}) = w_{k1}x_1 \quad w_{k2}x_2 \quad \ldots \quad w_{kn}x_n \quad w_{k,n\ 1}x_{n\ 1}$$

Linear discriminant functions only serve the class of patterns that are linearly separable. In other words, the decision surfaces need to be convex, since convexity guarantees linear separability.

One of the simplest but most effective discriminant functions is the minimum distance classifier. The function is described as

$$\boldsymbol{x} \ \omega_j \quad \text{if } D(\boldsymbol{x},\boldsymbol{z}_j) = \min_k D(\boldsymbol{x},\boldsymbol{z}_k) \quad k$$

where $\boldsymbol{x}$ is the set of feature vectors, $\boldsymbol{z}_k$ is the class prototype, and $D(\boldsymbol{x},\boldsymbol{z}_k)$ is any distance metric. One

common distance metric is the Euclidean distance:

$$D(\boldsymbol{x},\boldsymbol{z}_k) = \boldsymbol{x} \quad \boldsymbol{z}_k$$

For most distance measures, if $D(\boldsymbol{x},\boldsymbol{z}_k) > D(\boldsymbol{x},\boldsymbol{z}_j)$ then $D^2(\boldsymbol{x},\boldsymbol{z}_k) > D^2(\boldsymbol{x},\boldsymbol{z}_j)$.

$$D^2(\boldsymbol{x},\boldsymbol{z}_k) = (\boldsymbol{x} \quad \boldsymbol{z}_k)^T(\boldsymbol{x} \quad \boldsymbol{z}_k) = \boldsymbol{x}^T\boldsymbol{x} \quad 2\boldsymbol{x}^T\boldsymbol{z}_k \quad \boldsymbol{z}_k^T\boldsymbol{z}_k$$

The Minimum Distance Classifier (MDC) can then be equivalently expressed as

$$\min_k \; (\; 2\boldsymbol{x}^T\boldsymbol{z}_k \quad \boldsymbol{z}_k^T\boldsymbol{z}_k) = \max_k \; (\boldsymbol{x}^T\boldsymbol{z}_k \quad \frac{1}{2}\boldsymbol{z}_k^T\boldsymbol{z}_k)$$

In using the MDC, one has to be careful and make sure that the classes and feature vectors are standardized. If not, misclassification will result due to differences in class variances. Figure 2.9 illustrates a problem scenario in which the feature vector is misclassified as ω_1 since $D_1 < D_2$.

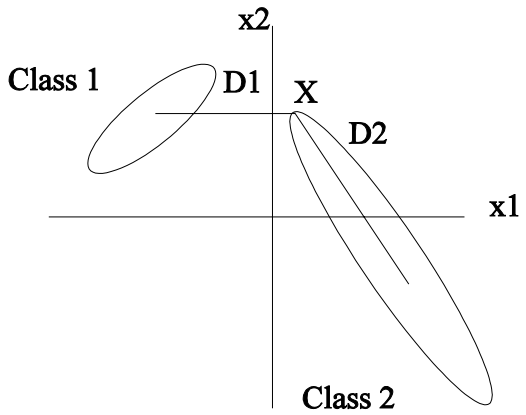


Figure 2.9. MDC Misclassification

An extension of the simple linear discriminant function is the quadratic discriminant function. Using the quadratic function is one way of treating nonlinear feature surfaces. The quadratic function has

the following form:

$$d(\boldsymbol{x}) = \boldsymbol{x}^T A \boldsymbol{x} \quad \boldsymbol{x}^T B \quad C$$

where A, B, and C are weight matrices. Other nonlinear discriminant functions include phi functions, potential functions, and piece-wise linear functions.

For some non-linear feature spaces, a piece-wise linear discriminant function can be used. In this case, there could be many class prototype discriminant functions $d_k^m(x)$ where m is an index to the particular class prototype. Given the additional discriminant functions, a classification rule can be developed that takes into account the totality of all the linear discriminant functions combined. Figure 2.10 shows a simple example of a two class piecewise linear discriminant function.

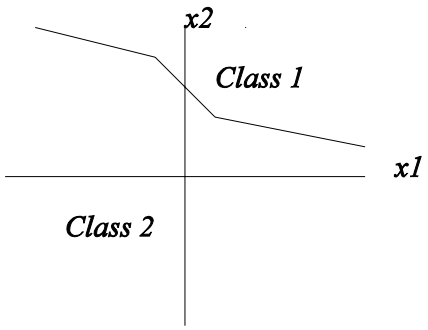


Figure 2.10. Piecewise Linear Discriminant Functions

One way of considering multiple linear functions is simply to take the greatest discriminant function. This is called the *nearest-neighbor* rule. So

$$d_k(\boldsymbol{x}) = \max_m d_k^m(\boldsymbol{x}) \quad m$$

An alternate approach is the *committee machine*, which considers the decision from each discriminant function and totals the votes. Figure 2.11 shows an example of this two-tiered machine.

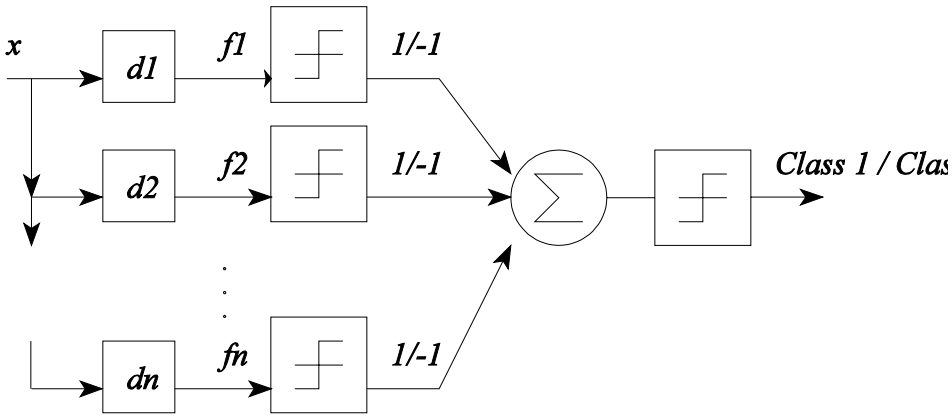


Figure 2.11. Committee Machine

An alternative classification rule when pattern categories are known a priori is the statistical discriminant function. This includes areas such as statistical decision theory, loss functions, Bayes' discriminant function, and maximum likelihood decision. In statistical classification, an underlying distribution density function needs to be assumed. In general, the pattern categories are characterized by a set of parameters.

This chapter presented an overview of approaches commonly used in system identification. However, not all approaches are suitable for the problem of vehicle reidentification. In chapter 4, the chosen approach will be discussed in detail and explanations will be given as to why one approach was chosen over another.

CHAPTER 3. FIELD INSTRUMENTATION AND DATA COLLECTION DESIGN

The discussion of field instrumentation and data collection is very important from two perspectives. First, the requirements for a field implementation of the proposed vehicle reidentification system are critical to agencies that wish to deploy such a system. One goal of this research is to enable government agencies to utilize the proposed technology in the near future. Second, the careful explanation of data collection procedures is provided for the use of other researchers who are interested in investigating the vehicle reidentification problem. This chapter is divided into two sections. One section discusses all the issues involved with field deployment and field data collection. The second section presents the details of the data that were collected as part of this research.

3.1 FIELD INSTRUMENTATION AND DATA COLLECTION

The data collection effort is a superset of a field implementation in the sense that in addition to collecting the vehicle signature data, there is the additional need to collect video ground-truthing information. The ground-truthing is needed for researchers to evaluate the performance of their vehicle reidentification systems. The discussion will therefore center on the development of the data collection setup, since the field implementation is a subset of this setup.

The first critical issue related to field data collection is the selection of an appropriate data site. This site will vary depending on the overall application of the vehicle reidentification system. There are several general recommendations that are critical in producing useful vehicle signature data. Also in this discussion, a site will consist of a single section or link of roadway. This link can be part of an arterial or a freeway. Once the requirements for a simple link are defined, the design can be extended to a general network that is composed of the links which are the basic building blocks.

The first requirement for a data collection link is that every vehicle that is traveling in the relevant

section must be captured by the datalogging detectors. In other words, the section needs to be "closed". If there are ingress and egress points within the data collection site, these additional vehicles will introduce vehicle signatures that will not be matched. This ideal requirement might not be feasible in certain locations due to engineering or institutional concerns, and one would be faced with losing reidentification accuracy. One would expect that arterial sections would necessitate instrumentation at each intersection block. The arterial sections might have lengths of approximately an eighth of a mile or a quarter mile, while freeways would only need instrumentation between ramps.

Figure 3.1 shows a sample arterial site in Irvine, California, that is suitable for data collection. This site consists of a two lane section of Alton Parkway between the intersections with Telemetry and Jenner streets. Each loop station has double loops for speed measurements. The distance between the two loop stations is 425 feet. The set of advance loops between the two stations was not used in the data collection effort. Arterial signalization which causes vehicles to decelerate during the red phase produces conditions that differ from freeway sections.

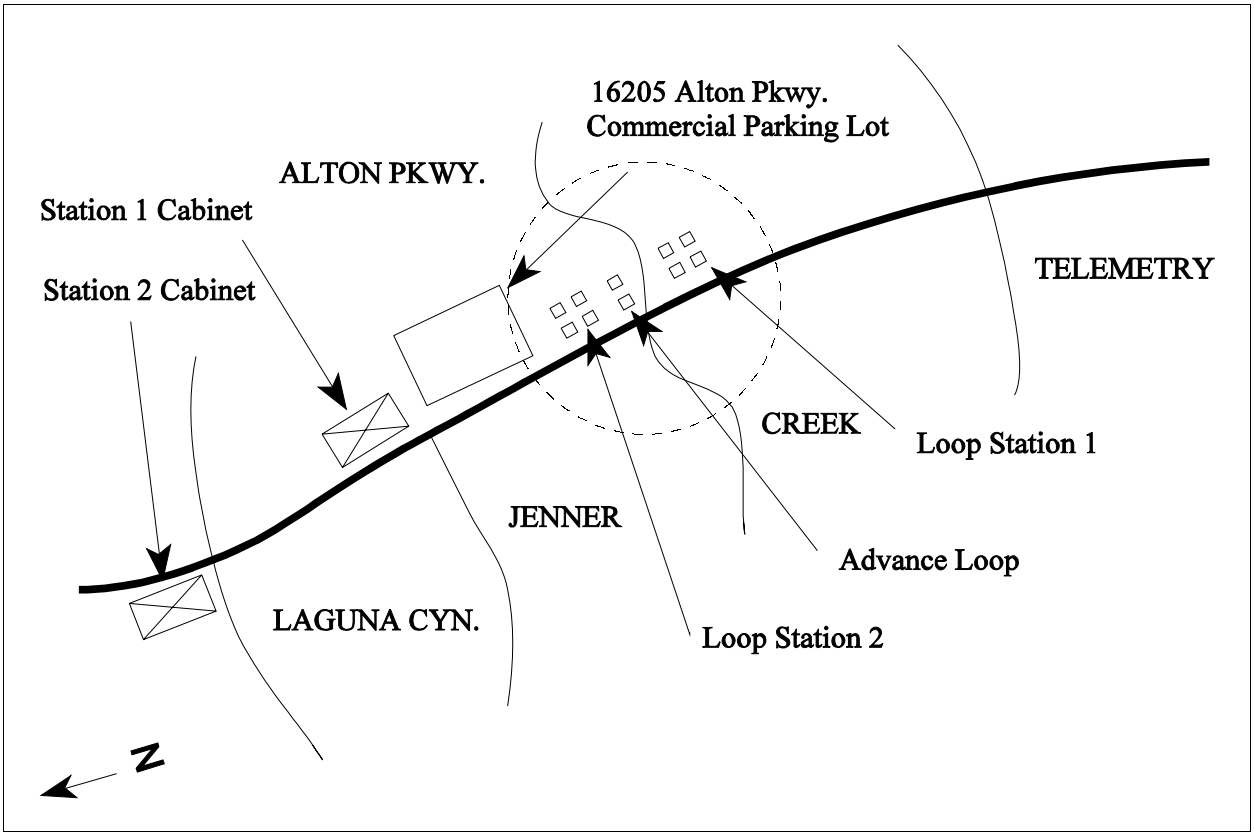


Figure 3.1 Sample Arterial Data Collection Site

Figure 3.2 shows a sample freeway site in Lafayette, California. The upstream loop station is located near the Central Lafayette on-ramp while the downstream loop station is before the Acalanes Road off-ramp. The proximity of this section to the on-ramp produces merging behavior that is suitable for algorithm testing. The four lane configuration also allows the study of "out-of-sequence" vehicles due to overtaking and lane changing.

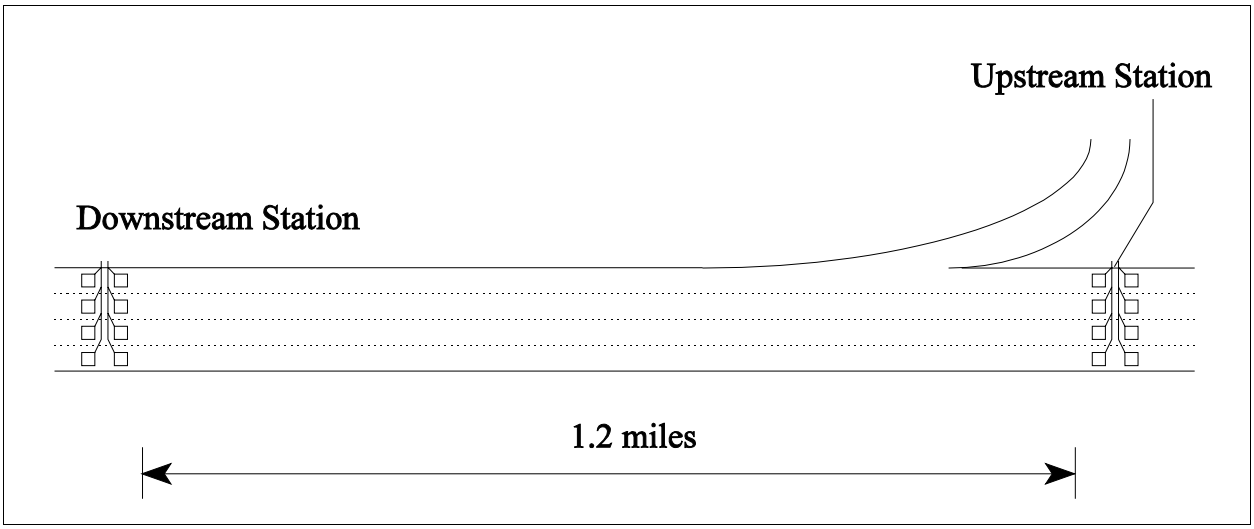


Figure 3.2 Sample Freeway Data Collection Site

The hardware interconnect of the test site is composed of four different components: the data processing computer, time synchronization source, physical loops, and the detector cards. Figure 3.3 illustrates the hardware interconnect for instrumenting a two lane highway. Additional loops and detectors could be added to accommodate more lanes. If communications are required for a real time implementation, then a modem connection would also need to be designed.

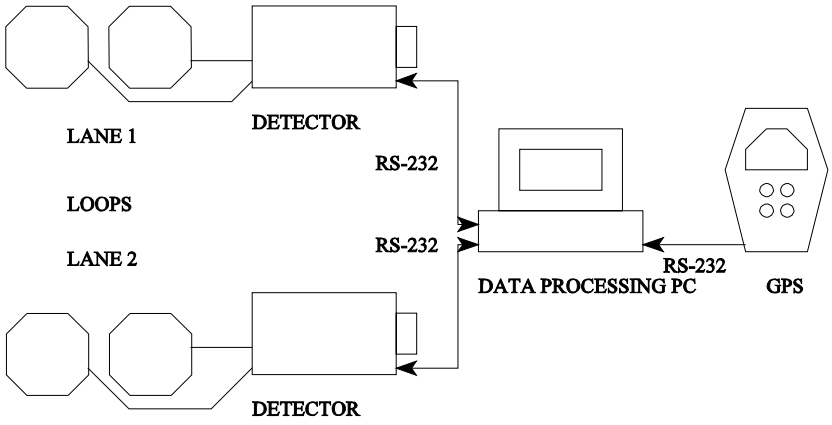


Figure 3.3 Sample Hardware Interconnect

The data processing computer is the central component in a local field site. If communications are designed for several local sites and a central location, then additional computers at the central site can perform more computationally intensive applications such as incident detection. The local computer can contain different hardware and software platform specifications. Some different hardware specifications include the Intel x86 processor or the Motorola 68K processor, and central bus or VME (Versa Modular Eurocard) bus. Many operating systems such as DOS, Windows, UNIX, Macintosh, and OS-9 all have the capability to fulfill the data processing duties. In the past field data collection effort for this research, VME bus computers were used for freeway data collection, while standard Intel Pentium based laptops were used for arterial data collection . However, in real time field implementations an approved traffic controller would need to be used. Currently, the only computer that is adequate for field implementation is the 2070 controller. The 170 controller or NEMA controllers would not have adequate capabilities for the proposed system.

The most critical requirement of the datalogging computer is adequate communications ports to interface with detectors, GPS receivers, and other peripherals. Serial communication is often specified for detectors, GPS receivers, video time encoders, and other peripherals. Most desktop computers are generally equipped with two serial ports, while most laptops have only one standard serial port. If multiple peripherals are attached to the data logging computer, additional serial ports are needed. The following is a non-exhaustive list of three expansion options.

The first option is ISA bus cards (or PCI). There are many low priced ISA bus cards that contain one or two additional serial ports. This option is only available to desktop systems.

The second option is PCMCIA communication cards. Even though the PCMCIA protocol was originally designed for laptop systems, adapters exist for even desktop systems. These adapters are more costly due to the small laptop market and their small form factor. The additional overhead comes with accessing the adapter through PCMCIA card services. The selection is also limited in terms of manufacturers. Depending on the software requirements, this may or may not introduce

more problems for the software developer.

The last option is the universal serial bus (USB). The USB is a brand new specification for serial communications. The USB port is flexible and allows the connection to numerous peripherals. However, since this is a new specification, the suppliers are limited and requires a newer revision of the Windows 98 operating system. More ports can be extended from a single port through the use of hubs.

Time synchronization might be needed to synchronize time between loop stations and between computers and video cameras. Even though computers and camcorders contain internal clocks, these clocks might not have adequate accuracy in terms of the granularity of time and the drift tolerance. Different time sources exist to act as an absolute time source. These sources include modem time from NIST (National Institute of Standards and Technology), broadcast radio time from WWVB (Colorado, USA), high-accuracy external clock, internet based (TCP/IP) time protocol, and satellite time from GPS (Global Positioning System) receivers.

The GPS receiver was utilized in this research because of its performance, price, and availability. The GPS receiver used must have the following capabilities in order to perform field data collection duties:

1. RS-232 output to communicate with data loggers. This format can be variable, although the ability to poll the receiver (i.e. full-duplex RS-232) is greatly desired. NMEA (National Marine Electronics Association) is a well known format that is suitable for field data collection. Other formats include proprietary protocols such as Trimble Standard Interface Protocol (TSIP).
2. RS-232 output to communicate with video time stamping equipment. NMEA, TSIP (Trimble Standard Interface Protocol), or SMPTE (Society of Motion Picture and Television Engineers) is required for this application. The current NMEA release is 0183, and some message formats for that release are described later in this chapter. The standard NMEA communications specifications is 4800 baud, 8 bit data, no parity, and one stop bit. The

SMPTE output in the Trimble ScoutMaster is at 1200 baud, 8 bit data, no parity, and one stop bit.

3. Time resolution greater than 1/30th of a second. UTC time sent by GPS satellites has an accuracy down to 1 microsecond. However, if video equipment is the limiting factor, then a resolution of a little less than 1/30th of a second is needed. The reason the accuracy is not exactly 1/30th of a second is due to the color NTSC frame rate of 29.97 frames per second.
4. Form factor appropriate for field data collection. Some GPS receivers are produced mainly for original equipment manufacturers (OEM), consequently they are not properly packaged for field testing.

There are currently four major GPS receiver manufacturers, not counting the ones who re-package OEM units. They are Trimble, Magellan (Ashtech), Garmin, and Lowrance. Table 3.1 shows a comparison between different GPS units.

The Trimble Scoutmaster GPS was found to be the most suitable for data collection needs in this research. This unit uses Master Clock or atomic clock performance to synchronize video and sound equipment in multiple camera shot.

Table 3.1. Evaluation of GPS Receivers

Model	Brand	Price	Timing Accuracy	Interface Capability	NMEA Sentence	Notes
ScoutMaster (land navigation and field mapping)	Trimble	$500	UTC +/- 1µs	NMEA183 TSIP 1PPS Clk. SMPTE	APA, APB, BWC, GGA, GLL, GSA, GSV, RMB, RMC, VTG, WCV, XTE, ZTG	
GeoExplorer II (GPS mapping system)	Trimble	$3500	UTC +/- 1 µs	ASCII, RTCM SC-104, XMODEM, NMEA183	(*)	

SveeSix (OEM module)	Trimble	$995	1 µs	TSIP, NMEA183, TAIP	GGA, VTG, optional: GLL, ZDA, GSA, GSV, RMC	Price is for development package that includes enclosure, antenna, software, and cables
Acutime (6 channel GPS smart antenna)	Trimble	$500	UTC +/- 1 µs	TSIP, RTCM, 1pps clk.		
Thunderbold	Trimble	$995	UTC +/- 50 nanoseconds (for 1pps clk.)	TSIP, binary protocol		Price is for development package that includes enclosure, antenna, software, and cables
G12 GPS Sensor	Ashtech	$2495 $2995 (base)	+/- 1 µs	NMEA183, RTCM, Ashtech OEM, 1pps blk.	(*)	20hz update rate available
G8	Ashtech	$2000	1 second +/- 1 µs for 1pps clk.	NMEA183, RTCM, Ashtech OEM	(*)	
GPS2000XL (4000XL)	Magellan	$300	1 second	NMEA183, RTCM	GSA, APB, GGA, GLL, GSA, GSV, RMB, RMC	
Garmin 12XL	Garmin	$300	1 second	NMEA183, RTCM	GGA, GSA, GSV, RMB, RMC, RTE, WPL	

(*) The NMEA sentence capability for some receivers were not specified in the receiver data sheets.

The NMEA produces the specifications for the NMEA 0183 standard which is used by several manufacturers of time synchronization equipment and GPS receivers. The NMEA 0183 was first released in March of 1983 and defines electrical signal requirements, data transmission protocol,

timing, and specific sentence formats for a 4800 baud serial data bus.

The NMEA sentence strings are the ASCII strings outputed by GPS receivers and contain different information dependent on the needs of the user. The following is a description of some relevant sentence formatters:

GGA (Global Positioning system Fix Data) - Time, position, and fix related data

$__GGA,hhmmss.ss,llll.ll,a,yyyyy.yy,a,x,xx,x.x,x.x,M,x.x,M,x.x,xxxx

hhmmss.ss	UTC of position
llll.ll	latitude of position
a	N or S
yyyyy.yy	longitude of position
a	E or W
x	GPS quality factor, 0=no fix, 1=GPS fix, 2=Dif. GPS fix
xx	number of satellites in use
x.x	horizontal dilution of precision
x.x	antenna altitude above mean-sea-level
M	units of antenna altitude in meters
x.x	geoidal separation
M	units of geoidal separation in meters
x.x	age of differential GPS data in seconds
xxxx	differential reference station ID

ZDA (Time & Date)

$__ZDA,hhmmss.ss,xx,xx,xxxx,xx,xx

hhmmss.ss	UTC of position
xx	day, 01 to 31
xx	month, 01 to 12

xxxx	year
xx	local zone description, 00 to +/- 13 hours
xx	local zone minutes description (same sign as hours)

The selection of the type of loop is often not a possibility, and one is obligated to use the existing loop infrastructure. One commonly used loop is a six feet by six feet square loop. Figure 3.4 shows such an example of the square loop configuration in an arterial test site. In Germany, there are loops which are only one meter in length along the travel direction. These loops perform less integration of the waveform signal and give a less smoothed vehicle waveform. If customization is possible, then a loop smaller in length in the travel direction should be used to avoid smoothing out transition points in the waveform. However, the advantage of using the existing inductive loop infrastructure is not exploited with cutting new loops.

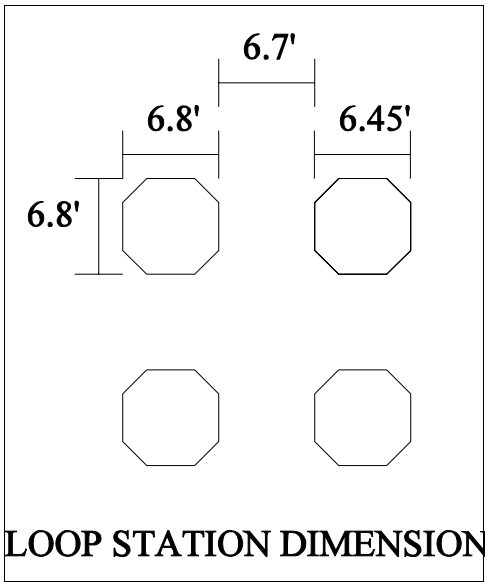


Figure 3.4 Example of Loop Station Configuration

The detectors that are required for collecting vehicle waveforms are newer generation NEMA TS-1/2 or Caltrans 170 compatible detectors. They need to have the capability of transmitting the

inductance count or inductance change via a serial port. Due to different traffic equipment specifications used in Europe and in the United States, an adapter is necessary for using NEMA form factor equipment in DIN racks.

The City of Irvine has installed advanced Speed Occupancy & Headway (S.O.H.) detectors in parts of the Irvine Field Operational Test (FOT) region, mainly on Alton between Sand Canyon and Technology East. These detector racks contain a power supply, a SC1 Communications Interface Module, and MTS38z two channel detectors. The MTS38z is a "EURO card" and as such uses a 64-pin DIN connector. A conversion adapter is needed in order to use standard NEMA detectors such as the 3M Canoga C824 or the Peek GP6. For the field data collection, only relevant pins are mapped with the remaining pins left disconnected. Since the detectors were only connected to data logging equipment and not controllers, all loop outputs were left disconnected.

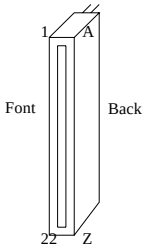


Figure 3.5. NEMA Connector

Table 3.2. NEMA Connector Pin Assignments

Pin Label	Pin Position	Standard Function	Field Test
A	1	DC Ground	DC Ground
B	2	+24 VDC	+24 VDC
C	3	Reset	N/C
D	4	Channel 1 Input	Channel 1 Input
E	5	Channel 1 Input	Channel 1 Input
F	6	Channel 1 Output (+)	N/C
H	7	Channel 1 Output (-)	N/C
J	8	Channel 2 Input	Channel 2 Input
K	9	Channel 2 Input	Channel 2 Input
L	10	Chassis Ground	Chassis Ground
M	11	N/C	N/C
N	12	N/C	N/C
P	13	Channel 3 Input	Channel 3 Input
R	14	Channel 3 Input	Channel 3 Input
S	15	Channel 3 Output (+)	N/C
T	16	Channel 3 Output (-)	N/C
U	17	Channel 4 Input	Channel 4 Input
V	18	Channel 4 Input	Channel 4 Input
W	19	Channel 2 Output (+)	N/C
X	20	Channel 2 Output (-)	N/C
Y	21	Channel 4 Output (+)	N/C
Z	22	Channel 4 Output (-)	N/C

Note that only back side signals (A-Z) are used and shown in Table 3.2. Front signals included communications pins that might be useful in other applications. Also note that only 11 pins are connected for field data collection.

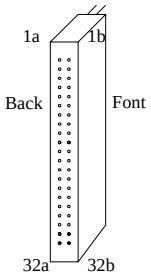


Figure 3.6. DIN Connector

Table 3.3. DIN Connector Pin Assignments

Pin Number	Field Test Function	Pin Number	Field Test Function
1a	N/C	2b	N/C
3a	N/C	4b	N/C
5a	Channel 1 Input	6b	Channel 1 Input
7a	N/C	8b	N/C
9a	N/C	10b	N/C
11a	N/C	12b	Channel 2 Input
13a	Channel 2 Input	14b	Chassis Ground
15a	N/C	16b	N/C
17a	N/C	18b	N/C
19a	N/C	20b	N/C
21a	N/C	22b	N/C
23a	N/C	24b	N/C
25a	N/C	26b	N/C
27a	N/C	28b	N/C
29a	N/C	30b	+24 VDC
31a	N/C	32b	DC Ground

In Table 3.3, note that only odd pins for the back side and even pins for front side are shown. Even pins for back side and odd pins for front side can be left disconnected.

Table 3.4. DIN to NEMA Adapter

DIN Pin Number (*)	NEMA Pin Label	Pos	Field Test Function	Wire Color
5a (1)	D	4	Channel 1 Input	Blue
6b (1)	E	5	Channel 1 Input	Blue
12b (1)	J	8	Channel 2 Input	White
13a (1)	K	9	Channel 2 Input	White
14b (1)	L	10	Chassis Ground	Green
30b (1)	B	2	+24 VDC	Red
32b (1)	A	1	DC Ground	Black
5a (2)	P	13	Channel 3 Input	Orange
6b (2)	R	14	Channel 3 Input	Orange
12b (2)	U	17	Channel 4 Input	Yellow
13a (2)	V	18	Channel 4 Input	Yellow

In Table 3.4, note that the number in parenthesis indicates the first or second connector. Since current City of Irvine wiring only connects two loops to each 2 channel MTS38z detector, each 4 channel NEMA detector will be mapped to 2 DIN connectors. For NEMA PCB edge connector use CINCH 50-44S-30 or equivalent 22 positions/44 contacts. For DIN use HIROSE PCN10-64P-4.54DSA (straight plug), AMP 533248-1 (right angle), or other 64 pin two row connectors.

The back of the City of Irvine detector racks is a EURO4 Detector Motherboard. Table 3.5 shows the relevant pins needed for data collection from this 21 pin connector.

Table 3.5. EURO Motherboard Pin Assignments

Pin Position	DIN Pin Number	Function
1	5a	Channel 1 Input
2	6b	Channel 1 Input
3	12b	Channel 2 Input
4	13a	Channel 2 Input
19	30b	+24 VDC
20	32b	DC Ground
21	14b	Chassis Ground

In addition to the proper hardware setup, data collection software is also needed to handle the data stream from the detectors. Figure 3.7 is a graphical representation of the data collection software

.

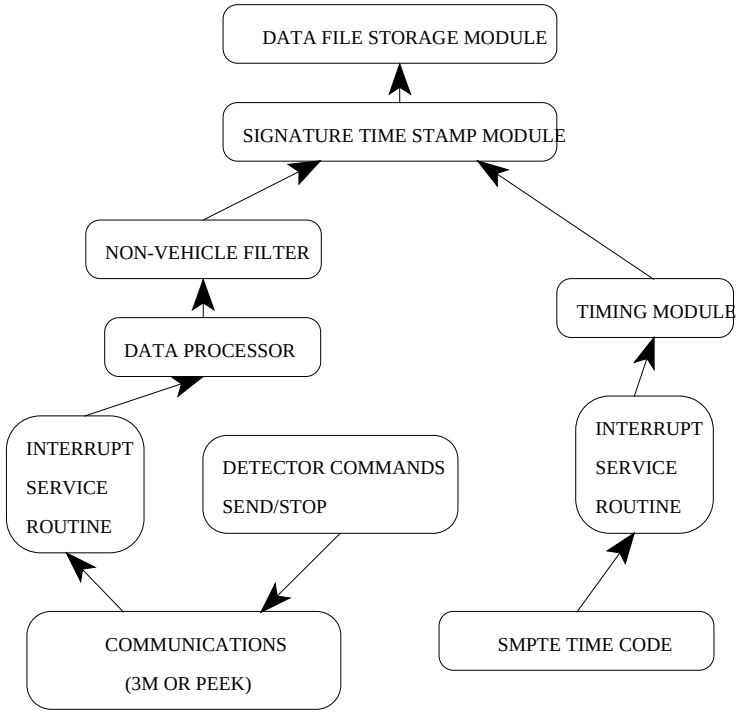


Figure 3.7. Data Collection Software Components

This software has two major components: datalogging routines and time synchronization routines. The datalogging portion of the software is a simple serial driver implementation for RS-232 communications between the computer and the inductive loop detector.

The software should have modules to perform the following tasks:

a. send detector commands (send signature/stop signature)
b. interrupt service routine to receive and store incoming signature messages
c. data processing routine to strip inductance change information from raw data stream
d. filter out non-actuations (a set threshold of inductance change) to save storage space
e. store resulting time stamped waveforms in data files

Note that items a-c are detector specific, while items d and e uses non-detector specific formatted inductance change data. In other words, there will be separate implementations for different manufacturers. The time synchronization portion of the software should receive reference time from a serial port and append it to the waveform data.

The careful design of the video ground truthing system will facilitate the process of manual correlation of vehicles in the laboratory. Whether an absolute or a relative time clock is used for the video, time stamping each frame will allow frame by frame analysis. Time stamping for video is usually in the form of SMPTE time code. SMPTE was originally developed for use in the television and motion picture industry. The original intent was to frame synchronize video and sound equipment. The SMPTE Time Code is formatted to provide a system wide clock for reference by each individual piece of equipment. The Time Code is usually distributed via standard audio equipment or encoded directly onto the video signal. The time code has the following format with four fields: HH:MM:SS:FF. Table 3.6 shows the value ranges for each time code field, and table 3.7 explains the four standard frame rates.

Table 3.6. Time Code Representation

Time Code	Representation
HH	Two digits representing hours. Values range from 00 to 23.
MM	Two digits representing minutes. Values range from 00 to 59.
SS	Two digits representing seconds. Values range from 00 to 59.
FF	Two digits representing frames or frame rate. Values range from 00 to 29 (for 30 frames per second).

Table 3.7. Frame Rates

Frame Rate	Resolution (second)	Explanation
24	1/24th	Frame rate is based on U.S. standard motion picture film.
25	1/25th	Frame rate is based on European motion picture film and video. Also know as SMPTE EBU, and is compatible with PAL/SECAM color and b&w.
30	1/30th	Also known as 30 frame non-drop. Frame rate based on U.S. NTSC black & white video.
30 Drop Frame	1/30th	Frame rate is based on U.S. color video. Used by most non-broadcast applications since it matches real time.

The present U.S. color television standard has a frame rate of 29.97. In order to correct for real time clock drift, a drop-frame scheme was devised in which the time code clock is advanced ahead two frames a minute except on minutes ending in 0 (i.e. 00, 10, 20, 30, 40, and 50). This compensates for the 108 frames an hour, or 3.6 seconds that are lost. In terms of accuracy, the worst case drop-frame time code drift would be approximately the two frames that are dropped or 66.7ms (33.37ms/frame). The drop frame format should be used in field data collection and data post-processing as it is the only format that is appropriate for real-time synchronization. The drop frame format is indicated in the Horita time code generator unit by a semicolon (;) before the last digit

places, while the non-drop frame is indicated by a colon (:).

Other possible sources of time error include a possible 1 frame error at the start of the recording and video color crystal oscillator error. If an FCC (Federal Communications Commission) standard specification of 3.579545 Mhz. +/- 10Hz. is assumed for the color oscillator generated frequency, then a +/- 0.01 seconds per hour or equivalently 1 frame per 3.3 hours error could result. Taking into account all the possible sources of error related to SMPTE generation equipment, the maximum amount of error would be 3.33 frames (1 start + 2 drop-frame + 0.33 oscillator) or 0.11 seconds for a recording time of one hour. Since the error sources are not linear, the error would be 0.12 seconds for a recording time for two hours for example. In order to understand the effects of such an error on traffic measurements, a sample scenario is assumed. Given a roadway section of 1 mile with average speeds of 40 mph. The resulting average travel times is 90 seconds. The previously computed maximum SMPTE time error would then result in 0.1% travel time error which is quite small for travel time applications.

Standard NTSC video signals are interlaced to produce the approximate 60 hertz cycle, so two fields are refreshed approximately every 30 hertz to form the complete image. In the Horita unit, field-1 is indicated by the display for a period (.) while field-2 is indicated with a semicolon (;) or colon (:).

SMPTE time code data include the time code, sync data, frame data, and user information data. The sync data marks the end of frames and notes whether the videotape is moving forward or in reverse. The frame data shows whether drop frame is used and/or if color framing is active. The user information is just additional data that can be encoded, such as filming date or location.

There are two types of SMPTE time code: Longitudinal Time Code (LTC) and Vertical Interval Time Code (VITC). LTC is an audible electronic digital signal recorded on an audio or time code channel of a VCR or audio recorder. VITC is a video frame identification code recorded in the vertical blanking interval of each video field. VITC must be recorded at the same time as the video and can be read during VCR pause mode, while LTC can only be read when the tape is in motion.

The object of the real-time synchronization is to be able to coordinate the field data from multiple sites. Each site would have a configuration illustrated in Figure 3.8. This coordination allows the derivation of more accurate parameters using the video data. The UTC time also serves as the master clock reference for all field data collection equipment.

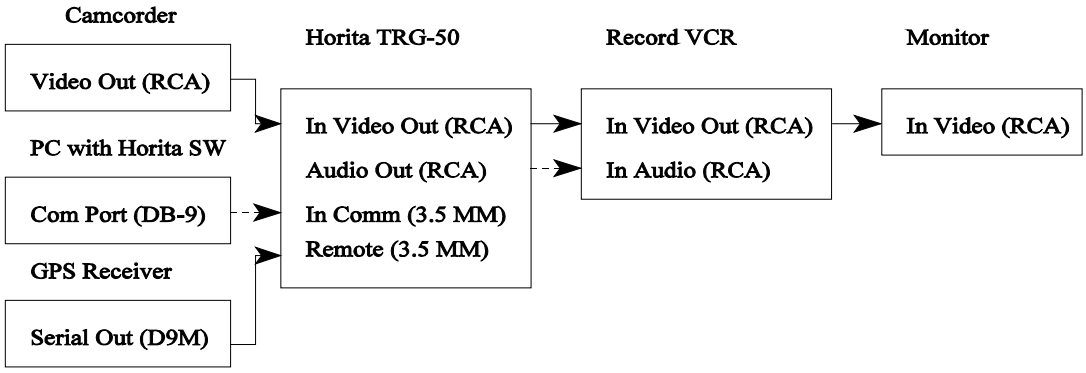


Figure 3.8. Real-time Synchronization

In Figure 3.8, the solid lines indicate necessary connections, while dotted lines indicate optional connections. The type of connector necessary is indicated in parenthesis. The video connections are the most critical in order to do a "window burn" of the time code on the original image. The SMPTE time code should be recorded on the audio channel if reading of the time code is desired. An optional PC unit can facilitate the initialization of the time code and the entering of user information. The monitor is used to verify the proper functioning of the time code generation and to help determine the location and size of the time code window. In real-time synchronization, UTC time is obtained from a GPS receiver.

The real-time synchronization proceeds in the following stages:

1. Enable the GPS receiver to allow the acquiring of clean satellite signals.
2. Start Horita TRG-50.
3. Verify that the GPS time matches the window burned time code.
4. Start recording of traffic data.

The object of the in-lab post-production is to append SMPTE time code in the original video image. This allows each frame of the video data to be uniquely identified, and consequently, individual vehicles can be located based on its loop crossing time. The SMPTE time code is both a source of reference and of time data for computation purposes. Figure 3.9 shows the connections for the in-lab post-production setup. The solid lines indicate necessary connections, while the dotted lines indicate optional connections.

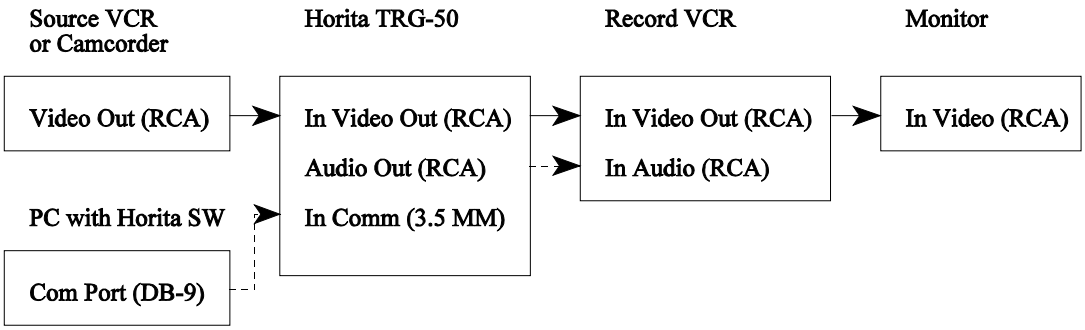


Figure 3.9. In-lab Post-production Setup

The post-processing has the following stages:

1. Determine the real time of the beginning of the video and find the first second transition frame. The camcorder time stamp with second resolution should be recorded in the original video.
2. Set the TRG-50 to the real video time. Make sure that drop-frame mode is enabled (i.e. a semi-colon is displayed in the last time position).
3. Start both the VCR and the TRG-50 generator at the same time. This will take some experience because of the VCR latency in responding to commands. The VCR can be started from both stop and pause modes; however, the TRG-50 does not perform as well in pause mode.
4. The delay in frames between real time and the time code can be found by frame advancing the "window burned" tape after its completion.

In evaluating video equipment the trade offs between resolution, compatibility, and cost need to be considered. Tables 3.8 and 3.9 summarize different video formats and connectors. Care must be taken to ensure that the proper video format is used for all equipment. A high quality recording system is useless if unaccompanied by a high quality display system. Direct baseband connections are recommended for all equipment.

Table 3.8. Video Tape Formats

Format	Upper Resolution	Explanation
Hi-8	425+	Improvement on the 8mm. Higher quality tape is also needed.
SVHS	400-425	Super VHS. Enhanced version of VHS by using higher quality tape formulation.
Standard US-TV	350	
8mm	300	Videotape format promoted by SONY.
VHS(SP)	240-250	Video Home System. 1/2" tape. Promoted by JVC and others.
VHS(EP)	220	

Table 3.9. Video Connector

Type	Explanation
RF	Radio Frequency. Combined audio and video signal as a broadcast signal. Lower quality than baseband connections.

RCA BNC	RCA or phone plugs, and BNC (British nut connector) carries baseband composite signals. BNC connectors are used with coaxial cables while RCA is used with standard cables. An adapter easily converts between the two types of connectors.
S-Video	Also known as a Y-C connector. Carries separate luminance (brightness) and chrominance (color) information. Is higher quality than composite.

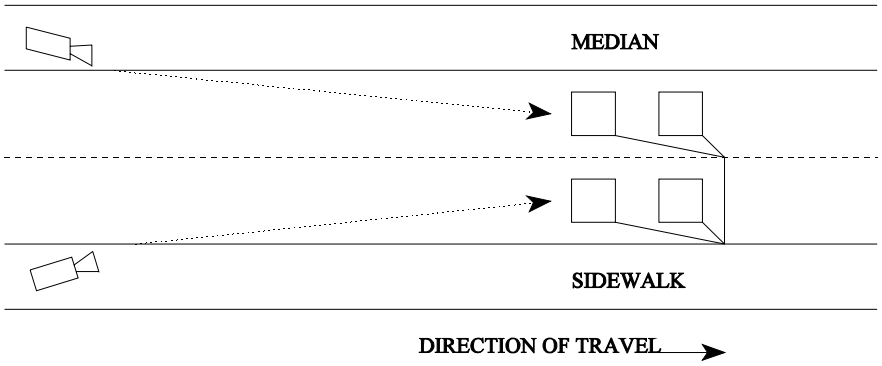


Figure 3.10. Camera Placement

To facilitate the manual vehicle correlation process, good resolution of the vehicles and a consistent field of view among different stations is necessary. The ideal camera setup is to have two cameras on either side of the roadway. In a two lane arterial case, each camera will be focused on a single lane. The back of the vehicles should be recorded instead of the front because many vehicles lack the front license plate. Figure 3.10 shows the recommended camera setup for arterial data collection.

A comparison was made between regular VHS and Hi-8mm format. Even after reduplication of the Hi-8mm tape in VHS format, the quality was found to be much better with the Hi-8mm recording system. Manual focus is recommended to avoid the auto-focusing mechanism from "hunting". If a camera is dedicated to a single lane, then the resolution will be adequate for recognizing most license plates.

Similar angles are needed so that vehicles from different lanes can be compared upstream and downstream. Thus, the camera should be situated as far away from the loop site as possible without loss of resolution. Note that digital zooming is of lower quality than true optical, and its use should be avoided. The ideal view results from the camera shooting straight at the back of vehicles. Also, the loop location should not be close to an intersection, since the proximity to the intersection will lead to significant occlusion from the left turn lanes in the upstream section.

3.2 SR-24 FREEWAY DATA

The test data were obtained from a field site on the westbound SR-24 freeway in Lafayatte, California in December 1996. Two data acquisition stations were instrumented with video, loop waveform dataloggers, and speed trap dataloggers. Standard 6ftx6ft (1.82mx1.82m) loops were used at both stations. Several hours of data were collected, but a smaller portion of the data was reduced into two datasets and used for this initial investigation into vehicle reidentification. The reduced dataset contains the waveforms of the upstream and downstream vehicles along with their speeds, electrical length (derived from occupancy time), arrival time at stations, and the ground truth vehicle identification number. One dataset was composed of moderate flow traffic (1000 VPHPL) and contained approximately 2000 vehicles. This dataset was recorded on December 6 at approximately 12:00pm. Another dataset was composed of congested flow traffic (1800 VPHPL) and contained approximately 3000 vehicles. The second dataset was record on December 12 at 8:00am during the morning rush hour. Both data sets were divided into training and testing data sets. The calibration of the algorithms was performed on the training data only. Figure 3.11 shows the data collection setup for the SR-24 site.

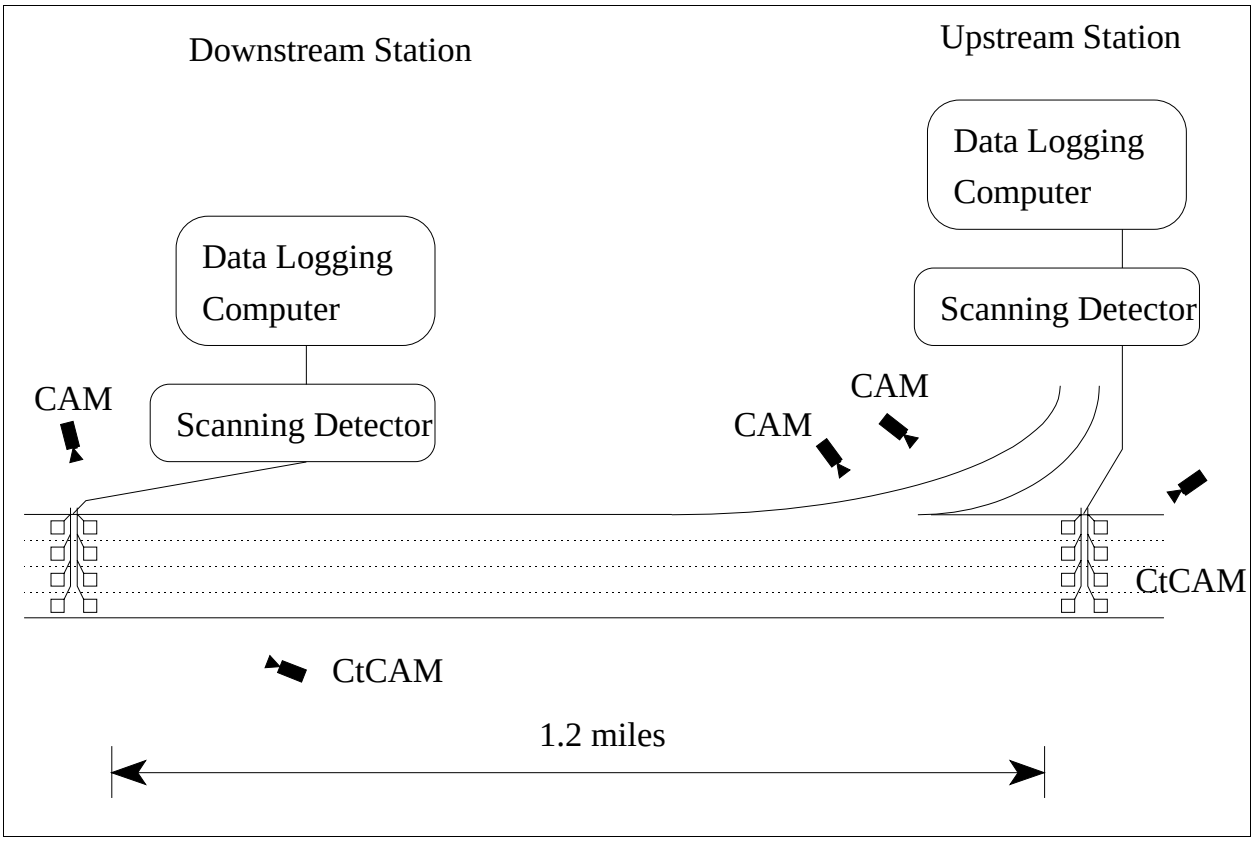


Figure 3.11. Lafayette Data Collection Site

Arterial data were also collected on Alton Parkway in Irvine, California. A similar setup to SR-24 was used where two loop stations were located on the direction of travel. These data were not used in the initial development of the vehicle reidentification algorithms.

The test data were stored in SIG (signature) files and ground truthed by video correlation. The unprocessed signature files and PVR (Per Vehicle Record) files were collected in the field. The video correlation database was produced in the laboratory from the video footage taken during data collection. Manual correlation of the downstream with the upstream vehicles using video was required for the development of the video correlation database. Due to discrepancies in time stamping the loop data and the video, another manual correlation was required to match the video correlation vehicle records with the vehicle loop signatures. The last step in the process of creating

the test data was to append unique identification numbers to the signature data that could be referenced back to the video database.

The format of the video correlation database will be described in terms of each table column. A sample video correlation database segment is contained in appendix A for reference. The video correlation database is stored in Microsoft Excel Worksheet format.

"A" - Consecutive vehicle ID number. These numbers are assigned sequentially to the upstream vehicles.

"B" - Indicates lane numbers as 1,2,3,4, OR (on-ramp), and S (shoulder). Lane 1 is closest to the median. Any combinations of lanes (eg. 2>3) indicates vehicle going from one lane to another.

"C" - Indicates upstream time at which the vehicle traverses the loop. Format is hour:minutes:seconds:frames (0 to 30, NTSC). Due to the limitations of Excel the following Excel number format was chosen for the time stamp: 00:00:00:00, where each 00 combination represent digit place holders. This format was chosen because it allows sorting according to time stamp value, and it is easy to enter in Excel without the need to type colons between fields. A custom format hh:mm:ss:ff was tried, but the frame count, ff, was non-standard. A third option was to use a separate column to store the framecount, but that was even more unconventional than the previous option. In all cases, simple arithmetic involving time stamps is not possible since the frame count field is non-standard.

"D" - X indicates any occlusions from the sideview angle (eg. color monitor). The auxiliary black and white monitor was almost never occluded for any vehicles.

Fields E-G are fields corresponding to the matched vehicles downstream.

"E" - same as "B" field but for downstream

"F" - same as "C" field but for downstream

"G" - same as "D" field but for downstream

"H" - vehicle types

C - car, passenger vehicle

SUV - sports utility vehicle (eg. Explorer, Cherokee, 4-runner, etc...)

MV - minivan (eg. Previa, Aerostar, Caravan, etc...)

P - pickup

V - van, delivery/cargo van

T - truck (eg. Ford F250, etc...)

SEMI - any tractor-semitrailer combinations

SW - stationwagon

B - bus

M - motorcycle

"I" - colors used and abbreviations

WH - white

BLK - black

GR - green

BR - brown

YEL - YELLOW

ORG - ORANGE

PURP - PURPLE

BLUE

RED

TAN

GOLD

GRAY

SILVER

LT - light

DK - dark

"J" - other helpful descriptions, such as make/model

STR - stripe or side molding

WPS - wooden side panels

HB - hatchback

Any problematic visual matches are denoted with "?". This means that this data point set should be dropped from the signature database.

Due to the limitations of the video resolution and camera differences, fields I and J are only best estimates of the vehicle color, make/model, and other features. These fields were mainly used during the video correlation procedure but are retained for completeness.

The next section is an explanation of the SIG and PVR files. Observations and assumptions for several parts are listed to help clarify problems that were encountered during the project. A sample PVR entry and the SIG entry of the same vehicle will be used to facilitate the explanation. More extensive samples of the PVR file and the SIG files are contained in appendix B and C.

```
Normal  12:27:12.114  1.83  1567    1          33.7    4.46       Car    208    208

# Record 3873, lane 1   1926
# Normal  12:27:12.114  0.00  1567    1          33.7    4.46       Car
3553.114        48
3553.127        136
3553.140        358
3553.153        580
3553.166        904
3553.179        1228
3553.192        1397
3553.205        1567
3553.218        1519
3553.231        1343
3553.244        1167
3553.257        830
3553.270        492
3553.283        336
3553.296        74
3553.309        17
```

Figure 3.12 Sample PVR and SIG Record

The meanings of each part of the SIG file will be listed first. The number signs on the first two lines indicate the message header portion of each record. The remaining lines comprise the data points which are used to create the vehicle signature waveforms.

On the first line, 3873 is a sequential record number assigned by the consultant's data collection software. This sequence is in order of the time the vehicle was last detected by the loops, not the order that the vehicles first crossed the loops. Therefore, many large vehicles are out of sequence by number, but not by time. The same can be said for some of the slower vehicles during the peak hour time. If it is assumed that the computer asked for the PVR information after the vehicle completely crossed the loop, then this sequence number makes sense. It is possible that the computer may have confused time and signatures if this was not the case. Lane 1 is the lane number that the vehicle was travelling in. The next field value 1926 is the matched vehicle ID from the video database. If the vehicle was matched successfully, then the upstream and downstream signature files should contain a signature identification number such as 1926. If the identification field has a question mark, it means that there is a problem with the signature and is not matched.

The second line is essentially a mirror image of the PVR file with a few notable exceptions.
The first field, Normal, lists whether the vehicle is in one or two lanes. If the vehicle was straddling lanes, the word "DOUBLE" appears instead of "Normal". Note that this did not necessarily mean that the vehicle was straddling lanes. We noticed that the word "DOUBLE" appeared whenever the maximum vehicle magnitude was very low and the vehicle length was small. The next field, 12:27:12.114, is the time stamp of the vehicle when it was first detected. The next field differs between the PVR and SIG files. The time offset 1.83 is computed from the last detection and is for PVR files only. The time offset did not appear in the SIG files. It is logical to assume that the PVR file was created in a spreadsheet in order to quickly calculate the time since the last detection, then that line was pasted as the second line on the SIG files erasing the formula. Therefore, for all SIG files, the offset is 0.00. The next field is the max amplitude value of 1567 recorded from the loop detectors. The lane number 1 is repeated in the next field. Next is the speed, 33.7, in meters/second for the front of the vehicle. This speed measurement does not take into account that the vehicle may

slow down or be stopped, and therefore some of the vehicle types listed by the computer are different than what we observed on the video. Next is the value 4.46, which is the length in meters. "Car" is the value for the field that is derived from a simple classification using vehicle length. The classification uses the following lengths: MotorCycle <3.0m (generally)
Car 3.0m-4.7m
SmallVan 4.7m-6.0m
SmallTruck 6.0m-8.0m
RigidTruck >8.0m
For signatures with two or more peaks:
RigidTrucks <11.0m
Artic >11.0m
The next two fields with the values 208 and 208 are proprietary fields.

The remaining lines on the signature files are the time in seconds and the magnitude. The time is referenced from car number 1 which receives time 0.XXX where XXX represents the decimal fraction of a second in real time. So if car 1 crossed the detector at 11:22:31.545, then the first time would read 0.545, and all other times would be referenced back to 11:22:31.000. Car # 3873 is therefore 3553.114 seconds later than the zero time. Magnitude was recorded every .012 to .014 seconds in most cases. For those vehicles that spent more than 10 seconds over the loop, the time was recorded backwards, recording a number every 0.001 seconds.

For most vehicles a graph of magnitude vs. time can be created. As long as the vehicle was moving at a constant speed or at a gradual change in speed, waveforms that represent the vehicle signature can be created from the SIG files. If the vehicle stopped or significantly changed speeds over the loops, then strange waveforms may appear.

The next section describes the process used to match vehicles at the upstream and downstream locations. To ensure correlation consistency and accuracy, a highly detailed procedure was followed throughout the entire process.

The workstation setup for the video correlation process is divided into two stations: the upstream station and the downstream station. Each station consists of two different size monitors and two VCRs. Each station is equipped with a different type of VCR that allows the two stations to work independently. Otherwise, the stations are mirror images of each other. The large monitors show the side view of the vehicles at each location. This footage was shot in color by the field assistants. The small monitors show the rear view of the vehicles at each location. This was shot in black and white by cameras already installed by Caltrans. The large monitors were placed right next to each other. Each of the large monitors had a yellow marking across the roadway to indicate loop detector locations.

The upstream data points such as the vehicle identification number, the upstream time stamp, occlusion status, type of vehicle, color, and noticeable features of each upstream vehicle were recorded directly onto the data reduction form, which has 25 records per sheet. Because the side view was at an angle from the direction of travel, it was very difficult to tell exactly when a vehicle first crossed the loop. Therefore, the front tire was used as the marker to identify the times of each vehicle. And since the upstream loop detectors were at an angle instead of straight across the road, it was difficult to determine exactly where the loop detectors were. The loop marking was placed as close as possible to the loop detector locations based on the cuts that could be seen on the road.

The downstream data points were recorded on the downstream data reduction assistance form which has 45 records per sheet. One sheet was used for each lane of traffic to simplify the locating of potential matching vehicles. The camera was aimed perpendicular to the flow of traffic on a slope above the freeway so that more accurate observations could be made. The time was recorded downstream when the vehicle just barely crossed the line.

A system was also developed to keep track of the time spent processing each sheet. For the upstream data, the time was noted on the upper left corner of the sheet. On the last sheet in a sequence the end time was recorded on the lower left corner of the sheet. For the downstream data, the start of the recording time was noted next to the first vehicle on all four sheets. When each sheet was finished

the time was noted on the beginning of the next sheet. After finishing the downstream data, the finishing time was noted on all four sheets.

Once the upstream and downstream data were available, each upstream vehicle was lined up to be matched at the downstream location. First, the lane the vehicle was in at the upstream station was looked at in the corresponding lane in the downstream station to find a match. Based on observations, many vehicles did not change lanes and stayed in order so it was not usually very difficult to find the matching vehicle. The monitors from both stations were used equally to correlate the vehicles. The information for the matched vehicle shown on the downstream station was then recorded on the data reduction form. Then, the vehicle id number was written next to the matched vehicle on the downstream form to indicate that the vehicle was matched.

All data were double-checked to ensure that they were recorded accurately. First, during vehicle matching, each vehicle on the upstream station was lined upon again on video to ensure that the upstream data was correctly written down. Once a set of cars was matched, the downstream data written on the data reduction form were compared to the information written on the downstream data reduction assistance form. Any discrepancy was corrected by reviewing both upstream and downstream videos.

In order to minimize confusion between several similar vehicles, the process was organized into sets of 200-400 vehicles. Each set was completed within 6-10 hours depending on the number of occluded vehicles.

The last step in the formation of the signature database was to establish a one-to-one correspondence between the vehicles recorded on video and the vehicles recorded from the loop detectors. This step is necessary due to the fact that the video was not time synchronized with the datalogging computer.

The video database is sequential in terms of the upstream arrival times. Thus the vehicle identification numbers are also sequential. However, the downstream time stamps are not sequential

with respect to the vehicle identification numbers because of traffic movements such as lane changes that occurred after the upstream station. Therefore the video database needs to be sorted according to the downstream timestamp in ascending order. Printouts of the sorted and unsorted versions of the video database will show the vehicle arrival sequences for both upstream and downstream.

The corresponding time periods from the PVR file are also produced. A column is inserted in the front of the PVR database for adding the vehicle ID number. This file is also sorted according to the time stamp. The result is one PVR file in the sequence of upstream vehicle arrivals, and another one in the sequence of downstream vehicle arrivals.

The following procedure was followed to match the video database with the PVR database:

1. Try to match lane sequences.
2. Check vehicles that are not passenger cars. Make sure magnitudes are lower for non-passenger vehicles.
3. Tag any "out-of-sequence" lane patterns and decide if they should be dropped.

After the correspondence is established between the two databases, the vehicle identification number is appended to the end of the first line of the signature vehicle header. Any dropped vehicles were deleted from the signature database.

The principle cause of deleted vehicles was due to the steeply superelevated on-ramp at Central Lafayette. Even though this on-ramp was instrumented, most vehicles missed the loop and were not detected. Consequently, these vehicles were identified in the video database but were not recorded by the loop detector.

CHAPTER 4. LEXICOGRAPHICAL OPTIMIZATION APPROACH

4.1 FEATURE EXTRACTION

The vehicle reidentification algorithm developed takes into consideration real-life implementation issues. Specifically, an algorithm was developed so that it could perform in real time and not be computationally intensive. The second consideration was that the number of feature vectors should be small enough to be able to be transmitted to neighboring stations or a central location. Because the algorithm is constrained by an anticipated limited communications bandwidth and possible need for real time performance, simplicity was a main goal in its design. Complicated feature extraction schemes such as parametric estimation schemes (Ljung, 1987) were eliminated because of the computational burden. Frequency domain methods were not found to be suitable in producing useful features and were not used.

The selected algorithm uses simple but effective feature vectors derived by direct methods. The main feature vector actually consists of a set of vectors corresponding to the signature waveform of any vehicle. Note that the term "waveform" is used here according to standard electrical signal processing parlance, though it refers to a detector inductance output signal pattern. There are three main steps in deriving this waveform from the raw waveform which is obtained from sampling the inductive loop magnitude at intervals of every 11ms to 14ms (this was the scan rate of the detector cards used during SR-24 data collection):

Step1. The signal magnitude is normalized with respect to its maximum amplitude to eliminate upstream and downstream variations.

Step 2. In order to remove the effects of vehicle speed on the waveform, we transform the time axis into length by multiplying by the speed of the vehicle which is obtained from each double loop speed trap configuration.

Step 3. The waveform was modified (stretched/shrunk) in step 2, so that there is a need for re-sampling of the waveform. This is needed because the waveform magnitude vectors need to be compared at the same intervals. A spline interpolation (Shoenberg, 1946), was used

to fit the waveform and provide re-sampled data points. These sampled ordinates become the values in the waveform feature vector. The spline interpolation is a piecewise polynomial interpolation which requires the endpoints of the approximating polynomial to equal the endpoints of the piecewise function. In other words, given a function $f(x)$ on an interval $a \le f(x) \le b$, partitioned into subintervals $a=x_0<x_1<...<x_n=b$, then $g(x_0)=f(x_0)$, $g(x_1)=f(x_1)$, ..., $g(x_n)=f(x_n)$ for all approximating polynomials, $g(x)$, where $g(x)$ should be several times differentiable. Spline interpolation is chosen because it generally avoids the problem of numerical instability and is easy to implement.

These steps are illustrated graphically in Figure 4.1.

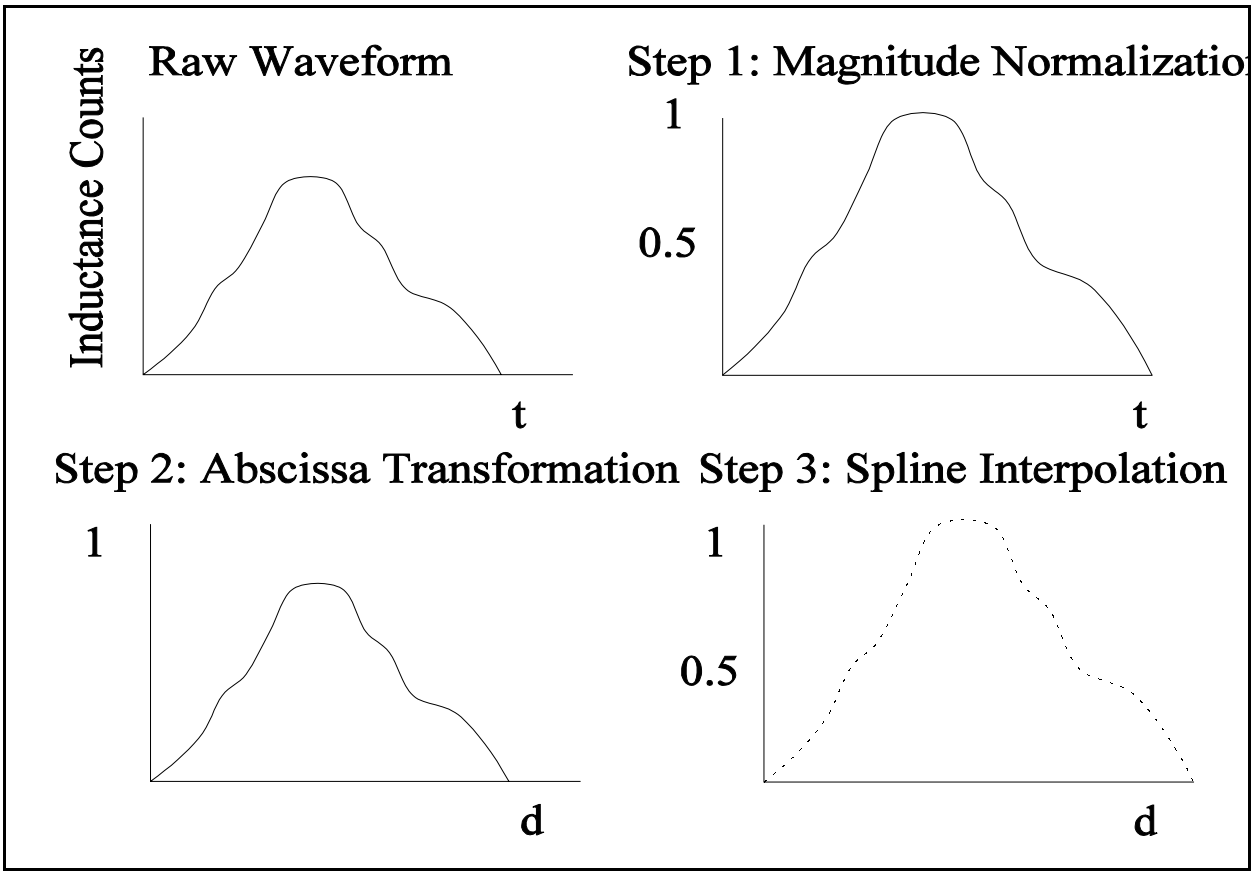


Figure 4.1. Steps to Waveform Processing

Even after the steps described above for waveform transformation, there is still variability in the upstream and downstream waveforms due to the following sources:

Input:

vehicle entrance angle into inductive field

vehicle offset from the physical loop center, even straddling (for example, a vehicle's inductance when covering half the loop can be reduced from 3500nH to 2000nH)

height from loop due to the suspension system

System:

inaccuracies in the measured speed from the speed trap configuration which might lead to an inaccurate length scale

errors in interpolation

aliasing due to sampling

quantization error in the analog to digital conversion (A/D) process

variability between upstream and downstream inductive loop systems including the physical loop installation and the loop energizing circuitry

non-LTI (Linear Time Invariant) characteristic of the ILD

In addition to the spline-interpolated waveform ordinates as part of the feature vector, several other feature vectors were also considered including vehicle speed, electronic length, maximum amplitude, Fourier transform, and lane number identification. As mentioned previously, the Fourier transform components were not found to be adequate to form a feature vector. Electronic length was also not used since its inclusion did not improve the accuracy of reidentification.

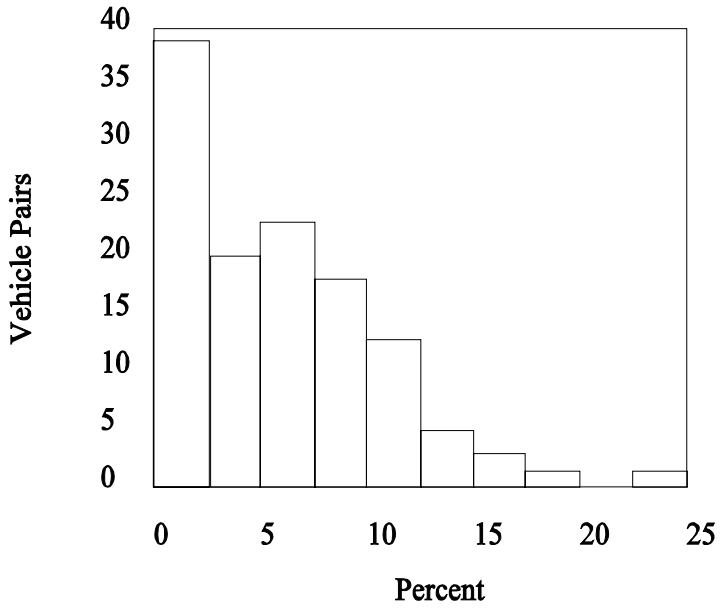
The vehicle speed was obtained from a speed trap configuration in each lane, both in the upstream and downstream stations. Other sources for speed such as Doppler radars can also be used to obtain speed. In Chapter 4, the use of single inductive loop speed estimation will also be discussed briefly.

The electronic vehicle length was computed by multiplying the speed by the total sampling period of the detector for that vehicle. The sampling period is simply the time of the first sample subtracted

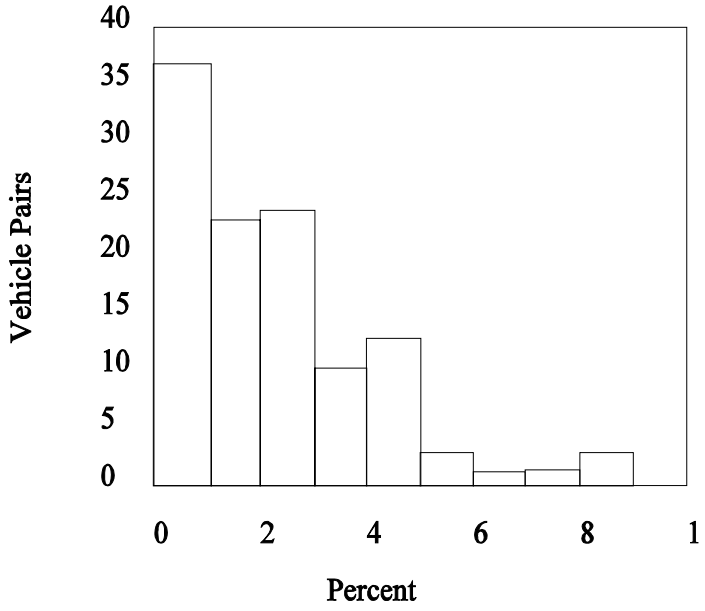
from the time of the last sample. This value is consistently higher than the physical length of the vehicle since it includes the magnetic length of the loop.

The value of the magnitude of the waveform is proportional to the inductance change produced by the presence of the vehicle. Even though the vehicle produces a net inductance decrease, the magnitude value is made positive to facilitate computation. The maximum amplitude is the highest magnitude value that was sampled for each individual vehicle.

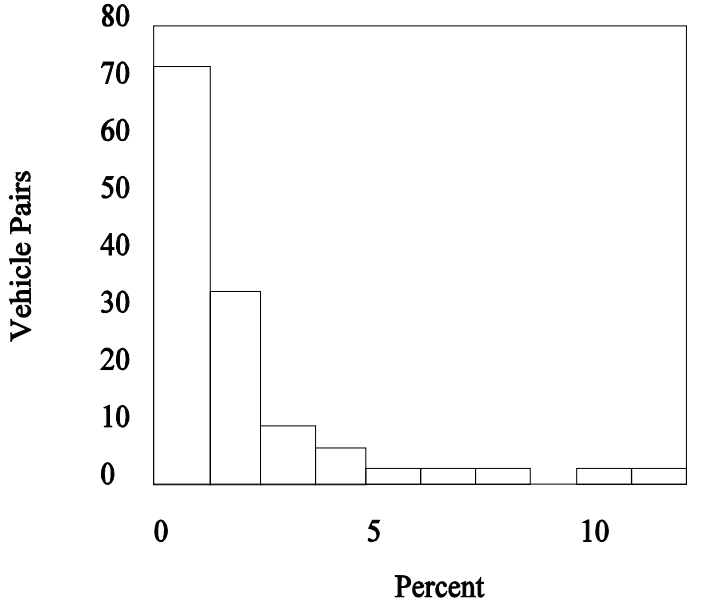
Lane number identification which can indicate the occurrence of a lane change from upstream to downstream is also an effective feature vector since the majority of the vehicles remained in the same lane over the 1.2 mile stretch of freeway in the SR-24 dataset. This was the case in both the moderate flow and the congested flow cases. A lane change is easily noted since both the upstream and downstream lane number identifications were recorded with each signature.



4.2. a) Speed Difference

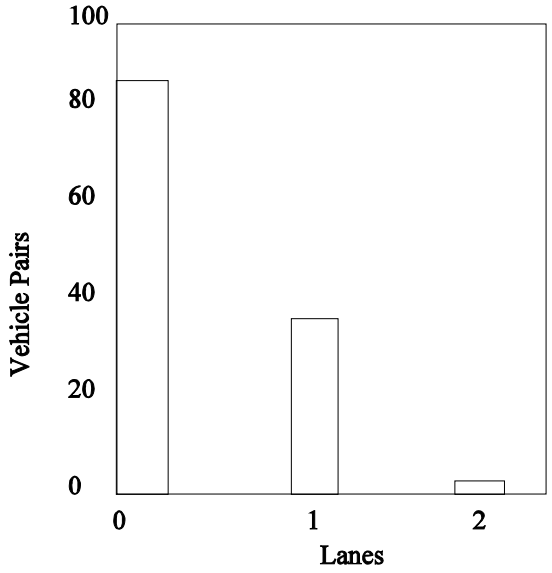


4.2. b) Length Difference



4.2. c)

Magnitude Difference



4.2. d) Number of Lane Changed

Figure 4.2. Traffic Characteristics of Matched Vehicle Pairs

Figure 4.2 is a set of four histograms that resulted from the matched upstream and downstream vehicle pairs for the moderate flow SR-24 dataset. The figure shows the inertia of traffic from upstream to downstream in this test section of 1.2 miles (1.92 km). It shows that traffic tends to stay in the same lane, maintain the same speed, and the measured vehicle characteristics of length and peak magnitude should be somewhat invariant.

4.2 CLASSIFICATION

A general transportation network can be represented by a collection of individual links where each link has only one entrance and one exit. The following are some examples of network links:

a section of multi-lane freeway where there are no on or off ramps.

a section of an arterial between two intersections

a freeway on-ramp

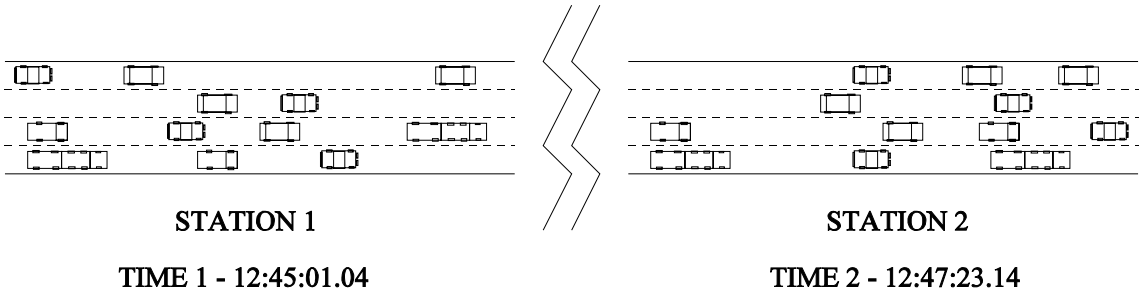


Figure 4.3. Sample Traffic Movement on a Link

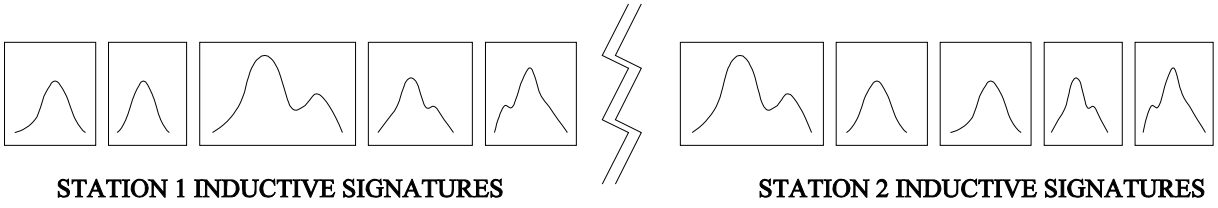


Figure 4.4. Sample Surveillance System Output

If a link is fully instrumented, then vehicle waveforms can be obtained from the beginning (upstream) and end (downstream) detectors. Figure 4.3 illustrates the sample traffic movement on a link, while Figure 4.4 shows the output of the inductive loop surveillance system. A vehicle waveform pair can then be formed using one downstream waveform and one upstream waveform. The vehicle reidentification problem is to find the matching upstream vehicle waveform from a set of candidate upstream vehicle waveforms given a downstream vehicle waveform. Instead of using upstream and downstream waveforms independently, one can form pairs of upstream and downstream waveforms. Another way of stating the vehicle reidentification problem is:

> given a set of vehicle waveform pairs $\boldsymbol{x}_i$, where $\boldsymbol{x}_i \quad S,\ i = 1, \ldots, N_c,$ and S
>
> is the set of vehicle waveform pairs, and N_c is the number of waveform pair combinations. Find the waveform pair which is produced by the same

vehicle.

Waveform pair $\boldsymbol{x_i}$ is actually the following matrix of feature vectors:

$$\boldsymbol{x_i} = \begin{matrix} x_{us1} \\ \begin{bmatrix} x_{us2} \\ \cdot \\ \cdot \\ \cdot \\ x_{usN_s} \\ x_{um} \\ x_{ul} \\ x_{uv} \\ x_{ud} \end{bmatrix} \\ x_{ut} \end{matrix} \quad \begin{matrix} x_{ds1} \\ \begin{bmatrix} x_{ds2} \\ \cdot \\ \cdot \\ \cdot \\ x_{dsN_s} \\ x_{dm} \\ x_{dl} \\ x_{dv} \\ x_{dd} \end{bmatrix} \\ x_{dt} \end{matrix}$$

The individual feature vectors are indicated with subscripts, s for shape, m for maximum magnitude, l for lane number identification, v for speed, d for vehicle length, and t for arrival time. The subscript u indicates upstream while d indicates downstream. The shape feature vector contains N_s number of spline-interpolated waveform ordinates.

A multi-objective optimization approach is used in the solution of the vehicle reidentification or classification problem. Note that even though "vehicle matching" is the overall objective, matching of each individual feature value pair in a pair of feature vectors can be considered an objective by itself. For instance "matching magnitude" is one objective and "matching speed" is another. A multi-objective approach combines these multiple objectives, the relative "importance" of each becoming significant. The "optimization" on any given objective here refers to minimizing the mismatches between feature vector pairs based on the given objective. Note that in this regard, the optimization based on any objective is equivalent to a matching based on the given feature as the "criterion". Thus the approach may also be called a "multi-criteria" approach. The multi-objective optimization in this case is performed over a matrix space instead of a vector space since the waveform pair $\boldsymbol{x_i}$ is actually a matrix.

Even though the optimization field is still dominated by single objective optimization approaches, the multi-objective approach merits consideration in many applications such as the vehicle reidentification problem. In many real world problems, more than one objective is often considered. For example in the field of transportation, a specific problem can involve minimizing several objectives such as financial cost, travel time, pollution, interjurisdictional conflict while maximizing the total throughput. In fact, the human decision making process seldom relies on a single objective but trades off between different factors in order to obtain an optimum. On the other hand, not every problem needs to be formulated as a multi-objective optimization problem. For example, the trivial case in which a point exists in the feasible region that simultaneously minimizes all objectives should be reformulated to a single objective optimization problem.

Mathematically, the multi-objective optimization problem can be written as:

$$\begin{aligned} \min \quad & f_1(x) = z_1 \\ \min \quad & f_2(x) = z_2 \\ & \vdots \\ \min \quad & f_m(x) = z_k \\ s.t. \quad & x \quad S \end{aligned}$$

where, f_1, ..., f_m are objective functions, z_1, ..., z_k are objective values, and x represent points in the feasible set S.

Qualitatively speaking, multi-objective optimization involves finding optimum points x^* within a feasible set x that are "as good as can be obtained" when judged according to multiple criteria (Rentmeesters, 1998). The notion of a Pareto set is used to describe optimum solutions for multi-objective problems. A pareto-set is also known as an "efficiency frontier", and the points in the set are called "efficient points". The Pareto set contains all points for which there does not exist any other point that would be uniformly better on all objectives. In other words, it is not possible to move from one efficient point feasibly to another so as to increase one of the objectives without necessarily decreasing at least one of the others (Steuer, 1986).

The lexicographic method is a sequential approach to solving the multi-objective optimization problem. In this approach, each objective is ordered according to its importance. In other words, the first level minimization is

$$\hat{f}_1 = \min [f_1(x) : x \quad S].$$

where, f_1 is the objective function, $\hat{f}_1$ is the optimum objective value, and x represent points in the feasible set S. The optimization yields the optimal decision set

$$S^1 = [x \quad S : f_1(x) = \hat{f}_1]$$

The next level minimization yields

$$\hat{f}_2 = \min [f_2(x) : x \quad S^1]$$

$$S^2 = [x \quad S^1 : f_2(x) = \hat{f}_2]$$

This process continues until all m objectives have been considered. Alternatively, a level in the lexicographic scheme could be a search-space reduction scheme using a criterion

rather than a pure minimization. The upper levels of the vector matching method in this research is incidentally of this type. It is interesting to note that in lexicographic optimization, the prior level optimization constrains the feasible set of the current level optimization. More importantly, each level ought to yield a set of multiple solution points for such a scheme to be implemented.

The existence of a hierarchical structure is at the core of the lexicographical method. Human decision making employs this structure when breaking down a complex problem into a series of simpler and related decisions (Lockett and Hetherington, 1985). The simple example of choosing

the type of manufacturing equipment in an electronics factory is shown in Figure 4.5. In this example, there are higher level priorities of safety and cost that must be considered first before the features of the equipment are even examined. The sequential decision making process is therefore represented in this structural approach.

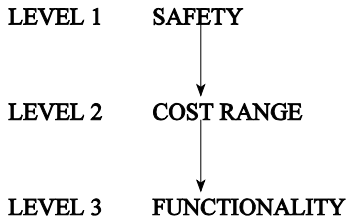


Figure 4.5. Simple Example of Hierarchy in Decision Making

Another approach to multiple-criteria techniques is to use any of several methods broadly classified as "weighted averaging schemes" where the different objectives are reconciled using "trade-offs" of "weights" between the objectives/criteria. One advantage that the lexicographical method has over weighted average schemes is the way the goals measured in different units can be placed at different priority levels. This avoids the problem of specifying relevant weight values when the objectives are measured in different units. For example, if the objectives are to maximize GNP measured in billions of dollars and to maximize the fraction of the population that is above the poverty level as measured by a number between zero and one, then equal weights of these two objective will actually render the second objective useless (Zionts, 1985).

Another advantage of using lexicographic versus simple multi-objective optimization is the characterization of the extremal points of the Pareto set. In a traditional weighted averaging solution to the multi-objective optimization problem, the optima is the result of the combination of all the objectives. This might lead to a compromised solution when the extremal points are actually needed to produce a better result.

Lexicographic optimization can be treated as "a form of multi-criteria optimization in which the

various objectives under consideration cannot be quantitatively traded off between each other" (Rentmeesters et al., 1996). Furthermore, under simple requirements of convexity of the objectives, every Pareto point of a vector valued objective function is the lexicographic minimum of some prioritized sequence of weighted averages of the original objectives under consideration. Therefore all Pareto optima in general can be satisfied by the basic lexicographic optimum characterization.

One concept that is used within the proposed optimization framework is lexicographic goal programming. Goal programming in its broadest definition refers to the establishment of a level of achievement for each criterion. The well known goal programming methods deal with setting target or "goal" values on each objective and optimizing a single objective which is the sum of deviations from the goals for all objectives. In the broadest definition of goal programming, however, a goal or set of targets can be used to form the constraint set. In the vehicle reidentification problem presented here, since certain target areas readily appear in the feature vector pair space, goals are introduced to reduce the search space to those areas. Since goals for such search space reduction are introduced in a hierarchical order with an explicit ordering among the features in the feature vector pairs, this may be called "lexicographic" goal programming. The search space reduction scheme presented here is also necessitated by the lack of a functional form for the objectives which will be discussed in the next section. Figure 4.6 illustrates a case in which a range of values (x_1 to x_2) is specified as a goal for a given objective, in this case a utility function U.

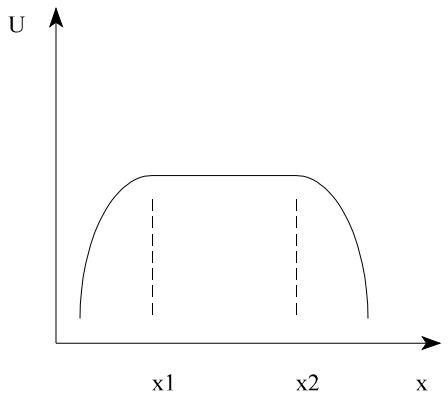


Figure 4.6. Goal Programming Utility Example

Lexicographic optimization is commonly used when the real cost function of an optimization problem is either unknown or too complex to describe. This is the case of the vehicle reidentification problem where the functional form of the real cost is not known, but the values of these functions in terms of differences in feature values for any pair of feature vectors is known.

The vehicle reidentification problem is formulated as a lexicographical optimization problem with five levels. The first three levels are implemented as goal programs having a range of values as solutions. These three levels can also be reformulated as constraints in a two level lexicographical optimization using the last two levels as hierarchical objectives, but the proposed formulation is more straightforward. The fourth level objective is a multi-objective optimization that is solved with the weighted averaging approach. The last level has a stochastic objective that is solved by using Bayesian analysis. It is recommended for the reader to first glance at the illustrative example at the end of the chapter before reading the problem formulation.

The first three levels of the optimization problem reflect the real-time requirements of the algorithm. These levels also result from the assumption that flow conditions are normal and no incidents have occurred. In other words, an upstream vehicle cannot have near infinite travel time to arrive downstream. Incident detection through the use of vehicle reidentification is a logical application, but it is not the topic of the current research.

The first level objective is the following goal program:

$$goal\ (\ f_1(x) = t(x) = z_1\ )\ such\ that\ (\ z_1 \quad [\ L_t,\ U_t\ ]\ ),\ x \quad S$$

$$S^1 = [\ x \quad S : f_1(x) = z_1\ ]$$

where, x represents a vehicle pair vector as defined in the beginning of Section 4.2, $t(x)$ is the measured time difference between the downstream and the upstream vehicle recording times at their

respective loop stations, L_t and U_t are the lower and upper bound for feasible travel time, and S is the feasible set of vehicle pairs which can be very large in practice. Note that rather than a "max" or "min", a "goal" is used in the notation above, since the optimization here is a goal-based search space reduction.

The first level optimization constrains the set of vehicle pairs composed of feasible upstream candidates to a probable set of vehicle pairs thus reducing the computational requirements of the algorithm. The first objective says that the travel times of vehicles cannot be either infinite or zero and are bounded by a lower and upper limit (i.e., a time window). Care must be taken in the determination of this time window because of the trade offs between computational efficiency and accuracy. If this time window is too wide, then the constraint set is enlarged so more computations are required and the possibility for misclassification is increased. If the time window is too small, then the matching vehicle waveform might be excluded from the constraint set, and once again misclassification can result. Reasonable time windows can either be determined a priori or set dynamically. Some methods for determining dynamic time windows include the use of current local speeds, the use of historical travel times, or the use of recently computed maximum and minimum travel times. Since the algorithm has only been used off-line, historical travel times were used in this determination for the results presented in Chapter 4.

The second objective is the following goal program that selects the pairs of vehicles that contain a candidate upstream vehicle magnitude that is within a certain percentage of the downstream vehicle that is being reidentified.

$$goal\ (\ f_2(x) = m(x) = z_2\ )\ such\ that\quad (\ z_2 < T_m\ ),\ x \quad S^1$$

$$S^2 = [x \quad S^1 : f_2(x) = z_2\ ]$$

where, $m(x)$ is percent magnitude difference between the upstream and the downstream vehicle waveforms, T_m is the percentage tolerance for inductance magnitude, and S^1 is the feasible constraint set that resulted from the first objective.

This optimization effectively limits the candidate vehicles to be from the same vehicle class as the downstream vehicle. This results from the fact that the inductance magnitude is inversely proportional to the height of the vehicle. For example, trucks would have much smaller magnitudes than passenger vehicles, so the candidate set is limited to trucks by using this constraint.

The last goal program is the following objective that limits the candidate vehicle's length to be within a certain percentage of the downstream vehicle.

$$goal\ (\ f_3(x) = l(x) = z_3\ )\ such\ that\quad (\ z_3 < T_l\ ),\ x \quad S^2$$

$$S^3 = [x \quad S^2 : f_3(x) = z_3\ ]$$

where, $l(x)$ is the percent difference of the electronic lengths between the downstream and upstream vehicles, T_l is the percentage tolerance for the electronic length, and S^2 is the feasible set derived from second objective. The magnitude tolerance used in the second objective and the length tolerances used in the third objective can be set by using historical data.

The fourth level multi-objective optimization is

$$\min f_{4,\omega_i} = \min (f_{a,\omega_i}(x),\ f_b(x),\ f_c(x),\ f_d(x)), \qquad i = 1,2,...,N$$
$$s.t.\quad x \quad S^3$$

$$f_{a,\omega_i}(x) = distance\ measure$$

$$f_b(x) = \lambda\, m(x)$$

$$f_c(x) = \lambda\, l(x)$$

$$f_d(x) = \lambda\, v(x)$$

$$x \quad S^3$$

where, ω_i indicates a particular distance measure, N is total number of distance measures used, $\lambda m(x)$ is the change in inductance magnitude between upstream and downstream, $l(x)$ is the lane change dummy variable, and $\lambda v(x)$ is the change in speed between upstream and downstream. Note that each distinct distance measure produces a different discriminant function composed of the objective functions $f_{a,\omega i}$, f_b, f_c, and f_d. The different discriminant functions used during this optimization produces different optimum points. The collection of the set of optimum points produces the feasible set for the next level optimization.

Some candidate distance measures defining $f_{a,\omega i}(x)$ are

$$\textit{Euclidean} \quad f_{a,\omega_1}(x) = \sum_{k=1}^{N_s} \quad x_{dk} \quad x_{uk}$$

$$\textit{Correlation} \quad f_{a,\omega_2}(x) = 1 \quad \frac{\sum_{k=1}^{N_s} (x_{dk} \quad \bar{x}_d)\,(x_{uk} \quad \bar{x}_u)}{\sum_{k=1}^{N_s} (x_{dk} \quad \bar{x}_d)^2 \sum_{k=1}^{N_s} (x_{uk} \quad \bar{x}_u)^2}$$

$$\textit{Similarity} \quad f_{a,\omega_3}(x) = 1 \quad \frac{x_d \quad x_u}{x_d \quad x_d \quad x_u \quad x_u \quad x_d \quad x_u}$$

where u and d are the indices for the upstream and downstream vehicle waveforms, and k is the index for a component of the waveform.

$$\textit{Lebesgue} \quad f_{a,\omega_4} = \sum_{k=1}^{N_1} [\ (y_{ik} \quad y_{jk}) \quad \lambda\bar{y}\ ]^2 \qquad \sum_{k=N_1 \quad 1}^{N_2} [\ (y_{ik} \quad y_{jk}) \quad \lambda\bar{y}\ ]^2 \qquad \sum_{k=N_{p-1} \quad 1}^{N_p} [\ (y_{ik} \quad y_{jk}) \quad \lambda\bar{y}\ ]^2$$

where, y_i and y_k are y axis transformed values of the variable x and p is the total number of piece-wise continuous segments that form the waveform x.

$$\textit{Neural Network} \quad f_{a,\omega_5}(x) = 1 \quad \textit{output neuron}$$

where, output neuron is trained to measure the closeness of waveforms. During training, output neuron=1, if the vehicle waveforms are from the same vehicle, and output neuron=0, if the vehicle waveforms are from different vehicles. Note that for correlation, similarity and neural network, one is subtracted from the distance measure in order to minimize the objective.

The use of the fourth level optimization exploits the electronic characteristics of individual vehicles as well as the traffic characteristics of vehicles within freeway sections. The fourth level optimization is formulated with the following objectives:

The most important objective is to minimize the distance between processed waveforms which represents the inductance shape of vehicles. Thus, distance metrics measure the "closeness" between two processed waveforms. Temporal variability of the individual vehicles and spatial variability among different detectors is removed by the waveform processing. The five measures that have proved to be effective are: Euclidean, correlation, similarity, Lebesgue, and neural network. However, it is not always necessary to use all five measures if a subset of these prove to be particularly effective.

The magnitude difference between the prototype vehicle and the match candidate should be minimized, since the amplitude of the inductance represents the height of the vehicle body from the ground. However, due to variability in suspension height and the possibility that the vehicle is offset from the center of the loop, this objective is less unique than the processed waveforms.

Lane changes should be minimized as the inertia of a high percentage of vehicles is maintained. This frequency of lane changes is a function of the length of the segment, the number of ingress/egress points near the segment, the current traffic conditions, and individual driver behavior. The importance of this objective varies depending on the aforementioned factors. However, from empirical evidence (Chang and Kao, 1990) including the Lafayette freeway data, a high percentage of vehicles on freeways tend to stay in the same lane.

Speed difference between the prototype vehicle and candidate vehicles should be minimized. This objective is similar to the previous objective of lane change minimization as it also

expresses the conservation of the traffic inertia.

Even though the distance measures used as $f_{a,\omega i}$ are not completely independent (i.e., the information in these measures overlap), they define "closeness" in different ways. The Euclidean distance is a straightforward "city-block" measure of two different waveforms. The correlation measures the linear association between two waveforms. The similarity metric resembles a normalized vector dot product, thus it produces a measure analogous to the angular displacement of two waveform vectors.

One major difficulty encountered with the use of the processed waveform is the fact that waveform shift was present due to the detector sampling period and detector thresholding. Figure 4.7 shows the plots of the same vehicle upstream and downstream. In the case of a waveform shift, the distance measure between the upstream and downstream waveforms would result in significant errors despite the fact that it is derived from the waveforms originating from the same vehicle.

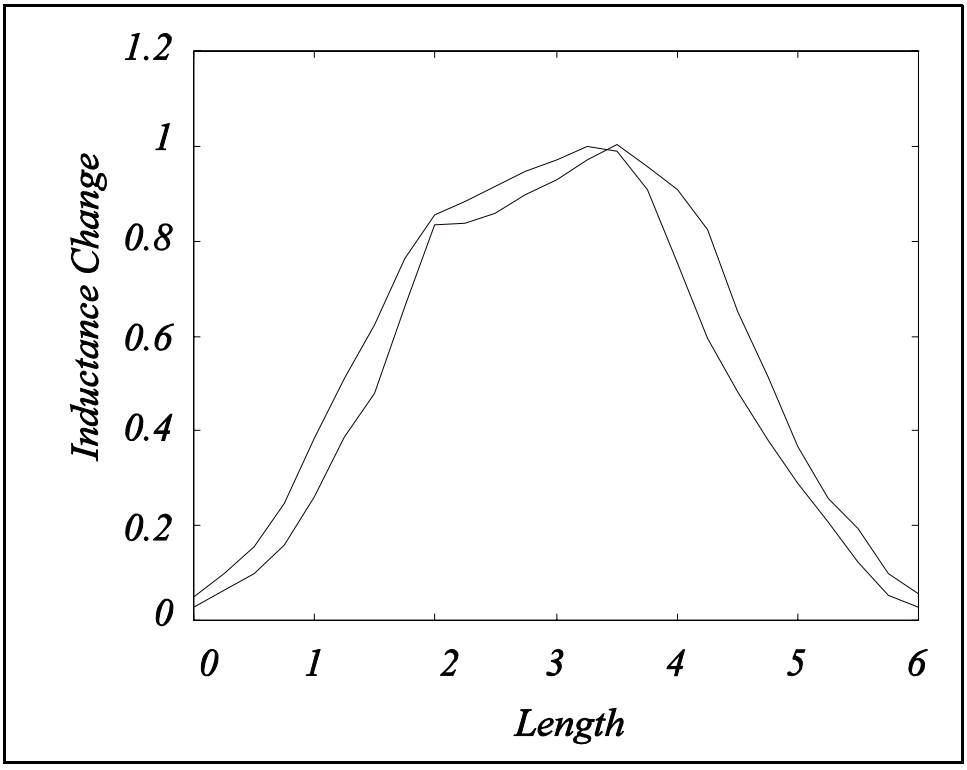


Figure 4.7. Example of Waveform Shift From Upstream to Downstream

One way of addressing the problem of waveform shifting is to use a technique of horizontal differencing motivated by the Lebesgue integral (Hawkins, 1970). For a function f defined on $[a,b]$, the Lebesgue integral partitions the range of function instead of the domain, $[a,b]$. Given that the greatest lower bound and the least upper bound of f on $[a,b]$ is m and M respectively, so $m=a_o<a_1<...<a_n=M$. Also consider the set e_i which partitions $[a,b]$**.** The Lebesgue integral then states that

$$\int_a^b f = \lim_{\|P\| \to 0} a_i m(e_i) = \lim_{\|P\| \to 0} a_{i-1} m(e_i)$$, where P is the maximum of the differences a_{i+1}-a_i.

The steps of the Lebesgue measure are as follows.

Step1. Separate the waveform into a minimum number of monotonic functions. This can be done in general by making the maxima the endpoints. So a given $f(x)$ on a domain $[a,b]$ is subdivided into piecewise functions $f_1(x)$ on $[a,a_1]$, $f_2(x)$ on $[a_1,a_2]$,..., $f_n(x)$ on $[a_n,b]$, and $f_1'(a_1)=f_2'(a_2)=...=f_{n-1}'(a_{n-1})=0$. Only the end curves, $f_1(x)$ and $f_n(x)$ were used in the actual algorithm and proved to be adequate for deriving a good representative of the waveform. One factor in using only $f_1(x)$ and $fn(x)$ is because many vehicles only have a single maximum in their waveform.

Step 2. Axis transformation is performed in which the ordinate and abscissa are switched. A resampling of the waveform along equal increments of the new ordinate is performed. This resampling is done with linear interpolation.

In order to use this transformed waveform, the variance of the function differences between the two waveforms is measured. This transformation removes the effect of waveform shift since waveform differences between waveforms of the same vehicle will show small variances even if they are shifted versions of one another.

The last candidate distance measure is the Neural Network classifier. The multi-layer feed-forward

(MLF) neural network was chosen for its nonlinear classification capabilities and for its well-developed formulation. (The reader is referred to Wasserman 1989 and Lippman 1987 for more details on the MLF.) Figure 4.8 illustrates the architecture of the MLF. The MLF performs like a traditional classifier and is trained on the test data set. The back propagation algorithm is used to train the MLF. The back propagation uses a gradient search technique to minimize the mean square difference between the desired and the actual net outputs.

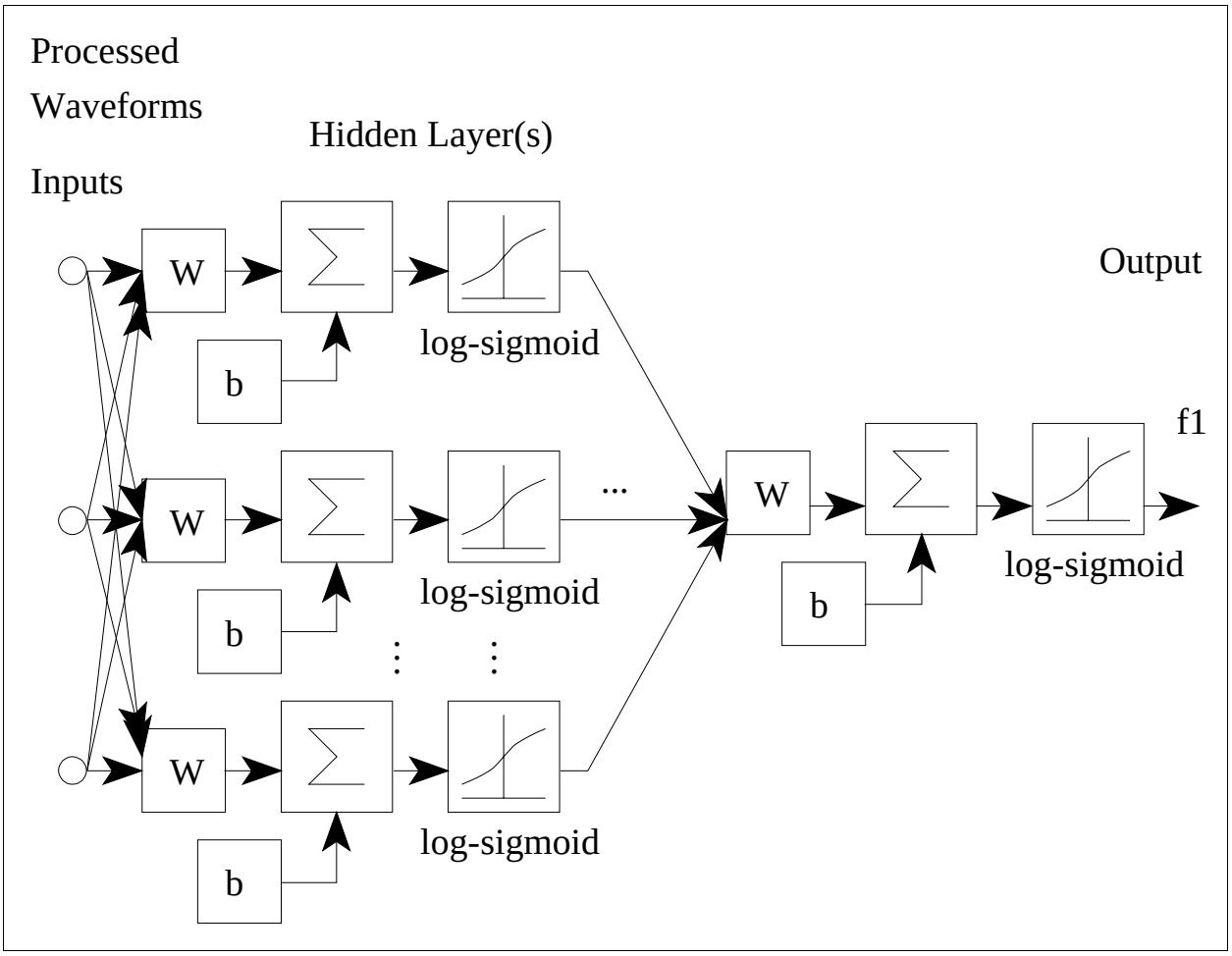


Figure 4.8. Multilayer Feed-Forward Network Architecture
Used as Distance Metric

The last level objective is expressed as follows:

$$\max_{x} f_5(x) = \max_{S^4} p\left(f_{4,\omega_i}(x)/\omega_i \right) p(\omega_i), \quad i = 1, 2, ..., N$$

where, $f_{4,\omega i}(x)$ is the minimum objective value obtained from distance measure ω_i from the previous level optimization, $p(f_{4,\omega i}(x)/\omega_i)$ is the conditional probability function of distance measure ω_i, and $p(\omega_i)$ is the a priori probability of distance measure ω_i. This objective then tries to optimize the distance measure with the maximum posteriori probability. Equivalently, this results in the choosing of the "most likely" distance measure that would identify the correct match between a vehicle pair. The selection of the optimum objective function is therefore done by Bayesian analysis.

This last objective is useful in a general formulation of the vehicle reidentification problem because it serves as an indicator of the current traffic conditions. When adequate historical data are available, probability functions for the different distance measures can be obtained. These different distance measures perform differently under different traffic conditions. For example in low flow conditions, vehicles travel faster and the total number of waveform samples is small. In this instance, the sampling error is much more significant since the waveform shift due to the sampling period is a much greater percentage of the total number of samples. The Lebesgue distance measure would then be a better distance measure for dealing with the waveform shift problem. In other traffic flow conditions, another distance measure might prove to be better.

Another benefit of using Bayesian analysis in the last objective is to address the stochastic nature of the vehicle reidentification problem. The randomness of the vehicle waveforms is introduced by both input and system errors. The input errors include differences in the vehicle entrance angle into the inductive field, varying vehicle offsets from the loop center, and different instantaneous suspension heights. The system errors can include inaccuracies in the speed measurement, errors in interpolation, aliasing errors, quantization errors, and loop station calibration differences. As a result of these errors, the feasible set of waveform pairs contain disturbances. The Bayesian approach compensates for these errors by utilizing historical data in the form of discriminant

probability functions.

The solution of the lexicographic optimization problem is straightforward since the vehicle reidentification problem is computationally tractable. The first goal program involves a sequential search of vehicle pairs (or equivalently the candidate upstream vehicles). This is simple computationally since the upstream vehicles are recorded sequentially in time. This critical optimization reduces a very large feasible set into one that can approximately range between twenty to sixty vehicle pairs.

The second and third goal programs involve the selection of vehicles that are within the magnitude and the length tolerances. Since the feasible set is now a manageable size, the percent magnitude difference and the percent length difference can be easily computed for the entire feasible set derived from the first optimization.

The fourth level multi-objective optimization problem is solved by forming a utility function from the various objective functions. The advantage of the utility function is that a multiple-objective problem is reduced to a single objective function problem for which many methods exist for solution (Kapur 1970). So the new fourth level problem becomes

$$\begin{array}{c} \textit{minimize } h\ (\ f_a(x),\ f_b(x),\ f_c(x),\ f_d(x)\ ) \\ \textit{s.t. } x \quad S \end{array}$$

By formulating h as a linear function using weighted averaging, the problem is reduced to the following simple optimization problem

$$\textit{minimize} \sum_{k=a}^{d} \lambda_k\ f_k(x)$$

$$\textit{s.t. } x \quad S,\ \lambda \quad \lambda$$

$$\lambda \ = (\ \lambda \quad \sum_{k=a}^{d} \lambda_k = 1,\ \lambda_k \quad 0,\ k = a,b,c,d\ )$$

Thus, λ represents the set of weights associated with the various objective functions. In order to attain feasible solutions, the weights are constrained to be finite and non-negative. In the current off-line case, the weights are determined a priori with historical data. However, there are no restrictions for this process to be done dynamically or at set time periods when the algorithm becomes implemented in real-time. The sequential determination of weights is done in the following manner:

$$i.\quad \textit{minimize}\ \lambda_a\ f_a(x) \quad \lambda_b\ f_b(x)$$
$$s.t.\ x \quad S^h,\ \lambda_a \quad \lambda_b = 1$$

$$ii.\quad \textit{minimize}\ \lambda_a\ f_a(x) \quad \lambda_b\ f_b(x) \quad \lambda_c\ f_c(x)$$
$$s.t.\ x \quad S^h,\ \lambda_a/\lambda_b = \lambda_a/\lambda_b,\ \lambda_a \quad \lambda_b \quad \lambda_c = 1$$

$$iii.\quad \textit{minimize}\ \lambda_a\ f_a(x) \quad \lambda_b\ f_b(x) \quad \lambda_c\ f_c(x) \quad \lambda_d\ f_d(x)$$
$$s.t.\ x \quad S^h,\ \lambda_b/\lambda_c = \lambda_b/\lambda_c,\ \lambda_a \quad \lambda_b \quad \lambda_c \quad \lambda_d = 1$$

At the first stage, the ideal weight ratio of the first two objective functions is found by a grid search. This ratio is then fixed for the second stage where a third objective function is added. Again, a grid search is used to minimize the linearized objective function and the weights ratios of all three objectives are set. At the last stage, the previous procedure is applied by keeping the previously optimized weight ratios. Once the weights are determined in the utility function, the computation of the utility function and the subsequent minimization is straightforward.

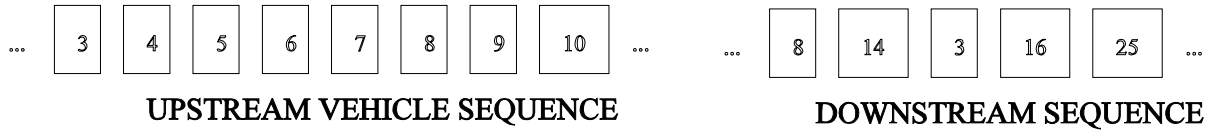
In order to choose the most appropriate distance measure for the fourth level multi-objective optimization, the fifth level optimization is performed. Since ground truthing is available for the entire data set, a priori classification probabilities can be obtained for each of the distance measures. The entire data set then is divided into a training (or historical) set and a testing set. The training set is also used for deriving the optimum weights in the fourth level multi-criteria optimization.

In Bayesian analysis, a priori information for each distance measure is estimated using the historical

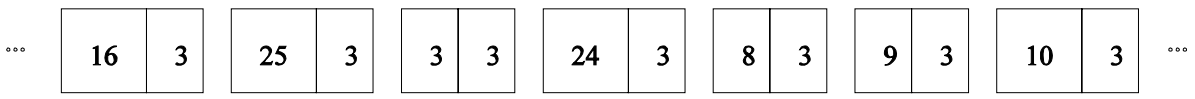
data set. A symmetric loss function is used since the cost of misidentifying any distance measure is the same. A Bayes decision function with symmetric loss then becomes a maximum likelihood decision.

A small illustrative example is presented to detail the process of the lexicographic optimization approach. Assume a physical link similar to the one where real world data was collected on a four lane mainline freeway in SR-24 in Lafayette, California. Such a link is composed of an entrance point (upstream) and an exit point (downstream) instrumented with surveillance equipment, and no intermediate ingress or egress locations exist. In this scenario, a sequence of vehicles is observed by the surveillance system both upstream and downstream. Note that when a vehicle number is mentioned, each vehicle's inductive signature, speed, lane information, time stamp, and other derived feature vectors are all contained in the vehicle record.

Therefore given a sample sequence of upstream and downstream vehicles:



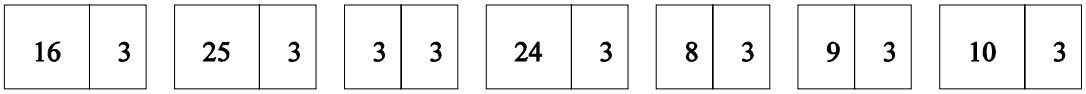
If the upstream sequence vehicles were numbered sequentially, then the downstream sequence can be in a rearranged order due to the difference in lane speeds, overtaking, or lane changing. A sample vehicle reidentification problem is stated as follows: "Given the particular downstream vehicle number 3, find the most likely upstream vehicle." The observed vehicle sequences can be written as the following set of vehicle pairs **x** for input into the lexicographic optimization algorithm.



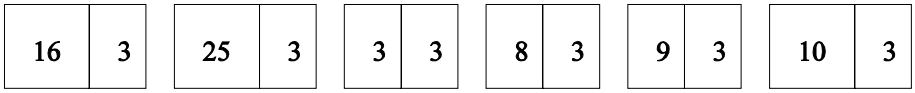
This large set of vehicle pairs is composed of individual vehicle pairs x_i.

The five levels in the lexicographic optimization algorithm would proceed by reducing the feasible set, as the previous level optimization is met.

1. Travel time optimization will reduce a large set of feasible vehicle pairs to a set of seven vehicle pairs that meets the travel time optimization criterion. Note that in the case of a real world dataset from SR-24, the optimum set is much larger and on the order of twenty to fifty vehicles.



2. For this example, assume every vehicle is a passenger vehicle except for vehicle 24 which is a truck. Magnitude optimization will eliminate vehicle 24 by further reducing the set of seven to six vehicles.



3. Now, assuming that vehicle 25 is a limousine, then its associated vehicle pair will be eliminated from the length optimum set. The result will be a set of five vehicle pairs.

4. At this optimization level, different discriminant functions will result in a different optimum point. The following optimum set can result from the evaluation of five different discriminant functions.

3	3

EUCLIDEAN

3	3

LEBESQUE

8	3

CORRELATION

8	3

SIMILARITY

3	3

NEURAL NETWORK

5. The final optimization level involves the use of Bayesian discriminant functions. If the highest posteriori probability results from the use of Lebesgue discriminant, then the optimum point will be the vehicle pair (3,3).

3	3

LEBESQUE

CHAPTER 5. RESULTS

5.1 SECTION MEASURES RESULTS

The solution of a single vehicle reidentification problem using the lexicographical optimization approach fixes the vehicle arrival times at the upstream and the downstream surveillance stations. When this problem is solved repeatedly for a group of vehicles, section measures such as travel time and density can be derived in a straightforward fashion.

Figure 5.1 is a flowchart of the travel time computation procedures for the true travel times, estimated travel times using reidentified vehicles, and local travel times using point measures. The true travel times and the local travel times are used for comparison with the estimated results from the lexicographical optimization method. The true travel times are derived from the manually correlated video database. The local travel times are derived using the upstream speed trap speed and extrapolated to the entire section.

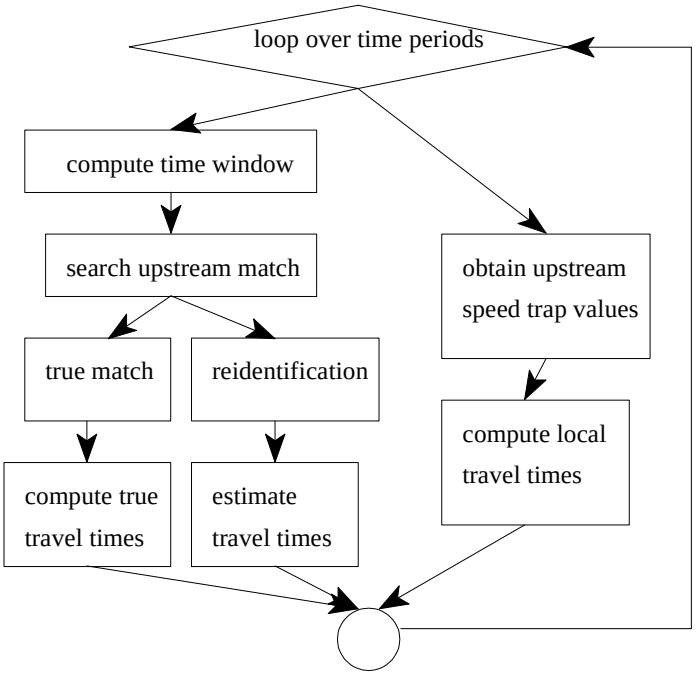


Figure 5.1. Travel Time Computation Pseudo-Flowchart

Figure 5.2 is a pseudo-flowchart of the density computation procedures for the true section densities, estimated section densities using reidentified vehicles, and local densities using point measures. The true section densities and the local densities are used for comparison with the estimated results from the lexicographical optimization method. The true section densities are derived from the manually correlated video database. The local travel times are derived by using the upstream speed trap speeds and the upstream flows.

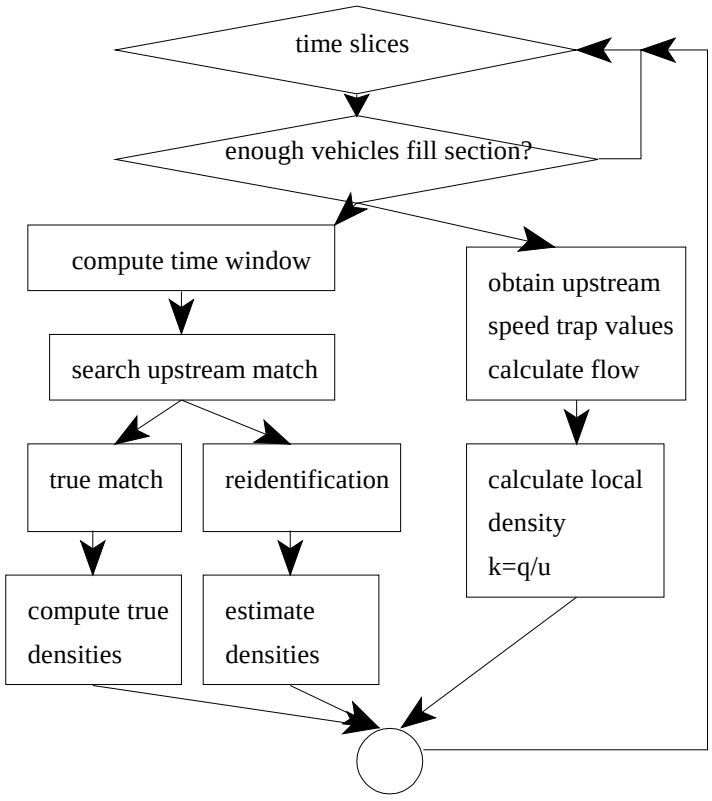


Figure 5.2. Section Density Computation Pseudo-Flowchart

The results of this approach based on data from the SR-24 freeway segment show that the use of vehicle reidentification performed much better than conventional methods that use point measures. Video ground truthing was used to measure true section measures.

Figures 5.3 and 5.4 show the 15 second average travel time comparisons for the moderate flow case. The moderate flow case involved flows of approximately 1000 vphpl. In both the case of the training data and the test data, the estimated travel times followed the true travel times closely. The moderate flow case is a less challenging case, as most traffic maintained the same speed from upstream to downstream in the absence of congestion. In this case, one would expect the extrapolation of point speeds to estimate travel times well for the section. The average percent error of the training results was 2.31%, while the average percent error of the test results was 1.88%. This shows that the training data set was adequate in the estimation of the conditional and discriminant probabilities used in the lexicographic optimization. The fact that the test data actually performed better than the training data shows that once the algorithm is calibrated, it can perform well under the same traffic conditions. The standard deviation of the test data of 2% shows that individual fifteen second travel times are being estimated accurately in addition to the accurate performance over the entire fifteen minute period. The entire moderate flow travel time results are tabulated in Appendix D.

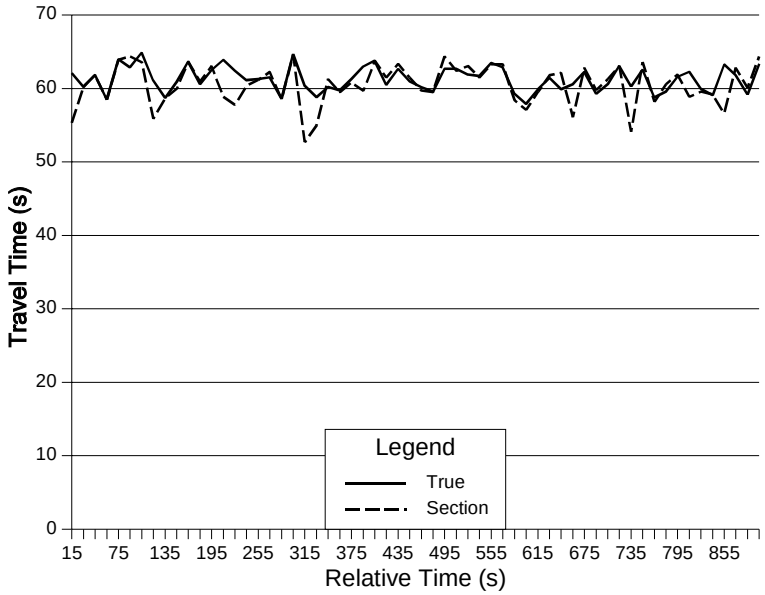


Figure 5.3. Comparison of Average Travel Times for Moderate Flow Training Data

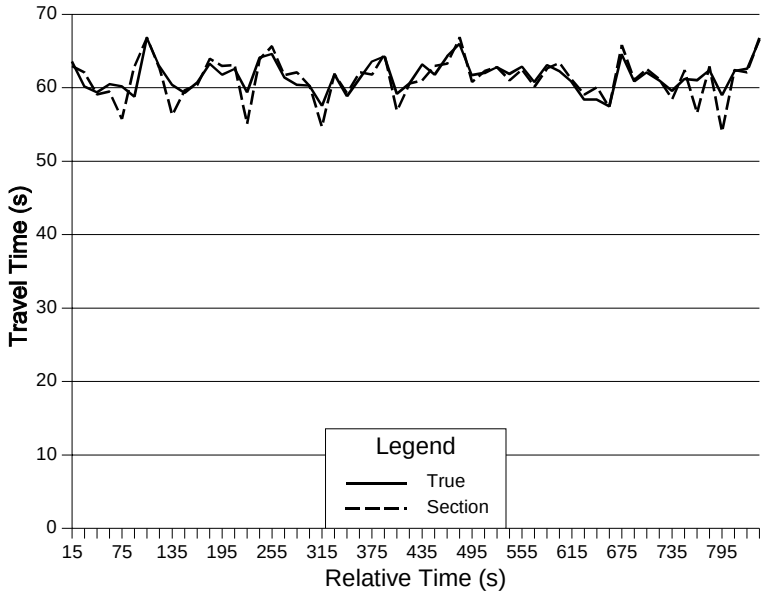


Figure 5.4. Comparison of Average Travel Times for Moderate Flow Test Data

Figure 5.5 shows a comparison of the estimated section travel time (section) results with the true travel times (true) and the locally estimated travel times (local) for congested data. The congested data averaged around 1800 vphpl. These data are more interesting than the moderate flow because congestion causes traffic to deviate from a constant speed. The congested case shows the inadequacy of using local or point measures for estimating section travel times. The figure shows that the point travel times or point speeds fluctuate significantly and should not be extrapolated over the entire section. On the other hand, the estimated section travel times once again followed the true travel times closely. Due to the more challenging traffic conditions the results had an average error of 3.24% which is just slightly higher than the moderate case results. The standard deviation of the error was 3.034%. The estimated local travel times has a significant average error of 45.82% with a standard deviation of 31.53%.

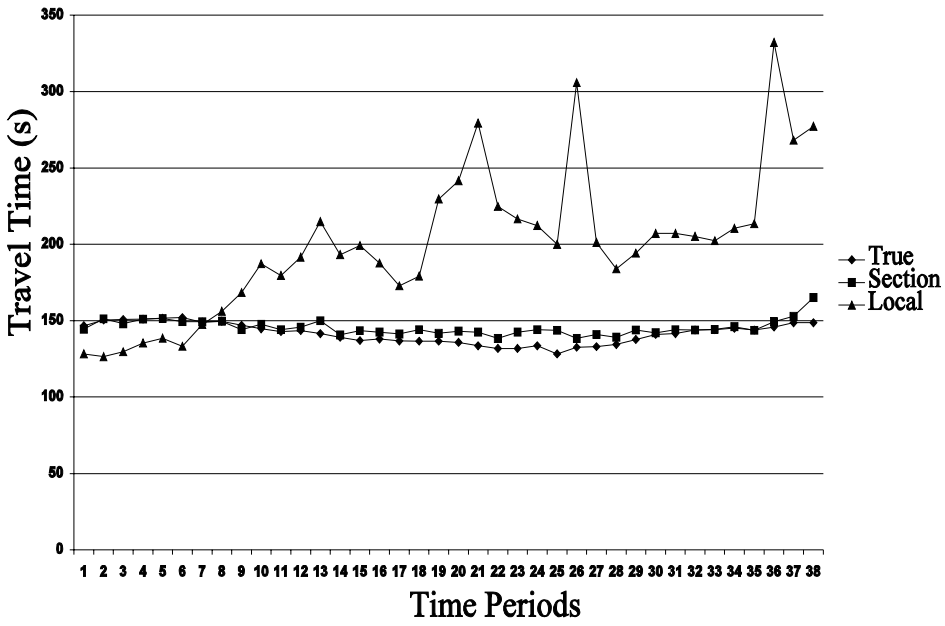


Figure 5.5. Comparison of Average Travel Times for Congested Flow Data

Instead of using all the reidentified vehicles, a subset of vehicles with higher match probability can be used for the derivation of section travel times. The true travel times used for comparison are still derived from the complete data set. Because discriminant probabilities are computed as part of the lexicographic optimization, a discriminant threshold can be used to retain only "highly probable" vehicle matches. The use of a subset of vehicles results in lower average error as shown in Table 5.1. The subset results show that the matching percentage increases with a lower discriminant threshold (higher discriminant probability), and the resulting average travel time error generally decreases with an increase in matching percentage. The size of the subset which is a function of the discriminant threshold should be chosen carefully since the variance might increase with a smaller match set. With a smaller subset there is also a danger of the vehicles being biased toward trucks which are slower but possess more unique signatures.

Table 5.1 Average Travel Times Computed from Subsets Using Moderate Flow Data

Discriminant Threshold	Additional Constraints	Average % Travel Time Error	σ % Travel Time Error	Size of Subset (veh)	% Total Vehicles (918)	% Match
0.1	none	3.01	2.83	490	53.4%	NA
0.07	none	2.62	2.89	262	28.5%	73.7%
0.07	>3 veh's	2.14	2.2			
0.06	none	2.17	3.09	191	20.8%	74.9%
0.06	>2 veh's	1.86	2.47			
0.05	none	2.30	4.60	127	13.8%	NA
0.04	none	1.69	3.66	56	6.1%	82%

Figures 5.6 and 5.7 show the 15 second average section densities for the moderate flow training and test data. The results show that the estimated section densities were tracking the true densities closely. The average percent errors for the training and test data are 1.25% and 1.30% respectively. The standard deviations for the errors were 1.53% and 1.18% respectively. The use of the training and the test data both resulted in similar performance. In contrast, the local or occupancy-based densities are clearly not able to follow the section densities. The average error was 31.1% and 33.5% for the training and test cases respectively. The standard deviations of the errors were 25.3% and 24.2%, respectively, which shows that the point densities are a poor surrogate for section densities.

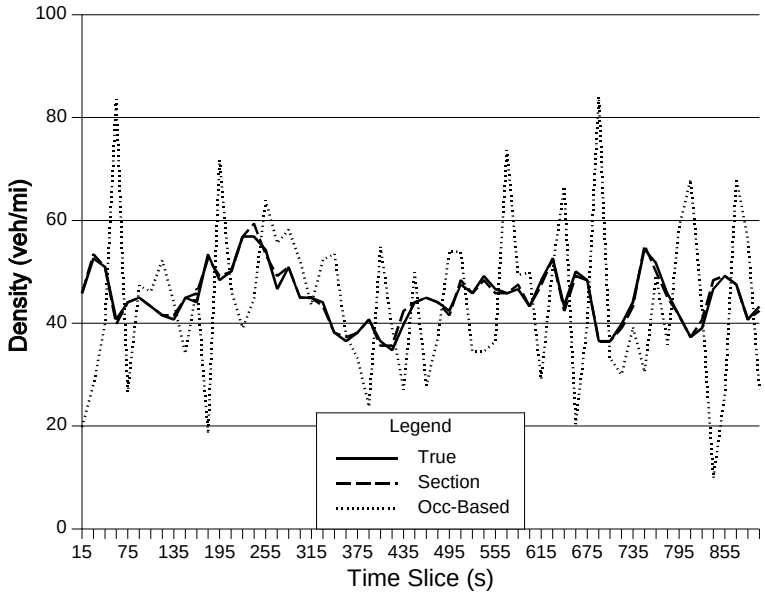


Figure 5.6. Comparison of Section Densities for Moderate Flow Training Data

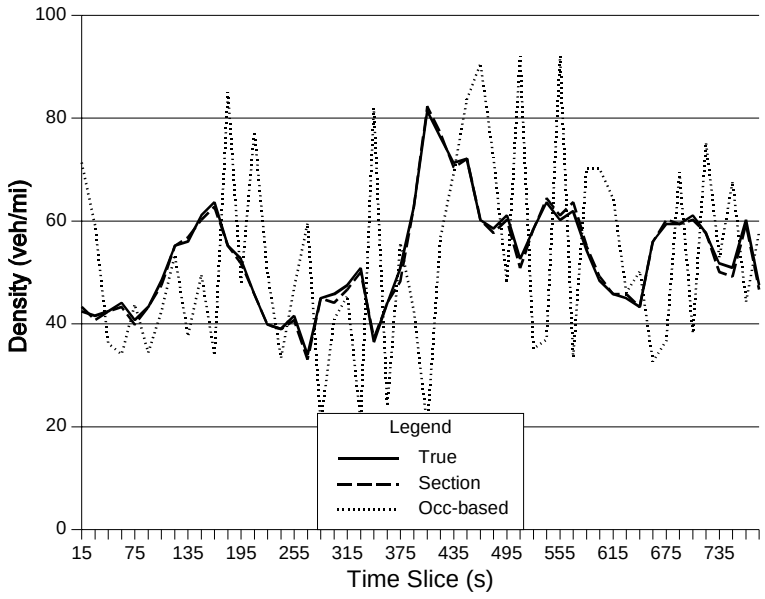


Figure 5.7. Comparison of Section Densities for Moderate Flow Test Data

Figure 5.8 shows a comparison of the estimated section densities (section) with the true densities (true) and the occupancy-based densities (occ) for congested flow data. The estimated section densities has an average error of 3.64% and a standard deviation of 2.15%. The error is slightly higher than the moderate case and is due to a lower vehicle reidentification rate. The congestion causes more variability in the traffic stream and translates into more mismatches. The point or occupancy-based densities showed an even worse performance than in the moderate flow case. The average error of 46.52% and standard deviation of 24.66% shows that the variability of the occupancy-based density is great while the accuracy is also poor.

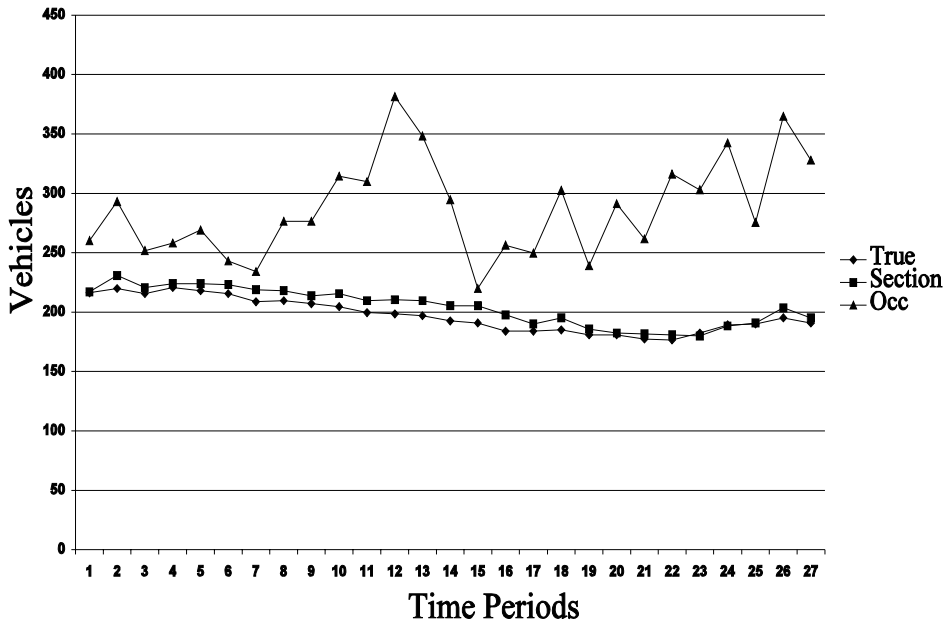


Figure 5.8. Comparison of Section Densities for Congested Flow Data

5.2 OTHER RESULTS

Sensitivity analysis was performed to see the effect of speed accuracy on the reidentification percentage. For this analysis, normally distributed disturbances were introduced into the vehicle speeds and the reidentification was recomputed using the erroneous speeds. One reason for performing this analysis is to assess the feasibility of using single loop stations in the vehicle

reidentification framework. This analysis is done using a smaller moderate flow dataset which had an average speed of around 62 miles per hour. Table 5.2 shows that the reidentification percentage decreases gradually with increased speed errors. Even with a normally distributed speed error variance of 5 $(mph)^2$, the reidentification system is still able to match over 50% of the vehicles.

Table 5.2. Speed Variance vs. Reidentification %

SPEED VARIANCE $(m/s)^2$	SPEED VARIANCE $(mph)^2$	MATCH%
0	0	74%
0.222	0.5	73%
0.444	1	69%
1.333	3	68%
1.777	4	65%
2.222	5	55%
4.444	10	41%
6.666	15	33%

An investigation using limited data was made on the estimation of speeds using single loops. The conventional approach necessitates the assumption of an average vehicle length and uses the turn on and turn off times of the loop detector. This method can lead to large errors if the variance of the vehicle lengths is great. The new method uses the vehicle signature and approximates speeds by using the waveform slew rate. The slew rate measures how fast the mass of the vehicle covers the entire loop, and is proportional to the speed. Since the slew rate is measured in terms of inductance over time, calibration is needed to translate the units to length over time. Polynomial curve fitting was used to estimate a curve that fitted the calibrated values of vehicle speeds against slew rate slopes. Table 5.3 shows the results of a study using a passenger vehicle for calibration and a sport utility vehicle for testing. The encouraging initial results show that this methodology should be further investigated once more data become available.

Table 5.3. Sample Speed Estimation Using Single Loops

ACTUAL SPD. (MPH)	ESTIMATED SPD. (MPH)	% ERROR
15	14.34	4.4
21	21	0
25	25.67	2.7
30	31.67	5.6

Table 5.4 shows the vehicle match percentages in terms of vehicle types. This analysis is only done for a small sample of the dataset because of the time consuming nature of manual classification using video. This analysis is necessary since the traffic characteristics between the different types of vehicles are different. For example, since semi-trailer trucks are restricted to the two slowest lanes, and since they tend to travel slower than the rest of the traffic, a high semi-trailer truck match percentage could bias the computed section travel times to be slower than the actual value. However, from the test results of actual data, this bias has not been found to be significant since semi-trailer trucks accounted for less than 6% of the vehicle traffic. The results listed in Table 5.4 are intuitive, because non-passenger vehicles tend to be longer than passenger vehicles and have more distinguishable features which leads to higher vehicle match percentages.

Table 5.4. Vehicle Reidentification Percentages by Type of Vehicle

Car	Station Wagon	Sports Utility	Pickup	Trucks	Van	Semi-Trailer
75%	75%	100%	80%	61%	92%	100%
Passenger Vehicles (49% of total number of vehicles)		Non-Passenger Vehicles (51% of total number of vehicles)				
75%		78%				

The match percentages are very significant because of the following two constraints. First, the physical loops were the standard Caltrans (California Department of Transportation) six-by-six foot

(1.82mx1.82m) loops. Because of their dimensions, these loops integrate the waveform signal over the six foot distance and essentially eliminate more distinctive edges. By contrast, loops one meter (3.3 foot) long in the traversal direction are used in Europe, and they are able to produce sharper curves in the waveforms. The second constraint is the sampling period of 13 to 14 ms. With this low sampling rate, sharp corners in the waveforms are also missed. In addition, the beginning of each of the upstream and downstream waveforms can be offset by this sampling period. Thus the upstream and downstream waveforms could be offset by as much as 28ms which produces a significant error when computing the distances between upstream and downstream waveforms.

CHAPTER 6. APPLICATIONS

6.1 FRAMEWORK FOR REAL-TIME ROAD TRAFFIC CONGESTION DETECTION

6.1.1 Introduction

Automatic road traffic congestion detection is one of the most important components of a freeway traffic management system. Early detection and warning of road traffic congestion may decrease the traffic impact as well as occurrence of an event related to the congestion such as an incident or queue overflow. It was suggested that the rise of secondary accidents can be significantly reduced by earlier detection and warning (Lindley 1987).

Despite conventional approaches which have been proposed for decades, development of an advanced approach to automatic road traffic congestion detection remains an active area of research due to lack of the capability of providing intra-lane and inter-lane traffic information for the use in ATMS and ATIS. Queue length derived from space mean speed was utilized to identify congestion levels in urban road networks in a study conducted by Shibata et al. (1984). Using their approach, only three congestion levels were categorized. Iwasaki et al. (1985) proposed a method of distinguishing congested flow from non-congested flow to improve congestion detection in urban motorway traffic surveillance and control systems where the detection output using their approach was limited to binary information (i.e., congestion and non-congestion). The McMaster incident detection algorithm (Persaud et al., 1990) appeared promising in distinguishing between incident and non-incident congestion; however, further tests were suggested in their study for special cases such as incidents immediately downstream and upstream to detector stations, and detection of incidents during already-congested operation.

6.1.2 Proposed Framework

The following symbols are used in this section:

L = link length;

n = number of lanes in a given link;

$o_j(k)$ = occupancy measured in lane j in time step k;

$o_j^{dn}(k)$ = downstream occupancy measured in lane j in time step k;

$o_j^{up}(k)$ = upstream occupancy measured in lane j in time step k;

$\Delta o_j(k)$ = difference between upstream and downstream occupancy values measured in lane j in time step k;

$qj(k)$ = traffic count measured in lane j in time step k;

$T(k)$ = link travel time in time step k;

$wj(k)$ = density-weighting value associated with lane j in time step k.

In this section, a new framework which is founded on the basis of a nonlinear stochastic system modeling approach and the Modified Sequential Probability Ratio Tests technology (MSPRT) for real-time road traffic congestion detection is proposed. The proposed framework is shown in Figure 6.1.

Data used as the input shown in Figure 6.1 include lane-by-lane raw traffic data (traffic count and occupancy), and processed section-related data (link travel time). Lane traffic count and occupancy data can be readily collected from point detectors in each time step. Link travel time is defined as the time required for any vehicles to cross a given section between two sequential detector stations (i.e., upstream and downstream detector stations) on a freeway. It can be noted that link travel time is time-varying since traffic conditions within a given section are time-varying as well. In this study, link travel time should be given in each time step. The vehicle reidentification algorithms described in the previous chapters is being considered for generating time-varying link travel time in our further research.

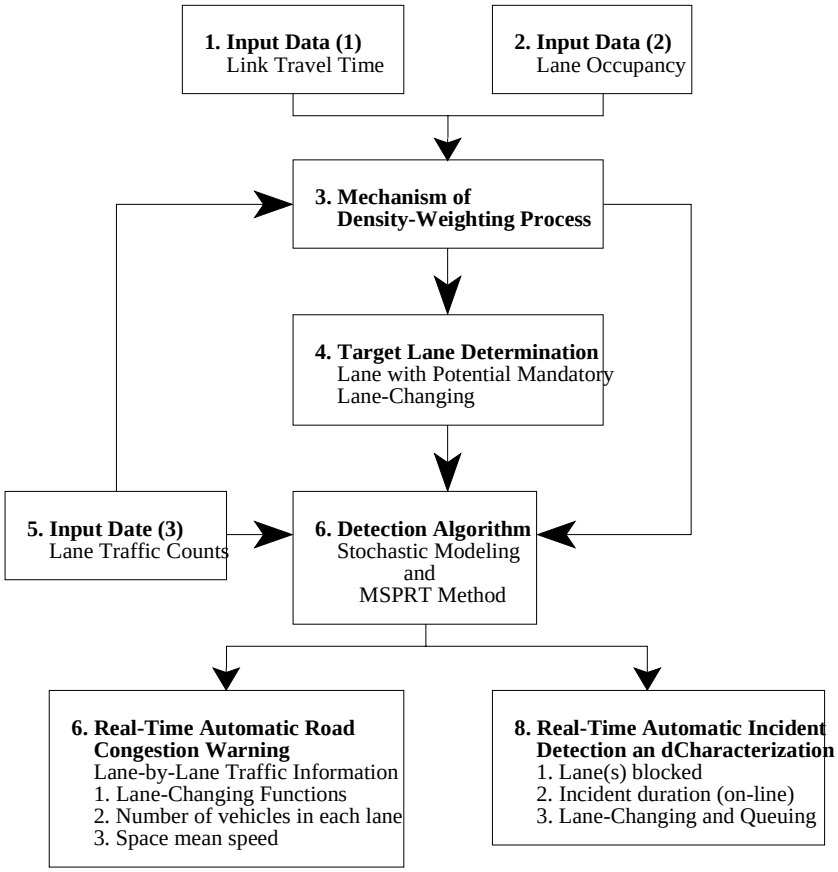


Figure 6.1 Proposed Framework for Real-Time Traffic Congestion Detection

A density-weighting process is designed in the framework used to decompose the uniform section density which is derived from traffic counts and link travel time. The weighting value associated with each lane density is given by:

$$w_j(k) = \frac{|\Delta o_j(k)|}{\sum_{\forall lane j} |\Delta o_j(k)|}$$

where:

$\Delta o_j(k) = o_j^{up}(k) \quad o_j^{dn}(k)$, corresponding to the difference between upstream and downstream lane occupancy values measured in lane j in time step k.

In contrast to queue length primarily used for non-incident congestion detection, mandatory lane changing is an important variable utilized for incident detection in this framework. Target lane determination is a procedure conducted in the framework to identify the lane with the highest density in which mandatory lane-changing may occur. Once a lane is coded as a target lane, the intra-lane and inter-lane traffic characteristics related to the target lane will be dynamically estimated in the detection algorithm for incident and congestion detection.

Three basic assumptions need to be postulated in order to facilitate construction of the nonlinear stochastic model proposed to predict real-time traffic characteristics. These are as follows:

1. Each vehicle in a target lane has the same lane-changing probability during a given time step. This implies that at a given time step, the value of the mandatory lane-changing probability for each vehicle in a target lane is identical to the lane-changing fraction in the target lane.
2. All state variables in the stochastic system follow homogeneous Gaussian-Markov processes. This assumption favors the setup of the recursive equations of the stochastic model. In the recursive equations of the model, the prediction of the next-step state variables depends on: 1) deterministic terms which follow homogeneous Markov processes, and 2) noise terms which follow Gaussian processes. Without the noise terms, the state variables are assumed to be homogeneous Markov processes. This means that in any time step the current-step state variables depend only on the state variables of the previous time step if the noise terms do not exist in the system.
3. Among the three types of state variables, each type of state variable is not correlated and independent on the other two types of state variables. Thus, the joint distribution of theses state variables follows a Gaussian-Markov process as well.

A stochastic system modeling approach is utilized to construct the lane-changing prediction models

in this research. The system constructed in the framework is specified by employing six types of state variables defined in this study. The following is a generalized form of the stochastic model:

$X(k\ \ 1) = f[x(k),k] \quad L[x(k),k]w(k)$ (recursive equation)

$Z(k\ \ 1) = h[x(k\ \ 1),k\ \ 1] \quad v(k\ \ 1)$ (measurement equation)

$0 \quad x(k\ \ 1)\ \ 1$ (boundary constraints)

The recursive equations represents a group of equations which imply the relationships of next-step and current-step state variables in the stochastic system. Recursive equations involve a deterministic vector *f[x(k),k]* and a noise term which is decomposed into two elements as shown in Matrix *L[x(k),k]* and Vector *w(k).* Vector *f[x(k),k]* is constructed on the basis of the second assumption described previously. If noise terms do not exist in the model, the state variables will follow Markov property only and the next-step state variables will depend only on the current-step variables. However, in the case of mandatory lane changing caused by unusual events such as incidents and queue-overflow occurrences, the state variables are unstable and affected sharply by traffic characteristics. The noise term of the recursive equations is thus developed to clarify the effects of traffic characteristics in the prediction of the state variables.

The measurement equation indicates the relationships among measured lane traffic counts and state variables. Matrix *Z(k+1)* represents a group of traffic counts in the lanes adjacent to target lanes collected at time step *k+1*. Each element in Vector *h[x(k+1),k+1]* indicates the components of traffic counts in each adjacent lane. A white noise vector, *v(k+1),* is added in the equation to take into account the errors of collected data due to malfunction of detectors, or inaccuracy of input data.

The boundary constraints are associated with each state variable in the system. In the boundary constraints, x(k+1) refers to the estimated value of any state variable at time step k+1. Each state variable has upper and lower boundary values, and thus, a generalized form of the boundary constraint is added in the model to keep the estimate of each state variable within its upper and lower

bounds.

To compute the unbiased estimates of the state variables of the model such that the mean square error of the estimation is minimized, a recursive estimation algorithm is used. It involves three major procedures: 1) an extended Kalman filter, 2) truncation and normalization procedures for the boundary restrictions on the state variables, and 3) a density-updating procedure for updating the number of vehicles either in target lanes or in adjacent lanes.

A MSPRT-based detection algorithm is then developed to distinguish incident-related congestion from non-incident congestion utilizing two decision variables: real-time estimates of mandatory lane-changing probabilities and queue lengths previously generated in the framework. The advantages offered by the proposed algorithm concerning real-time incident detection are addressed elsewhere in (Sheu, 1997).

6.1.3 Preliminary Tests for Incident Detection

Test scenarios associated with different congestion cases were designed in the research. The scope of this section is limited to the case of incident-related congestion. Preliminary tests in this scenario were conducted in order to demonstrate the feasibility of the proposed method for real-time incident detection on freeways.

The results were presented according to three performance measures: detection rate (DR), false alarm rate (FAR), and time to detection (D). In addition, three different persistence tests were conducted associated with each simulation, including none-interval persistence test (an incident is identified once the decision associated with incident-occurrence is made), one-interval persistence test (an incident is not identified until two consecutive decisions associated with incident-occurrence are made), and two-interval persistence test (an incident is not identified until three consecutive decisions associated with incident-occurrence are made). Tables 6.1 and 6.2 are the test results based on 36 simulated data sets generated from the INTRAS simulation model and 12 I-880 field data sets

(collected on the I-880 freeway in Oakland, California) respectively.

Overall, the test results, either using simulated data or using real data, demonstrated the feasibility of real-time incident detection on freeways using the proposed models. More importantly, the proposed method proved to remarkably improve performance in terms of time to detection.

Table 6.1 Results of Off-Line Tests for AID on Freeways (Based on Simulated Data)

Algorithm Performance			
Persistence	TTD (seconds)	DR (%)	FAR (%)
0	20.83	100	0
1	57.92	100	0
2	87.08	100	0
3	110.12	100	0

Table 6.2 Results of Off-Line Tests for AID on Freeways (Based on Field Data)

Algorithm Performance			
Persistence	TTD (seconds)	DR (%)	FAR (%)
0	9	100	0
1	47	100	0
2	71	100	0
3	90	95	0

6.1.4 Conclusions

Despite the early technologies which have been proposed for road traffic congestion detection, inadequate provision of intra-lane and inter-lane traffic information remains in developing

technologies for the use in ATMS and ATIS. Without providing real-time lane traffic information, conventional methods are restricted to road congestion detection. The result is that unsatisfying detection performance continues to be a source of frustration in implementation of road congestion detection systems.

In this section, a new framework which is capable of estimating real-time intra-lane and inter-lane traffic characteristics for the use of real-time automatic road congestion detection is proposed. The proposed framework is founded on the basis of a nonlinear stochastic system modeling approach, and a MSPRT-based algorithm. Further tasks concerning this study will be focused on model tests for the case of non-incident congestion as well as the use of congestion characterization.

6.2 DYNAMIC ORIGIN/DESTINATION ESTIMATION USING TRUE SECTION DENSITIES

6.2.1 Introduction

The problem of dynamic origin/destination demand estimation is vital to all aspects of ATMIS (Advanced Transportation Management and Information Systems) research. This estimation problem involves the determination of the number of trips going from an origin to a destination during a given time step. Any traffic assignment algorithm is dependent on accurate origin/destination demands as input.

The proposed solution uses a framework with an embedded dynamic traffic assignment model. Kalman filtering and other time series analysis methods will be used to track the demands and to minimize the error between estimated and observed section densities. This project therefore uses the section densities derived from the vehicle reidentification algorithms.

6.2.2 Background

An origin/destination demand matrix is a vector with components that denote the average number

of trips going from an origin to a destination. Traditionally, the method for obtaining an origin/destination matrix is to employ the use of household surveys coupled with roadside surveys. However, this method is costly and also is not feasible for real-time applications. Therefore there has been research in origin/destination matrix estimation problem by using network-wide link traffic counts and combining it with other available information. One source of information is household activity surveys, another source is the knowledge of the a priori probability distribution of the origin/destination matrix, and yet another is the use of previously estimated origin/destination matrices.

Recently, there has been increased research in the area of dynamic origin/destination demand matrix estimation. There have been a number of procedures proposed. Most of them assume that the time taken by vehicles to traverse the network under consideration is smaller than the time step of the procedure. In the case of an intersection, origin/destination matrix estimation becomes the problem of finding the turning proportions.

A simple way of classifying dynamic origin/destination demand estimation is to separate the approaches into two camps: assignment-based and non-assignment based. This classification scheme was used in Tao and Chang (1996).

The non-assignment-based approaches by definition do not use dynamic traffic assignment. They estimate origin/destination parameters directly from time-series input/output flows. Sometimes prior origin/destination matrices are used for initialization. Cremer and Keller (1987) highlighted the causal relationships that exist between the time variable sequences of entrance flow volumes and the sequences of short-time exit flow counts. They proposed four methods which were ordinary least squares estimator, constrained optimization, Kalman filter, and "simple recursive" formulation. Nihan and Davis (1987) proposed using recursive prediction error techniques and input/output volume counts. They discussed different search methods and concluded that Gauss-Newton appear superior to gradient-based search methods. Later in 1989, they formulated a method of estimating intersection turning and through movement probabilities. van der Zijpp[sic] and Hamerslag (1994)

augmented previous formulations by incorporating inner-link count data. Furthermore, van der Zijpp (1997) added partial origin/destination information by using Automatic Vehicle Identification (AVI) technology. van der Zijpp also proposed Bayesian updating schemes that used multivariate normal and truncated multivariate normal assumptions for the subjective probability distribution. Bell (1991) formulates an approach using platoon dispersion with the assumption that the time taken to traverse a junction is either small compared to the chosen time interval or is equal to a fixed number of time intervals. Chang and Wu (1994) uses macroscopic traffic flow characteristics in addition to input/output flows and inner flows. Wu and Chang (1995) increases observability of dynamic relations through arbitrary selected screenlines which reside in the path of an origin/destination pair. The resulting screenline flows are estimated from normal distribution assumptions and a mean arrival time. Chang and Tao (1996) expand the freeway case to urban networks by adding signalization.

The assignment-based approaches depend on a reliable and descriptive dynamic model for generating the network link flow usage pattern. Willumsen (1984) extends static entropy to multi-time intervals while using CONTRAM . Cascetta et al. (1993) discusses two types of estimators: simultaneous and sequential. Camus et al. (1994) added historical information and combined it with traffic counts at on-ramps. Ashok and Ben-Akiva (1993) build upon Okutani's work (1987) by formulating an algorithm using deviations of historical origin/destination flows. They use travel time and path choice fractions to determine the link incidence matrix and even discuss origin/destination demand prediction. Van Aerde's (1993) QUEENSOD backcalculates origin/destination demand estimates from the dynamic link flows derived from dynamic link use probabilities from simulation. Although this is not a strictly dynamic assignment-based approach, it performs the temporal and spatial mapping of link flows of the assignment-based approaches.

The challenge in dealing with dynamic formulations is the inclusion of short time steps in addition to the usual spatial patterns. Consequently, static methods that use flows aggregated over time are no longer valid. To determine the dynamic origin/destination demand matrix, the number of vehicles exiting from sources and entering into sinks during a small time period needs to be

estimated. This demand results in a traffic pattern that is distributed onto the network. Therefore, the demand is what produces the traffic pattern, but the traffic pattern does not cause the demand even though it changes the route choice of the demand. Past formulations use vehicle flows (or even occupancy) as the input variable to both dynamic o/d estimation and traffic assignment formulations. However, the danger with this approach is the fact that a point measure such as flow, is assumed to be spread evenly across the link. PATH research has shown that point values are rather unstable and do not accurately reflect the true traffic distribution over the entire roadway section (Sun, 1998). A more appropriate parameter to use in dynamic formulations is section density. By knowing the number of vehicles on all the links of a network at small time steps, one can reconstruct the o/d demand pattern that produced such traffic distributions through time series tracking of the demand pattern.

This proposal presents a practical approach for dynamic origin/destination demand estimation. The proposed dynamic origin/destination demand estimation framework addresses many of the shortcomings of existing formulations and presents a formulation for general networks and not just corridors. One unique feature of this framework is its use of section density as a variable instead of flow. If section densities are not available, then estimates can be derived by using spatial averages of occupancies corrected with measured vehicle lengths.

The objective of this project is to formulate a dynamic origin/destination framework that is both theoretically sound and field implementable. This involves fundamental research in both the dynamic assignment problem and the dynamic estimation problem.

6.2.3 Methodology

This proposal will try to build upon the foundation of static origin/destination matrix estimation by adding the temporal aspect. From the vehicle reidentification algorithms, we hope to obtain temporal section densities and also partial real time origin/destination demand information. With this information we focus on the temporal aspect of estimators and work with the classical methods

such as maximum entropy, maximum likelihood, or least squares estimators.

The dynamic origin/destination framework is composed of two interacting modules: DOE (Dynamic Origin/Destination Estimation) and DTA (Dynamic Traffic Assignment). The overall framework is illustrated graphically in figure 6.2. The DOE will estimate the OD matrix that generated observed section densities. The DTA in turn uses these demands and returns to the ODE newly computed section densities and a dynamic assignment matrix. The equilibration between the two modules will minimize the discrepancies between observed section densities and the ones computed by the framework.

The DTA model that is proposed for use in the estimation of origin/destination matrix is a bi-level optimization framework that incorporates fundamental traffic flow relationships (Jayakrishnan et al., 1995). The use of traffic flow relationships is significant since it bypasses the necessity to use a flow dependent link cost function such as the well know BPR (Bureau of Public Roads) link cost function.

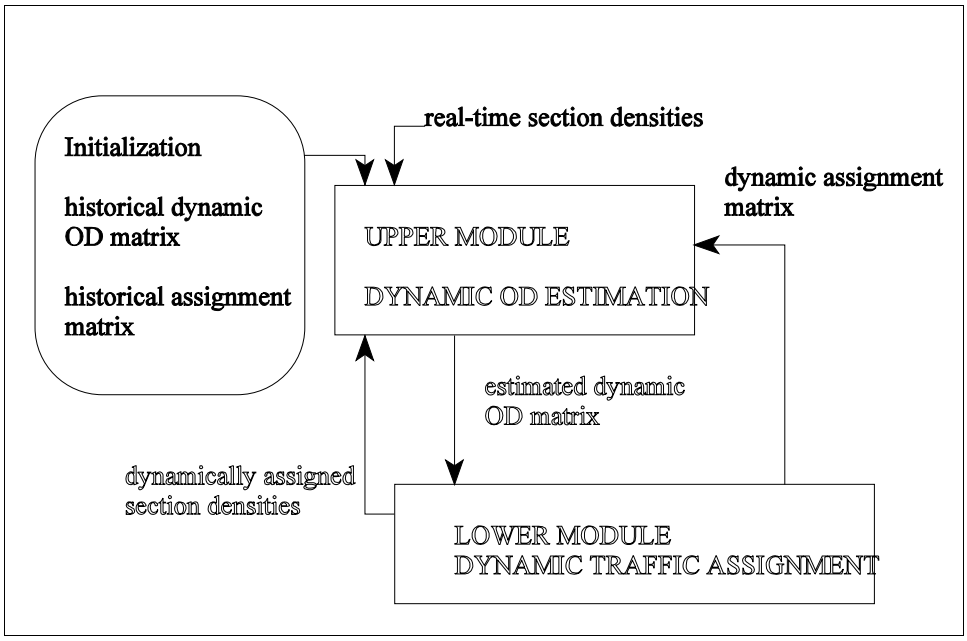


Figure 6.2 Graphical Illustration of the Dynamic Origin/Destination Estimation Framework

The formulation of the DOE uses the Kalman filtering technique. This technique has many advantages including the following benefits:

1. It is suitable for linear systems. The OD matrix estimation problem can be formulated with a linear mapping from the state variable to the measurement variable.
2. The formulation is stochastic which allows for measurement and state errors, and the initial state could be described stochastically.
3. It is dynamic and is suited for signal tracking, or more specifically, OD matrix tracking over time.
4. The formulation is recursive, and allows for each additional measurement to be incorporated sequentially. With the previous estimate and the new measurement, a new estimate is produced.

Both measured and simulated data will be utilized to develop and test the resulting framework. Initial investigations will focus on small experimental networks. These networks include grid networks that represent the characteristics of actual urban networks. Larger networks will be utilized for studying computational issues once the framework is adequately developed.

Even though densities are the primary input variables of the estimation system, travel times can be used to calibrate part of the estimation process. The proposed dynamic o/d estimation system embeds a dynamic traffic assignment module which determines reasonable paths and assigns traffic onto links. One task of the traffic assignment is to determine the travel times of each link and each route. If measured travel times are available, then this can be used to correct the travel times and produce more accurate route choice decisions by travelers. Testbed modeling of both dynamic traffic assignment and dynamic o/d estimation can benefit greatly from measured travel times for links and entire paths. In addition, other measures can also be incorporated to further constrain the solution of the estimation problem. For example, if partial o/d demand can be measured (either by signature analysis, toll tag, license plate, etc...), then it can be used in conjunction with densities in a stochastic system (van der Zijpp, 1997).

6.3 AN INVESTIGATION IN THE USE OF INDUCTIVE LOOP SIGNATURES FOR VEHICLE CLASSIFICATION

6.3.1 Introduction

This proposal presents an advanced traffic surveillance technique based on pattern recognition and the use of current inductive loop technology. The focus is to investigate the feasibility of using inductive loop signatures for obtaining vehicle classification information on a network-wide level. The potential benefits from the vehicle classification information include improvements in vehicle reidentification algorithms, roadway maintenance, vehicle emissions management, roadway design, and automatic toll collection. The compilation of a vehicle classification database is also a valuable resource for researchers in the areas of transportation planning and control. This effort continues the efforts in vehicle signature analysis started from the research in vehicle reidentification for the derivations of section-related measures of traffic system preformance.

This application calls for the use of different pattern recognition techniques for vehicle classification. Both classical template matching and advanced neural networks will be employed in this research. The trade-offs between the different methods will be studied.

6.3.2 Background

Vehicle classification is the process of sorting vehicles according to the vehicle type. The Federal Highway Administration (FHWA), California Department of Transportation (Caltrans), County Transportation Authorities, and even private organizations such as tollway agencies all use vehicle classification in diverse ways. Therefore the set of vehicle types is defined differently for each agency or organization. Research in vehicle classification will assist public and private agencies in different ways including road management, road maintenance, data collection, multi-modal transport, emissions control, bus preemption, and toll assessment functions. For example, Caltrans can use vehicle classification data to predict roadwear (which is mainly due to trucks), and to schedule

appropriate maintenance. Another example would be tollway agencies using the vehicle class to assign different toll rates. In addition to the aforementioned benefits, vehicle classification can also help to improve vehicle re-identification for the derivation of section related measures such as travel times. Since each vehicle class has different inductive signature characteristics, the derivation of a confidence probability for a vehicle class can improve the vehicle re-identification algorithm. The identification of vehicle classes also results in the derivation of section related measures for each vehicle class.

Previous research in this field include inductive loop detectors and other types of detectors. Pursula and Pikkarainen (1994) demonstrated the feasibility of using inductive loops for classification in Europe. Lu et al. (1992) demonstrated the use of infrared image analysis for vehicle classification. Also, there are no theoretical constraints on the use of other types of detectors for this purpose. Therefore Weigh-In-Motion sensors or video could also serve this purpose in the future. There are also commercial classification systems such as the IVHS2000 by Intersection Development Corporation. The existence of commercial products should not undermine other research efforts in this area. One goal of this applitcation is to consider the best applications of this technology. A commercial system might only be designed with certain applications or agencies in mind. Another problem with commercial systems is the fact that detailed performance analysis and methodology is not published. The development of such a system will allow agencies the ability to understand the theory behind the operation of such systems, and will allow agencies to even develop their own field software according to their needs.

One advantage of the proposed system is the utilization of current infrastructure. Since inductive loops are so prevalent on both city and state roadways, there exists the possibility of a network-wide implementation of the system on freeways and arterials. Current commercial vehicle classification systems require the use of a multi-beam light curtain and a Doppler radar, or treadles and loops. Compared to existing commercial implementations, the proposed system has much simpler equipment needs.

There is strong evidence that a given inductive signature will have a similar inductive signature at different sites. A site is composed of the roadway geometry, physical inductive loop detector, lead-in cables, and detector circuitry. This signature changes with vehicle speed, vehicle offset (both horizontally and vertically), loop type, and other factors. However, there are some changes that can be eliminated. For example, the signature can be transformed to be speed invariant by using a speed measurement (double loops, radar, single loop estimate). Also for example, the nominal site inductance level, which is a function of the loop geometry and detector settings, can be normalized for preserving the shape of vehicle signatures. In addition, the nominal site inductance level can also be calibrated for each individual site so that the total amount of signal (power) can be meaningful. Since the shapes of the vehicles are preserved to some extent by different transformation and normalization methods, a value proportional to the total power is the maximum inductance recorded. This maximum inductance is a easily computable value that can be used as a feature vector. More complex Fourier or Wavelet analysis can also be used for deriving the power spectrum. One evidence for the assumption that vehicle signatures at different classification sites can be similar is the research on vehicle reidentification. The vehicle reidentification results show that individual vehicles can be reidentified from one site to another site with accuracy of between 60% to 75%. In order for this to be possible, the signature features had to be preserved from one site to another. Because it was feasible to classify individual vehicles, a coarser classification according to vehicle types should also be possible.

6.3.3 Methodology

The derivation of vehicle classes is accomplished by using pattern recognition techniques. The inductive loop signatures from different vehicle classes normally exhibit differences in magnitude, length, and shape. Classical methods such as clusterization and template matching, and advanced neural networks can be used effectively with inductive loop signatures to yield relevant vehicle classes.

One classical method of accomplishing classification is by template matching. A template is a

waveform that is representative of the entire vehicle class. Therefore the number of vehicle classes or templates is determined beforehand. The template can be the average of a fixed number of vehicle signatures or an idealized signature such as a fitted polynomial curve. Thus a vehicle signature is classified as a particular class if the signature results in the closest match to the particular template.

One type of neural network that will be used for this system is the self-organizing network. This type of network uses neurons of competitive networks to learn to identify groups of similar input vectors (Demuth, 1994). Another network is the Carpenter/Grossberg classifier. This adaptive resonance network is trained without supervision and forms clusters of vectors without apriori determination of the number of total clusters. Figures 6.3 and 6.4 illustrate the Carpenter/Grossberg and the self-organizing network architectures (Lippmann, 1987).

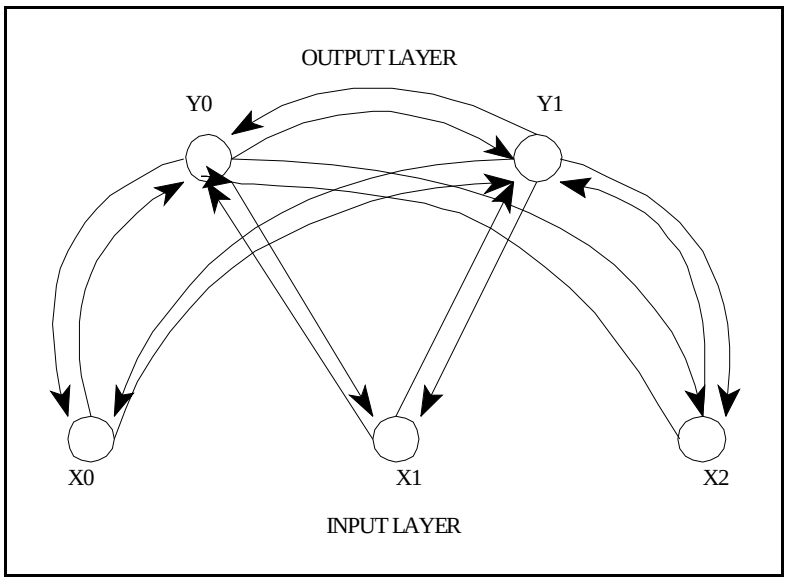


Figure 6.3 Major Components of the Carpenter/Grossberg Classification net

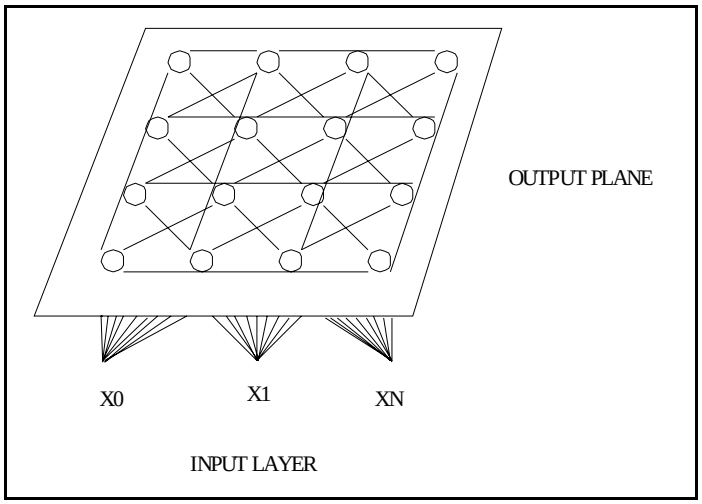


Figure 6.4 Self-organizing Network Architecture

There is a significant database of inductive loop signatures obtained from a field site on the westbound SR-24 freeway in Lafayatte, California in December 1996. Two data acquisitions stations were instrumented with video, loop waveform dataloggers, and speed trap dataloggers. Standard 6ftx6ft (1.82mx1.82m) loops were used at both stations. The reduced dataset contains the waveforms of the upstream and downstream vehicles along with their speeds, electrical length (derived from occupancy time), arrival time at station, and the ground truthed vehicle identification number. This data set contains vehicles from various classes ranging from semi-trailer trucks to passenger vehicles.

A second database has been collected from an arterial in Irvine, California as part of Phase II. This data collection used more advanced inductive loop detector cards that sampled more frequently than previous detectors. The signatures from this database are expected to have better resolution than the previous database.

It is important to note that research in developing inductive signature classification can also be applied to other types of signatures. For example, if infrared or acoustic signatures become available, the inductive signature algorithms can be a basis for building classification algorithms for

new types of signatures.

CHAPTER 7. CONCLUSION

7.1 SUMMARY OF RESEARCH CONTRIBUTIONS

This research on lexicographic optimization for vehicle reidentification on freeways involved a wide range of issues including field instrumentation, mathematical formulation, and useful applications. Therefore, there was an opportunity to explore not just the theoretical aspects of problem formulation and solution, but also practical aspects of field implementation and applications. The original contributions of this research include the definition of the vehicle reidentification problem, the identification of critical feature vectors, the formulation and solution of the lexicographic optimization, the consideration of dynamic aspects of real-time algorithms, the derivation of section-related measures of traffic system performance, the design of field instrumentation, and the discussion of ATMIS applications.

The discussion on fundamental traffic flow variables is important in investigating dynamic algorithms for ITS. First, it is important to understand the difference between point measures and section traffic measures. The results of this research show that point measures over short time periods yield values that fluctuate much more than section measures. This research also shows that significant errors can result from the estimation of section measures by extrapolating point measures. Thus, the granularity of time versus the granularity of space must be considered carefully in the development of dynamic algorithms. For example, in short time step applications such as dynamic origin/destination demand estimation, section density should be used as the state variable since the origin/destination demand is distributed over the entire network. Section measures converge to point measures as the section is reduced to be infinitely small. The particular advantage in using section density versus flow is in its monotonic relationship with travel time. Many ITS algorithms use BPR-like (Bureau of Public Roads) cost functions with vehicle flow as the independent variable. Since flow versus travel time is not a monotonic function, this presents problems in determining cost values.

A significant original contribution of this research is the definition of the vehicle reidentification problem. This problem is defined by using downstream and upstream waveform pairs of vehicles but is equivalent to scanning the upstream vehicle set for a given downstream vehicle. The unique usage of pairs of vehicles as variables enables the use of an optimization formulation.

Another significant aspect of the vehicle reidentification problem is the fact that it is defined to match individual vehicles and not just platoons. This is a more difficult problem than platoon vehicle signature reidentification investigated by previous researchers, because a group of vehicle waveforms is more unique than an individual waveform. The proposed vehicle reidentification problem is not subject to the constraint that upstream and downstream vehicle sequences need to be similar. For example, the SR-24 freeway data contains numerous vehicle movements that disrupt platoon sequences from one station to the next. In fact, as the sequence or platoon length reduces to a single vehicle, the vehicle reidentification problem results. In addition, the solution of the vehicle reidentification problem allows the measurement of lane-by-lane traffic movement and "partial" origin/destination demands which is not achievable with previous platoon matching methods.

A critical contribution in system identification is the determination of appropriate feature vectors. A feature vector set needs to be not only effective but also parsimonious. Redundancy of feature vectors need to be eliminated to reduce communication and computation burdens in real-time field implementations. This report details the process of deriving the processed raw signature or "shape" feature vector by using time and length transformations. The use of feature vectors that are dependent on traffic flow characteristics was described, and ineffective feature vectors such as the waveform frequency spectrum were identified.

Another significant contribution was the formulation and solution of the lexicographical optimization of the vehicle reidentification problem. Since this is a multi-objective optimization approach, it allows the incorporation of many optimization criteria. Goal programming was used to reduce the feasible set which on one hand improves the reidentification accuracy, and on the other

hand reduces the computational burden of the algorithm. The attractiveness of the lexicographical approach is that the previous level optimization serves as a constraint for the next optimization. Bayesian analysis was incorporated in the lexicographic optimization to utilize historical data and to compensate for the stochasticity of the vehicle waveforms. The Bayesian posteriori probability can also be used as a confidence measure of the vehicle reidentification. In other words it can be used as origin/destination demand measurement probability. An analysis of the sensitivity of speed measurements was performed for the lexicographic method. A new single loop speed computation method was derived for use in situations where the detector infrastructure is limited to single loops.

The discussion of the dynamic aspects of the algorithm is important for real-time implementation. One dynamic aspect is the specification of optimization goal programs using real-time information. Another aspect is the presentation of a set of discriminant functions that can be selected based on the traffic conditions. The reidentification algorithms are robust because of the aforementioned dynamic adjustments.

A direct consequence of the lexicographic solution is section-related measures of traffic system performance. Three particular measures are section density, travel time, and origin/destination demand measurement. Section density is very costly to measure with other technologies such as aerial photography or automatic license plate reading. The proposed system is cost effective, since it utilizes the current surveillance infrastructure. The current inductive loop detector infrastructure is extensive which allows network-wide measurements of section density. Likewise, network-wide travel times can be measured and used in ATMIS strategies. Traveller information systems, traffic routing systems, and transportation management centers can all use accurate travel times as a direct input. Dynamic origin/destination demand is another measure that is extremely costly to obtain. This research presents a way where partial origin/destination demands can be measured with a level of confidence. This information can be used in conjunction with historical data to estimate current demands.

In order to be comprehensive, this report presents discussions on field data collection and

implementation. This discussion includes the various requirements for deriving reliable test data and the associated ground truth data. Some requirements that are detailed include proper video instrumentation, time synchronization, and surveillance hardware/software design. Details are given for building a system to collect and store vehicle signature data. Details are also given for the formulation of a ground truth database. A real-time field implementation can be extended from the discussions on data collection design.

A further significant contribution is the discussion on three ATMIS applications that utilize the vehicle reidentification technology. The first application is the development of a framework for real-time road traffic congestion detection. The second is a formulation of a dynamic origin/destination estimation scheme using true section densities. The third is an investigation in the use of inductive loop signatures for vehicle classification.

7.2 DIRECTIONS FOR FUTURE RESEARCH

One natural direction for future research is to continue refining the vehicle reidentification algorithm. This involves investigation into the derivation of other efficient feature vectors, the refinement of goal programs, and improvement in discriminant analysis in the lexicographic optimization.

An extension of the current algorithm involves the method of "per lane matching" of a sequence of vehicles for field data that does not contain much lane changing. The testing of this algorithm extension will be feasible once the arterial data collected from the City of Irvine are processed. Since a sequence contains information about relative vehicle positions, it should theoretically improve matching accuracy over matching single vehicles when vehicle sequences are more or less preserved.

Figure 7.1 is a rough flowchart of a sequence matching algorithm. The vehicle reidentification problem is expanded to be the reidentification of groups of vehicles. But the same basic approach as the original lexicographic algorithm can still be used with the consideration of the combinations

of vehicles.

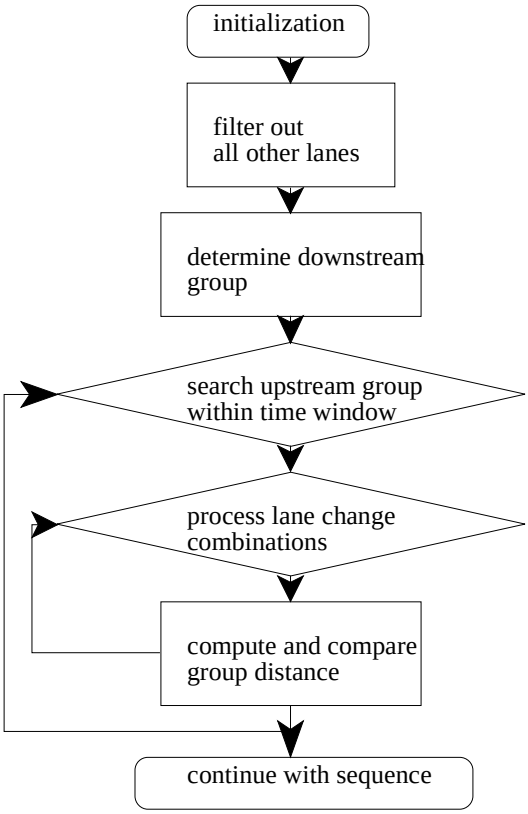


Figure 7.1. Vehicle Combination Matching Flowchart

Figure 7.2 serves to illustrative the vehicle combination matching algorithm. Suppose one is interested in matching a downstream group number of g_d=5 vehicles from a time window size of g_u=8 upstream vehicles. Of course with field data applications, the sizes of the downstream group and the upstream time window can be significantly larger than the example provided. Four different upstream combinations result from shifting the downstream number of vehicles by a single vehicle.

In other words, g_u-g_d+1=4. If lane changing occurs between upstream and downstream station, then the combinations of when a vehicle was absent in the downstream sequence needs to be considered. This results in $g_{d=}5$ combinations for a single vehicle being absent. In total, there needs to be 20 comparisons between different combinations of upstream and downstream groups. For simplicity, only the case where vehicles have changed lanes from the upstream platoon is considered. To generalize this method, the number of combinations is simply extended. In the limit that the downstream group size becomes 1, the original single vehicle matching algorithm is obtained.

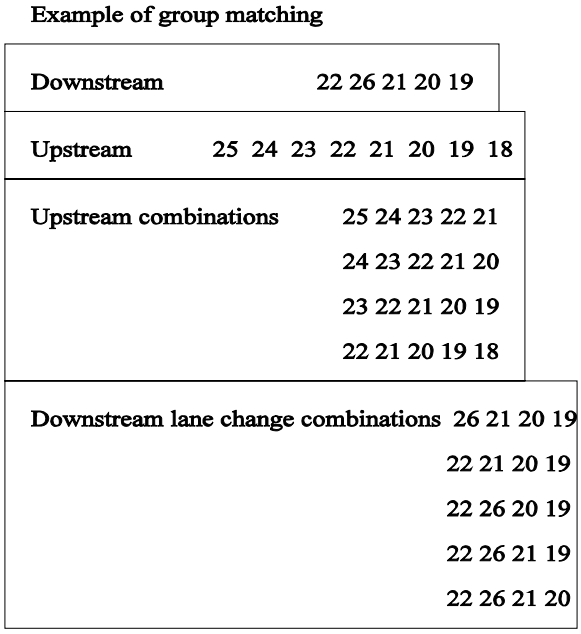


Figure 7.2. Illustrative Example of Vehicle Combination Matching

In order to optimize the group size for the sequence matching algorithm, the trade offs between accuracy and computational burden need to be considered. As the group size increases, the group distances becomes more unique, but at the same time the number of vehicles being absent also increases. The consideration of the different combinations within a time window is only considered in the beginning to find a reference point. After a reference point is established in the upstream sequence, a much smaller time window is needed.

Apart from refinements to the vehicle reidentification algorithm, there also needs to be further research on real-time implementation issues in order to apply the off-line algorithms in the field. One aspect of this implementation involves investigating the dynamic aspects of the algorithm, such as refining goal programs. The distribution of processing power and the resulting communications needs for the real-time system have to be investigated. One real-time scenario involves local processing of waveforms and the execution of the reidentification algorithm at the downstream station. Another option involves central processing of all waveforms in the TMC (Traffic Management Center). The different communications infrastructure and bandwidth associated with each scenario have to be carefully studied.

This research can be applied in various applications. One application is to use the proposed waveform analysis method for vehicle classification. This can result in benefits in road maintenance, toll collection, vehicle emissions management, bus preemption, and traffic data collection. Another application is in the area of incident detection and congestion management. The proposed system is particularly attractive in the potential for fast detection times. Unlike point measurement systems, the section-based algorithm does not need to wait for the upstream arrival of the shockwave resulting from the incident or congestion. Travel time is a section measure that has great potential in traveller information systems. Direct measurement of travel time will result in smaller error and variance compared with the current means of extrapolating travel time. Along the same lines, route guidance systems could also benefit from the measurement of travel times for determining dynamic shortest paths.

REFERENCES

3M. (1997) Canoga Vehicle Detection System Installation Instructions. 3M Intelligent Transportation Systems. 223-3N 3M Center. St. Paul, MN 55144.

Ashok, K. and Ben-Akiva, M.E. (1993) Dynamic Origin-Destination Matrix Estimation and Prediction for Real-Time Traffic Management Systems. Transportation and Traffic Theory, Elsevier Science Publishers B.V., pp. 465-484.

Bahler, S., Minge, E., and Kranig, J. (1998) Field Test of Non-Intrusive Traffic Detection Technologies. Paper No. 980666. Preprints. Transportation Research Board, 77th Annual Meeting. January 11-15. Washington, D.C.

Bell, M.G.H. (1991) The Real Time Estimation of Origin-Destination Flows in the Presence of Platoon Dispersion. Transpn. Res. B. Vol. 25B. Nos. 2/3, pp. 115-125.

Böhnke, P. and Pfannerstill, E. (1986) A System for the Automatic Detection of Traffic Situations. ITE Journal. Vol. 56.

Böhnke, P. (1995) ave Innovative Verkehrs-und Informationstechnik. (ave Innovative Traffic Engineering and Information Systems). Profile and MAVE information. Dennewartstraße 27. D-52068. Aachen. GmbH.

Bow, S. (1992) Pattern Recognition and Image Preprocessing. Marcel Dekker. New York.

Bowerman, R. and Hall, B. (1995) A Multi-Objective Optimization Approach to Urban School Bus Routing: Formulation and Solution Method. Transpn. Res.-A. Vol. 29A, No.2, pp. 107-123.

Brown, R.G. (1983) Introduction to Random Signal Analysis and Kalman Filtering. John Wiley&Sons. New York.

Camus, R. et al. (1994) O-D Prediction for Real-Time Traffic Management. IFAC.

Cascetta, E. and Nguyen, S. (1988) A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts. Transpn. Res. B. Vol 22B. No. 6. pp. 437-455.

Cascetta, E. (1993) Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts. Transportation Science. No. 4. November 1993. Operations Research Society of America.

Chang, G., and Kao, Y. (1990) An Empirical Investigation of Macroscopic Lane-Changing Characteristic on Uncongested Multilane Freeways. Transpn. Res.-A, Vol 25A, No.6, pp. 375-389.

Chang, G. and Wu, J. (1994) Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors. Transpn. Res. Vol. 28B, No. 2, pp.141-160.

Chang, G.L. and Tao, Xianding. (1996) Estimation of Dynamic O-D Distributions for Urban Networks. Thirteenth International Symposium on Transportation and Traffic Theory. Lyon. France.

Cheng, D.K. (1985) Field and Wave Electromagnetics. Addison-Wesley. Reading.

Coifman, B. A. (1996) A New Methodology for Smoothing Freeway Loop Detector Data an Introduction to Digital Filtering. Proceedings of the 1996 Transportation Research Board Meeting. Washington. D.C. January.

Coifman, B. A. (1998) Vehicle Reidentification and Travel Time Measurement in Real-Time on Freeways Using the Existing Loop Detector Infastructure. Preprints. Transportation Research Board. 77th Annual Meeting. January 11-15. Washington, D.C.

Cremer, M. and Keller, H. A New Class of Dynamic Methods for the Identification of Origin-Destination Flows. Transpn. Res. B. Vol. 21B. No. 2.

Dailey, D. (1993) Travel-Time Estimation Using Cross-Correlation Techniques. Transpn. Res.-B. Vol.27B, No.2, pp.97-107. Pergamon Press Ltd.

Detector Systems. *Product Manual: Two Channel Digital Loop Detector Model 222.* Stanton, California.

Gartner, N.H. (1983) OPAC. A Demand-Responsive Strategy for Traffic Signal Control. Transportation Research Record 906, TRB, pp. 75-81, 1983.
Gartner, N.H. and Stamatiadis, C. (1997) Integration of Dynamic Traffic Assignment with Real Time Traffic Adaptive Control. Paper No. 97-0857 presented at 76th Annual TRB Meeting.

Hawkins, T. (1970) Lebesgue's Theory of Integration. The University of Wisconsin Press. Madison.

Institute of Transportation Engineers. (1990) Traffic Detector Handbook. Second Edition.

Intersection Development Corporation. (1995) Models 262&272 Two Channel Digital Inductive Loop Vehicle Detector Technical Manual. 1511 E. Orangethorpe Ave., Suite A. Fullerton, CA 92631. 1511 E. Orangethorpe Ave., Suite A. Fullerton, CA 92631.

Intersection Development Corporation. (1997) Model IVS-2000 Intelligent Vehicle Sensor Technical Information.

Janson, B. N. and Southworth, F. (1992) Estimating Departure Times from Traffic Counts Using Dynamic Assignment. Transpn. Res. B. Vol. 26B, No. 1, pp. 3-16.

Jayakrishnan, R., Tsai, W. K., and, Chen A. (1995) A Dynamic Traffic Assignment Model with Traffic Flow Relationships. Transpn. Res. C. Vol. 3, No. 1.

Kapur, K. (1970) Mathematical Methods of Optimization for Multi-Objective Transportation Systems. Paper presented at the Operations Research Society of America Meeting. Washington, D.C. April 20-22.

Klein, L. (1995) Modern Detector Technology for Traffic Mangement. Presentation for University of California, Irvine. February 28. Unpublished work.

Koutsopoulos, H.N. and Xu, H. (1993) An Information Discounting Routing Strategy for Advanced Traveler Information Systems. Transpn. Res. C. Vol. 1. No. 3.

Kühne, R. D. (1991) Freeway Control Using a Dynamic Traffic Flow Model and Vehicle Reidentfication Techniques. Transportation Research Record 1320. Pp. 251-259.

Kühne, R. D and Immes, S. (1993) Freeway Control Systems for Using Section-Related Traffic Variable Detection. Pacific Rim TransTech Conference, July 25-28, Seattle, Washington.

Kühne, R. D. (1994) First Experiences with Real Time Motorway Control Using Section-Related Data Collection. IFAC. Tianjin, PRC. Pp. 651-653.

Kühne, R., Palen, J., Gardner, C., and Ritchie, S. (1997) Section-Related Measures of Traffic System Performance. Presented at the 1997 Transportation Research Board Annual Meeting. Washington D.C.

Lippmann, R. (1987) An Introduction to Computing with Neural Nets. IEEE Acoustics, Speech and Signal Processing Magazine. April. Pgs. 4-22.

Ljung, L. (1987) System Identification: Theory for the User. Prentice Hall. New Jersey.

Lockett, A. and Hetherington, B. (1985) The Analytic Hierarchy Process: Experiments in Stability. Mathematics of Multi Objective Optimization. Pps. 317-352. Springer-Verlag. New York.

Nihan, N.L. and Davis, G.A. (1987) Recursive Estimation of Origin-Destination Matrices from Input/Output Counts. Transpn. Res. B. Vol 21B. No.2.

Nihan, N.L. and Davis, G.A. (1989) Application of Prediction-Error Minimization and Maximum Likelihood to Estimate Intersection O-D Matrices from Traffic Counts.

Transportation Science. Vol. 23. No.2. May.

Oppenheim, A.V. and Schaffer, R.W. (1989) Discrete-Time Signal Processing. Prentice-Hall. New Jersey.

Parkany, E. and Bernstein, D. (1993) Using VRC Data for Incident Detection. Pacific Rim TransTech Conference, July 25-28, Seattle, Washington.

Peek Traffic. (1992) Sarasota Model 222&224 GP5 Multi Channel Loop Detector Sensor Unit Technical Manual. 1500 N. Washington Blvd. Sarasota, FL 34236-2723.

Rentmeesters, M., Tsai W., and Lin K. (1996) A Theory of Lexicographic Multi-Criteria Optimization. Proceedings of the Second IEEE International Conference on Engineering of Complex Computer Systems. Pps. 76-79. Los Alamitos, California. IEEE Computer Society Press.

Rentmeesters, M. (1998) A Theory of Multi-Objective Optimization: Comprehensive Kuhn Tucker Conditions for Lexicographic and Pareto Optima. Ph.D. Dissertation. University of California, Irvine.

Riinguest, J. (1992) Multiobjective Optimization: Behavioral and Computational Considerations. Kluwer Academic Publishers. Boston.

Rioul, O. and Vetterli, M. (1991) Wavelets and Signal Processing. IEEE Signal Processing Magazine. October. pp. 14-37.

Santina, M. S. et.al. (1994) Digital Control System Design. Fort Worth: Saunders College Pub. pp.605-653.

Schoenberg, I. J. (1946) *Quarterly of Applied Mathematics*. Vol. 4. pp. 45-99, 112-141.

Stadler, W. (1988) Multicriteria Optimization in Engineering and in the Sciences. Plenum Press. New York.

Steuer, R.E. (1986) Multiple Criteria Optimization: Theory, Computation, and Application. Wiley. New York.

Strang, G. (1994) Wavelets. American Scientist. Volume 82. May-June. pp. 250-255.

Sun, C., Ritchie, S., and Tsai, K. (1998) Algorithm Development for
Derivation of Section-Related Measures of Traffic System Performance Using Inductive Loop Detectors. Preprints. Transportation Research Board. 77th Annual Meeting. January 11-15. Washington, D.C.

Van Aerde, M. et al. (1993) QUEENSOD: A Method for Estimating Time Varying Origin-Destination Demands for Freeway Corridors/Networks. Annual Transportation Research Board Meeting. January. Washington, D.C.

van der Zijpp, N.J. and Hamerslag, R. (1994) Improved Kalman Filtering Approach for Estimating Origin-Destination Matrices for Freeway Corridors.

van der Zijpp, N.J. (1997) Dynamic OD-Matrix Estimation from Traffic Counts and Automated Vehicle Identification Data. Preprint. Transportation Research Board. 76th Annual Meeting. January 12-16. Washington, D.C.

Van Zuylen, H. J. and Willumsen, L. G. (1980) The Most Likely Trip Matrix Estimated from Traffic Counts. Transpn. Res. B. Vol 14B, pp. 281-293.

Wasserman, P. (1989) Neural Computing Theory and Practice. Van Nostrand Reinhold. New York.

Willumsen, L.G. (1984) Estimating Time-Dependent Trip Matrices from Traffic Counts. Ninth International Symposium on Transportation and Traffic Theory.

Wu, J. and Chang, G.L. (1995) Estimation of Time-Varying O-D Matrices with Dynamic Screenline Flows. Preprint. Transportation Research Board. 74th Annual Meeting. January 22. Washington, D.C.

Zionts, S. (1985) Multiple Criteria Mathematical Programming: An Overview and Several Approaches. Mathematics of Multi Objective Optimization. Pps. 227-274. Springer-Verlag. New York.

APPENDIX A. SAMPLE VIDEO CORRELATION DATABASE SEGMENT

Date Reduction Form
Site: 24 Westbound in Lafayette between Moraga (Central Lafayette) and Acalanes
C:\SECTION\12_6\VIDEO.XLS
Film Date: 12/6/96

	Upstream			Downstream			Other Notes (e.g. make/model)		
ID	Lane	Time Stamp	Occ	Lane	Time Stamp	Occ	Type	Color	
800	2	12:12:03:23		2	12:13:07:20		P	BLUE	
801	1	12:12:04:03		2	12:13:04:17		C	LT BLUE	
802	3	12:12:05:03		3	12:13:14:18		P	BLUE	CAB
803	1	12:12:05:16		1	12:13:06:25	X	C	WH	
804	3	12:12:07:03		2	12:13:12:26		P	GRAY	
805	2	12:12:07:20		3	12:13:11:25		SUV	BLK	PATHFINDER
806	4	12:12:07:22		4	12:13:28:26		SEMI	YEL	2 AXLE
807	4	12:12:09:16		3>2	12:13:19:22		P	GRAY	CAB
808	OR	12:12:09:25		4	12:13:24:28		MV	WH	
809	3	12:12:10:23		3	12:13:16:02		P	WH	
810	2	12:12:13:27		2	12:13:15:14		C	WH	
811	3	12:12:15:01		4	12:13:14:29		P	BLK	EXTENDED CAB
812	OR	12:12:15:10		3	12:13:29:07	X	C	BLK	?
813	2	12:12:16:04		1	12:13:15:29		MV	BLUE	
814	4	12:12:17:12		3	12:13:27:26		T	ORG	CALTRANS
815	2	12:12:17:25		2	12:13:21:09		P	BR	RANGER
816	OR	12:12:18:24		4	12:13:33:15		SW	GRAY	
817	1	12:12:19:22		1	12:13:15:05		MV	LT GRAY	
818	4	12:12:20:07		3	12:13:31:15		SEMI	BLUE	LUMBER- 3 AXLES
819	3	12:12:22:23		2	12:13:26:10		C	WH	SAAB
820	OR	12:12:22:27		2>1	12:13:31:05		C	WH	
821	4	12:12:23:04		2	12:13:29:22		MV	BLUE	
822	3	12:12:24:16		1	12:13:28:01	X	C	WH	HONDA
823	2	12:12:27:08		2	12:13:28:02		C	BLK	
824	2	12:12:28:21		1	12:13:29:04		V	BLUE	
825	3	12:12:29:03		3	12:13:34:17		C	SILVER	

APPENDIX B. SAMPLE PVR FILE SEGMENT

```
Normal  12:12:22.984  1.71   821          4       27.6    6.71 SmallTruck
322    321
Normal  12:12:22.957 -0.03   702      2           31.4    4.28        Car
208    195
Normal  12:12:24.924  1.97  1346    1            35.7    4.72   SmallVan
209    208
Normal  12:12:25.824  0.90   488          4       26.2   10.71 RigidTruck
484    483
Normal  12:12:28.370  2.55  1346        3         31.5    4.48        Car
218    218
Normal  12:12:28.704  0.33   906          4       28.4    4.95   SmallVan
264    254
Normal  12:12:30.144  1.44  1097        3         31.3    4.48        Car
218    218
Normal  12:12:32.456  2.31  1267      2           32.3    4.12        Car
195    195
Normal  12:12:33.851  1.39   872      2           33.2    4.84   SmallVan
209    195
Normal  12:12:34.752  0.90  1832        3         30.9    4.24        Car
230    231
Normal  12:12:36.595  1.84  1072        3         29.2    4.48        Car
230    242
Normal  12:12:36.353 -0.24   720      2           30.0    4.49        Car
208    221
Normal  12:12:40.860  4.51   909    1             34.9    5.70   SmallVan
248    234
Normal  12:12:42.400  1.54  1590        3         31.5    5.24   SmallVan
252    252
Normal  12:12:42.254 -0.15  1290    1             36.9    4.61        Car
182    182
Normal  12:12:43.009  0.76  1474      2           31.4    4.52        Car
221    221
Normal  12:12:44.618  1.61  1517        3         31.6    4.04        Car
195    207
Normal  12:12:47.907  3.29  1177      2           30.7    4.59        Car
222    221
Normal  12:12:48.720  0.81   888        3         32.4    4.43        Car
205    206
Normal  12:12:49.109  0.39  1063          4       29.6    4.79   SmallVan
228    240
Normal  12:12:53.442  4.33  1556      2           31.4    4.53        Car
232    232
Normal  12:12:54.702  1.26   764        3         31.3    6.25 SmallTruck
274    275
Normal  12:12:55.086  0.38  1212    1             32.9    4.41        Car
207    194
```

APPENDIX C. SAMPLE SIGNATURE FILE SEGMENT

```
# Record 2847, lane 4   814
# Normal  12:12:22.984  0.00   821        4       27.6    6.71 SmallTruck
2664.984        25
2664.995        56
2665.007        88
2665.018        216
2665.030        334
2665.041        463
2665.053        581
2665.064        706
2665.076        769
2665.087        785
2665.099        778
2665.110        790
2665.122        801
2665.133        813
2665.145        817
2665.156        800
2665.168        784
2665.179        767
2665.191        752
2665.202        735
2665.214        719
2665.225        658
2665.237        520
2665.248        370
2665.260        233
2665.271        109
2665.283        53
2665.294        25
# Record 2848, lane 2   815
# Normal  12:12:22.957  0.00   702      2         31.4    4.28       Car
2664.957        29
2664.970        131
2664.983        234
2664.996        379
2665.009        524
2665.022        669
2665.035        702
2665.048        680
```

2665.061 605
2665.074 529
2665.087 433
2665.100 337
2665.113 211
2665.126 84
2665.139 30
Record 2849, lane 1 817
Normal 12:12:24.924 0.00 1346 1 35.7 4.72 SmallVan
2666.924 77
2666.937 119
2666.950 334
2666.963 695
2666.976 1056
2666.989 1170
2667.002 1222
2667.015 1346
2667.028 1202
2667.041 1130
2667.054 947
2667.067 764
2667.080 443
2667.093 267
2667.106 91
2667.119 30
Record 2850, lane 4 818
Normal 12:12:25.824 0.00 488 4 26.2 10.71 RigidTruck
2667.824 27
2667.836 55
2667.847 85
2667.859 114
2667.870 182
2667.882 245
2667.893 310
2667.905 324
2667.916 341
2667.928 383
2667.939 429
2667.951 467
2667.962 478
2667.974 487
2667.985 472
2667.997 454

2668.008	402
2668.020	354
2668.031	302
2668.043	256
2668.054	225
2668.066	196
2668.077	170
2668.089	165
2668.100	160
2668.112	155
2668.124	150
2668.135	146
2668.147	150
2668.158	154
2668.170	158
2668.181	167
2668.193	193
2668.204	221
2668.216	247
2668.227	261
2668.239	234
2668.250	196
2668.262	138
2668.273	75
2668.285	30
2668.296	18

APPENDIX D. ESTIMATED TRAVEL TIME RESULTS FOR MODERATE FLOW DATA

	Training		Test	
Time	True	Section	True	Section
15	62.10	55.30	63.60	63.00
30	60.20	60.30	60.20	62.10
45	61.90	61.80	59.40	59.10
60	58.50	58.50	60.50	59.50
75	64.00	63.90	60.20	55.80
90	62.90	64.40	58.80	63.00
105	64.90	63.60	66.80	66.90
120	61.10	55.90	63.00	62.70
135	58.70	58.70	60.40	56.30
150	60.90	60.00	59.30	59.70
165	63.70	63.70	60.70	60.40
180	60.60	61.00	63.30	64.00
195	62.50	63.10	61.80	63.00
210	63.90	58.90	62.60	63.10
225	62.40	57.80	59.40	55.10
240	61.10	60.40	64.10	63.90
255	61.30	61.20	64.60	65.70
270	61.50	62.30	61.40	61.70
285	58.60	58.60	60.40	62.10
300	64.70	64.60	60.30	60.30
315	60.40	52.70	57.60	54.70
330	58.80	55.00	61.90	61.90
345	60.20	61.30	58.90	59.30
360	59.70	59.50	61.30	62.10
375	61.30	60.80	63.60	61.80
390	63.00	59.70	64.30	64.60
405	63.80	63.80	59.20	56.90
420	60.50	61.50	60.60	60.60
435	62.70	63.40	63.20	61.00
450	61.00	61.50	61.80	63.00
465	60.20	59.70	64.30	63.30
480	59.60	59.50	66.00	66.90
495	62.70	64.40	61.70	60.80
510	62.70	62.40	62.00	62.30
525	61.90	63.10	62.80	62.80
540	61.70	61.50	61.90	61.00
555	63.50	63.30	62.90	62.50

570	62.90	63.30	60.80	60.20
585	59.30	58.40	63.10	62.70
600	57.90	57.10	62.30	63.40
615	59.90	59.60	60.80	61.10
630	61.40	61.80	58.40	59.10
645	59.90	62.10	58.40	60.10
660	60.60	56.10	57.50	57.40
675	62.30	62.80	64.60	65.80
690	59.30	59.80	60.90	61.10
705	60.60	61.30	62.10	62.60
720	63.10	63.10	61.00	61.20
735	60.30	54.10	59.60	58.50
750	62.60	63.60	61.20	62.40
765	58.80	58.20	61.00	56.60
780	59.60	60.60	62.40	62.90
795	61.60	61.90	59.00	54.00
810	62.30	58.90	62.30	62.40
825	60.00	59.60	62.60	62.10
840	59.10	59.20	66.60	66.80
855	63.30	56.60	61.60	61.31
870	61.80	62.80		
885	59.20	60.10		
900	63.40	64.40		