PB2002-100535

Neural Network Model for Automatic Traffic Incident Detection

by

Hojjat Adeli

Knowledge Engineering Lab
Department of Civil and Environmental Engineering
and Geodetic Science
The Ohio State University

Sponsored by
Ohio Department of Transportation
and
Federal Highway Administration

REPRODUCED BY: NTIS
U.S. Department of Commerce
National Technical Information Service
Springfield, Virginia 22161

Neural Network Model for Automatic Traffic Incident Detection

Principal Investigator

Hojjat Adeli
Professor

Knowledge Engineering Lab
Department of Civil and Environmental Engineering
and Geodetic Science
The Ohio State University

August 2001

Sponsored by
Ohio Department of Transportation
and
Federal Highway Administration

Prepared in Cooperation with the Ohio Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration

"The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do no necessarily reflect the official views or policies of the Ohio Department of Transportation or the Federal Highway Administration. This report does not constitute a standard, specification or regulation."

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/HWY-03/2002		

4. Title and SubtitleL	5. Report Date
Neural Network Model for Automatic Traffic Incident Detection	August, 2001
	6. Performing Organization Code
	8. Performing Organization Report No.
7. Author(s)	10. Work Unit No. (TRAIS)
Hojjat Adeli	
9. Performing Organization Name and Address Department of Civil and Environmental Engineering & Geodetic Science The Ohio State University 470 Hitchcock Hall 2070 Neil Avenue Columbus, OH 43210-1275	11. Contract or Grant No. State Job No. 14690(0)
	13. Type of Report and Period Covered Final Report
12. Sponsoring Agency Name and Address Ohio Department of Transportation 1980 W Broad Street Columbus, OH 43223	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Automatic freeway incident detection is an important component of advanced transportation management systems (ATMS) that provides information for emergency relief and traffic control and management purposes. In this research, a multi-paradigm intelligent system approach and several innovative algorithms were developed for solution of the freeway traffic incident detection problem employing advanced signal processing, pattern recognition, and classification techniques. The methodology effectively integrates fuzzy, wavelet, and neural computing techniques to improve reliability and robustness. The specific accomplishments of this research are a) Development of an effective traffic feature extraction model using discrete wavelet transform and linear discriminant analysis, b) Development of a computational model for automatic traffic incident detection using discrete wavelet transform, linear discriminant analysis, and adaptive conjugate gradient neural network of Adeli and Hung, c) Development of a fuzzy wavelet radial basis function neural network (RBFNN) model for automatic detection of freeway incidents, d) Development of a two-stage single-station freeway incident detection model based on energy representation of the traffic pattern in the wavelet domain, e) A comprehensive parametric study of the performance of the single-station fuzzy-wavelet RBFNN freeway incident detection model and comparison with the benchmark California algorithm #8 based on three quantitative measures of detection rate, false alarm rate, and detection time, and the qualitative measure of algorithm portability using both real and simulated data, f) A comprehensive evaluation of the single-station wavelet energy neural network freeway incident detection algorithm and comparison with the California algorithm #8, and g) Evaluation of the wavelet energy neural network freeway incident detection algorithm on rural freeways where flow rates are low and detector stations are spaced further apart. It is demonstrated that both fuzzy-wavelet RBFNN and wavelet energy neural network freeway incident detection algorithms are computationally efficient, produce excellent detection rates and very low false alarm rates on urban freeways, and can readily be implemented on-line in any ATMS without any need for re-calibration and without any performance deterioration. Considering the difficulty in automatic detection of incidents on rural freeways, the wavelet energy algorithm performs well on rural freeways as well. The algorithm is fast as it detects an incident on urban freeways in less than two minutes and on rural freeways in less than three minutes.

17. Key Words	18. Distribution Statement
advanced transportation management systems, California algorithm, freeway incident detection, fuzzy logic, highway, intelligent transportation systems, ITS, neural networks, rural freeways, signal processing, traffic engineering, urban freeway, wavelets.	No Restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

Neural Network Model for Automatic Traffic Incident Detection

Principal Investigator: Hojjat Adeli, Professor, The Ohio State University

Executive Summary

Automatic freeway incident detection is an important component of advanced transportation management systems (ATMS) that provides information for emergency relief and traffic control and management purposes. Earlier algorithms for the freeway incident problems have produced unreliable results especially in recurrent congestion and compression wave traffic conditions. In this research, a multi-paradigm intelligent system approach and several innovative algorithms were developed for solution of the freeway traffic incident detection problem employing advanced signal processing, pattern recognition, and classification techniques. The methodology effectively integrates fuzzy, wavelet, and neural computing techniques to improve reliability and robustness. The specific accomplishments of this research are

- Development of an effective traffic feature extraction model using discrete wavelet transform and linear discriminant analysis.
- Development of a computational model for automatic traffic incident detection using discrete wavelet transform, linear discriminant analysis, and adaptive conjugate gradient neural network of Adeli and Hung.
- Development of a fuzzy wavelet radial basis function neural network (RBFNN) model for automatic detection of freeway incidents.
- Development of a two-stage single-station freeway incident detection model based on energy representation of the traffic pattern in the wavelet domain.
- A comprehensive parametric study of the performance of the single-station fuzzy-wavelet RBFNN freeway incident detection model and comparison with the benchmark California algorithm #8 based on three quantitative measures of detection rate, false alarm rate, and detection time, and the qualitative measure of algorithm portability using both real and simulated data. The new algorithm outperformed the California algorithm consistently under various scenarios. False alarms are a major hindrance to the widespread implementation of automatic freeway incident detection algorithms. The false alarm rate ranges from 0 to 0.07 % for the new algorithm and 0.53 to 3.82% for the California algorithm.
- A comprehensive evaluation of the single-station wavelet energy neural network freeway incident detection algorithm and comparison with the California algorithm #8.
- Evaluation of the wavelet energy neural network freeway incident detection algorithm on rural freeways where flow rates are low and detector stations are spaced further apart.

It is demonstrated that both fuzzy-wavelet RBFNN and wavelet energy neural network freeway incident detection algorithms are computationally efficient, produce excellent detection rates and very low false alarm rates on urban freeways, and can readily be implemented on-line in any ATMS without any need for re-calibration and without any performance deterioration. Considering the difficulty in automatic detection of incidents on rural freeways, the wavelet energy algorithm performs well on rural freeways as well. The algorithm is fast as it detects an incident on urban freeways in less than two minutes and on rural freeways in less than three minutes.

Summary and Organization of the Report

This report consists of seven parts presented as seven different manuscripts. Each manuscript is summarized in the following paragraphs. Automatic freeway incident detection is an important component of advanced transportation management systems that provides information for emergency relief and traffic control and management purposes. Earlier algorithms for the freeway incident problems have produced unreliable results especially in recurrent congestion and compression wave traffic conditions. Traffic incidents are non-recurrent and pseudo-random events that disrupt the normal flow of traffic and create a bottleneck in the road network. The probability of incidents is higher during peak flow rates when their system wide impact is most severe. Model-based solutions to the incident detection problem have not produced practically useful results primarily because the complexity of the problem does not lend itself to accurate mathematical and knowledge-based representations.

To eliminate false alarms an effective traffic incident detection algorithm must be able to extract incident related features from the traffic patterns. A robust feature extraction algorithm also helps reduce the dimension of the input space for a neural network model without any significant loss of related traffic information, resulting in a substantial reduction in the network size, the effect of random traffic fluctuation, the number of required training samples, and the computational resources required to train the neural network. In Part 1, an effective traffic feature extraction model is presented using discrete wavelet transform (DWT) and linear discriminant analysis (LDA). The DWT is first applied to raw traffic data and the finest resolution coefficients representing the random fluctuations of traffic are discarded. Next, LDA is employed to the filtered signal for further feature extraction and reducing the dimensionality of the problem. The results of LDA are used as input to a neural network model for traffic incident detection.

Artificial neural networks are known to be effective in solving problems involving pattern recognition and classification. The traffic incident detection problem can be viewed as recognizing incident patterns from the incident-free patterns. A neural network classifier has to be trained first using incident and incident-free traffic data. The dimensionality of the training input data is high and the embedded incident characteristics are not easily detectable. In Part 2, a computational model is presented for automatic traffic incident detection using discrete wavelet transform, linear discriminant analysis, and neural networks. Wavelet transform and linear discriminant analysis are used for feature extraction, de-noising, and effective preprocessing of data before an adaptive neural network model is used to make the traffic incident detection. Simulated as well as actual traffic data are used to test the model. For incidents with duration of more than five minutes, the incident detection model yields a detection rate of nearly 100% and false alarm rate of about 1% for two- or three-lane freeways.

Researchers have presented freeway traffic incident detection algorithms by combining the adaptive learning capability of neural networks with imprecision modeling capability of fuzzy logic. In Part 3, it is shown that the performance of a fuzzy neural network algorithm can be improved through preprocessing of data using a wavelet based feature extraction model. In particular, the discrete wavelet transform de-noising and feature extraction model presented in Part 1 is combined with the fuzzy-neural network approach presented by Hsiao et al. (1994). It is shown that substantial improvement can be achieved using the data filtered by DWT. Use of the wavelet theory to de-noise the traffic data increases the incident detection rate, reduces the false alarm rate and the incident detection time, and improves the convergence of the neural network training algorithm substantially.

In Part 4, a multi-paradigm intelligent system approach is presented for the solution of the freeway traffic incident detection problem employing advanced signal processing, pattern recognition, and classification techniques. The methodology effectively integrates fuzzy, wavelet, and neural computing techniques to improve reliability and robustness. A wavelet-based de-noising technique is employed to eliminate undesirable fluctuations in observed data from traffic sensors. Fuzzy c-mean clustering is used to extract significant information from the observed data and to reduce its dimensionality. A radial basis function neural network is developed to classify the de-noised and clustered observed data. The new model produced excellent incident detection rates with no false alarms when tested using both real and simulated data.

In Part 5, a two-stage single-station freeway incident detection model is presented based on advanced wavelet analysis and pattern recognition techniques. Wavelet analysis is used to de-noise, cluster, and enhance the raw traffic data, which is then classified by a radial basis function neural network. An energy representation of the traffic pattern in the wavelet domain is found to best characterize incident and non-incident traffic conditions. False alarm during recurrent congestion and compression waves is eliminated by normalization of a sufficiently long time-series pattern. The model is tested under several traffic flow scenarios including compression wave conditions. It produced excellent detection and false alarms characteristics. The model is computationally efficient and can readily be implemented on-line in any ATMS without any need for re-calibration.

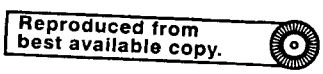
In Part 6, the performance of the fuzzy-wavelet radial basis function neural network (RBFNN) freeway incident detection model presented in Part 4 is evaluated and compared with the benchmark California algorithm #8 using both real and simulated data. The evaluation is

based on three quantitative measures of detection rate, false alarm rate, and detection time, and the qualitative measure of algorithm portability. The new algorithm outperformed the California algorithm consistently under various scenarios. False alarms are a major hindrance to the widespread implementation of automatic freeway incident detection algorithms. The false alarm rate ranges from 0 to 0.07 % for the new algorithm and 0.53 to 3.82% for the California algorithm. The new fuzzy-wavelet RBFNN freeway incident detection model is a single-station pattern-based algorithm that is computationally efficient and requires no re-calibration. The new model can be readily transferred without re-training and without any performance deterioration.

In Part 7, a comprehensive evaluation of the single-station wavelet energy neural network freeway incident detection algorithm of is presented. Quantitative performance measures of detection rate, false alarm rate, and detection time as well as the qualitative measure of portability are investigated for both urban and rural freeway conditions. Further, the performance of the algorithm is compared with that of the California algorithm #8. This research demonstrates the portability of the wavelet energy algorithm and its excellent performance for urban freeways across a wide range of traffic flow and roadway geometry conditions regardless of the density of the loop detectors. Rural freeways present additional challenges in that flow rates are low and detector stations are spaced further apart. Considering the difficulty in automatic detection of incidents on rural freeways, the new wavelet energy algorithm performs well on such freeways. The algorithm is fast as it detects an incident on urban freeways in less than two minutes and on rural freeways in less than three minutes.

Part 1

PROTECTED UNDER INTERNATIONAL COPYRIGHT
ALL RIGHTS RESERVED
NATIONAL TECHNICAL INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE



FEATURE EXTRACTION FOR TRAFFIC INCIDENT DETECTION USING WAVELET TRANSFORM AND LINEAR DISCRIMINANT ANALYSIS

A. Samant[1] and H. Adeli[2]

[1]Graduate Research Associate, [2] Professor

Dept. of Civil and Environmental Engineering and Geodetic Science

The Ohio State University, 470 Hitchcock Hall

2070 Neil Avenue, Columbus, OH 43210.

ABSTRACT: To eliminate false alarms an effective traffic incident detection algorithm must be able to extract incident related features from the traffic patterns. A robust feature extraction algorithm also helps reduce the dimension of the input space for a neural network model without any significant loss of related traffic information, resulting in a substantial reduction in the network size, the effect of random traffic fluctuation, the number of required training samples, and the computational resources required to train the neural network. This article presents an effective traffic feature extraction model using discrete wavelet transform (DWT) and linear discriminant analysis (LDA). The DWT is first applied to raw traffic data and the finest resolution coefficients representing the random fluctuations of traffic are discarded. Next, LDA is employed to the filtered signal for further feature extraction and reducing the dimensionality of the problem. The results of LDA are used as input to a neural network model for traffic incident detection.

1. INTRODUCTION

Reliable automatic detection of traffic incidents is required for efficient traffic management on freeways. Travel time delays occur due to lane blockages and the corresponding reduction in the capacity of the freeway following the incident. Our research goal is to create computational models which take into account the traffic flow variations and detect the traffic incidents automatically, by distinguishing the traffic incident patterns from the incident-free ones. Since there are lots of traffic fluctuations in the traffic flow for various reasons the incident and incident-free decision regions cannot be divided easily. This is the main cause for the poor performance of the existing traffic incident detection algorithms.

Until early 90s, the two-station comparative algorithms such as California algorithm[11] were widely used, where the differences between traffic flow parameters (e.g. traffic volume and occupancy) at upstream and downstream stations are used for the detection of the operating problems in the traffic flow. Persaud et al.[12] proposed a single station algorithm known as McMaster algorithm, where congestion is detected using traffic volume, occupancy, and vehicle speed (if available) values at a single station. Though these computational models were easy to implement, they could not achieve the desired level of accuracy. Consequently, new approaches such as artificial neural networks[3,16] and fuzzy logic[7] have been investigated to improve the performance. Research also has been carried out to filter out the random fluctuations of the traffic using moving average or median plus average methods[15] in an attempt to minimize the occurrence of false alarm (i.e., false detection of incidents) but with limited success.

Most of the existing incident detection algorithms based on conventional statistical methods compare the traffic flow values or the differences between the values at various locations with some fixed threshold values to recognize congested traffic patterns from non-congested ones. But this single threshold value may not represent the traffic flow accurately because of the random fluctuations and the time-dependent nature of the freeway traffic. This is the major cause of the unreliability in such incident detection algorithms.

To eliminate the false alarms an effective incident detection algorithm must be able to extract features from the traffic patterns, which are related to the incident. In this work we use the discrete wavelet transform (DWT) and Linear Discriminant Analysis (LDA) for feature extraction. Actual traffic data obtained from the sensors on the freeways are not well suited as a direct input to a neural network model to be used to detect incidents. The dimensionality of the training input data is generally high as various traffic parameters (e.g. traffic volume and occupancy) at different locations (e.g. upstream and downstream of the various incident locations) and at many instances of time are required to be inputted, and the embedded incident characteristics may not be easily detectable. Also, the training of a neural network incident detection algorithm requires input patterns containing sufficient incident data. Thus, effective pre-processing of the sensory data is essential before they can be used in a neural network model. In this work, we perform feature extraction in two steps. In the first step, the data is filtered and the high frequency signals representing noise, which may not be related to an incident, are removed using wavelet transform. In the second step, the features are enhanced using LDA. The feature extraction algorithm also helps reduce the dimensionality of the input

space to a neural network model without any significant loss of related traffic information. In the companion paper we use the feature extraction algorithm to develop a robust traffic incident detection model[1].

2. DISCRETE WAVELET TRANSFORM

The wavelet transform is found to be an effective tool in signal and image processing due to its attractive properties such as time-frequency localization (obtaining a signal at particular time or frequency), multi-rate filtering (differentiating the signals having various frequencies), scale-space analysis (extracting features at various locations in space at different scales), and multi-resolution analysis[4,8,9]. Using these properties one can extract the desired features from an input signal characterized by certain local properties in time and space. In this research, we view the traffic flow as a signal, with traffic incidents as well as other traffic patterns such as traffic bottleneck or compression wave having different time-space properties. Most of the previous incident detection algorithms performed unsatisfactorily because they can not distinguish the traffic incident patterns from other similar traffic patterns such as recurrent congestion, and specially the compression wave, consistently. We use the wavelet transform to extract the specific features distinguishing such traffic patterns as it can extract features from different time scales having different resolutions quite effectively.

For the traffic incident detection problem, we consider various traffic data (e.g. traffic volumes and occupancies at various locations) recorded at a fixed time interval (e.g. 20-30 seconds). Each of these data series can be represented by x[j], where $j \in Z$ and Z is a set of integers (square brackets represent a series, a sequence or a vector and

circular brackets represent functions). The vector space of square summable sequences is defined as follows:

$$L^2(Z) = \left(\gamma_j\right)_{j\in Z} : \sum_{j=-\infty}^{j=+\infty} (\gamma_j)^2 < \infty \tag{1}$$

where γ_j represents a sequence of real numbers and Z is the set of all the integers. That means the inner product of a sequence with itself converges to a finite value. We denote the orthonormal wavelet bases of $L^2(Z)$ by $\{\phi_{I,l}\}_{l\in Z}$ and $\{\psi_{j,l}\}_{l\in Z}$ (the brackets {} denote a set of series) where $\{\phi_{I,l}\}_{l\in Z}$ and $\{\psi_{j,l}\}_{l\in Z}$ represent scaling and wavelet functions[4,5], respectively and I is a positive integer. The value of I is chosen such that the desired level of resolution is obtained and j = 1, 2,..., I. The output of the DWT consists of the coordinates, $\beta_I[l]$ and $\lambda_j[l]$ of the orthonormal wavelet bases

$$\beta_I[l] = \left\langle x[k],\, \phi_{I,l}[k] \right\rangle \quad \text{and} \quad l = \frac{k}{2^I} \tag{2}$$

$$\lambda_j[l] = \left\langle x[k],\, \psi_{j,l}[k] \right\rangle \;;\; l = \frac{k}{2^j} \;\text{ and }\; j = 1, 2, \ldots, I \tag{3}$$

where $\langle\ \rangle$ denotes the inner product of the two sequences in the vector space $L^2(Z)$, k represents the total number of input data points, l represents the number of coefficients of each data series such as traffic volume or occupancy. The coordinates β and λ are in fact, low and high-resolution coefficients of the given data series $x[k]$, respectively. The inner product of any two data series f[n] and g[n] is calculated as follows:

$$\left\langle f[n],\, g[n] \right\rangle = \sum_{i=-\infty}^{i=+\infty} f[i]\, g[i] \tag{4}$$

In our traffic incident detection problem, we use 8-minute traffic patterns with data recorded in intervals of 30 seconds as multi-resolution analysis using DWT requires at least 16 data points at a time. As such, in Eqs. (2) and (3) k=16. We choose, I = 2, which means the traffic patterns are divided into three types of *signals*: low-resolution (β_2), medium-resolution (λ_2), and high-resolution (λ_1). In this case l in Eq. (2) = $\frac{16}{2^I}$ = 4. Consequently, we have 4 low-resolution coefficient ($\beta_2[l]$). Similarly l in Eq. (3) is equal to 8 for j=1, and equal to 4 for j=2, which yields us 8 fine-resolution coefficients ($\lambda_1[l]$) and 4 medium resolution coefficients ($\lambda_2[l]$).

The coordinates of the wavelet bases (βs and λs) are computed using a concept called the quadrature mirror filters[18]. Quadrature filter is an operator that performs signal convolution and downsampling[17]. We use mirror filters (a pair of filters) so that the original traffic signal can be reconstructed without any loss of related information (One filter yields the high-resolution components of the signal and the other filter yields the low-resolution components.) The convolution of any two sequences f[n] and g[n] is calculated as follows:

$$(f * g)[n] = \big(f[m] * g[m]\big)[n] = \sum_{i=-\infty}^{i=+\infty} f[i]\, g[n-i] \qquad (5)$$

The downsampling part is discussed in the following section.

To extract the traffic incident pattern from the traffic data we perform multi-resolution analysis of the wavelet transforms of traffic patterns. Multi-resolution analysis involves dividing the original *signal* (e.g. traffic volume or occupancy) into *signals*

having different frequencies and time localizations and analyzing the signal in different scales.

To carry out a multi-resolution analysis of a traffic pattern we need to define a two-dimensional set of scaling functions, $\phi(t)$, and wavelet functions, $\psi(t)$. A two-dimensional family of scaling functions is obtained by scaling and translating the basic scaling function $\phi(t)$ as follows:

$$\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t-k) \text{ where } j,k \in Z \text{ and } \phi(t) \in L^2(Z) \quad (6)$$

where t is an integer representing the number of time intervals (such as 30-sec. intervals). Integers j and k are called scaling and translation parameters respectively. The corresponding subspaces spanned by $\phi_{j,k}(t)$ are[8]:

$$V_j = \overline{\underset{k}{Span}\{\phi_{j,k}(t)\}} \qquad \text{for all } k \in Z \quad (7)$$

The over-bar indicates that V_j is a closed subspace (i.e. boundaries are included in the subspace). Equation (7) means any function $f(t) \in V_j$ can be represented as a weighted sum of the scaling functions with scale j as follows:

$$f(t) = \sum_k \gamma_k \phi_k(2^j t - k) \qquad \text{for any} \qquad f(t) \in V_j \quad (8)$$

The scale j can be varied from $-\infty$ to $+\infty$ to obtain signals having various resolutions.

Similar to the scaling functions a two-dimensional family of wavelet functions is obtained from the mother wavelet $\psi(t)$ by scaling and translation as follows:

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \quad (9)$$

The corresponding subspaces spanned by wavelets $\psi_{j,k}(t)$ are

$$W_j = \overline{\underset{k}{Span}\{\psi_{j,k}(t)\}} \qquad \text{for all } k \in Z \tag{10}$$

3. MULTI-RESOLUTION ANALYSIS

A *Multi-resolution Analysis* in $L^2(Z)$ consists of finding out wavelet transforms using a sequence of closed subspaces V_j in $L^2(Z)$ with the following properties[13]:

$$V_{j+1} \subset V_j \qquad \text{for all} \quad j \in \mathbf{Z} \tag{11}$$

$$V_\infty = \{0\} \qquad \text{(indicating the empty space)} \tag{12}$$

$$V_0 = L^2(Z) \qquad \text{(contains the original input signal)} \tag{13}$$

Equation (11) indicates that V_{j+1} is a subset of V_j. The subspace V_j contains all the signals included in V_{j+1} plus additional high-resolution signals. These additional high-resolution signals are contained in the wavelet-spanned subspace W_{j+1}:

$$V_j = V_{j+1} \oplus W_{j+1} \tag{14}$$

where $\oplus$ indicates that both subspaces V_{j+1} and W_{j+1} are part of V_j and orthogonal to each other[8,13]. For $j = 0$ in Eq. (14) we obtain

$$V_0 = V_1 \oplus W_1 \tag{15}$$

and by combining Eqs. (11), (14) and (15) the sequence can be generalized as

$$V_0 = V_I \oplus W_I \oplus W_{I-1} \oplus W_{I-2} \ldots . \oplus W_2 \oplus W_1 \tag{16}$$

The value of I can be varied to obtain the desired level of resolution. We choose $I = 2$, for the same reasons explained for Eqs. (2) and (3). In that case, the original signal is divided into three parts, each one lying in a different subspace as follows:

$$V_0 = V_2 \oplus W_2 \oplus W_1 \tag{17}$$

where V_2, W_2, and W_1 contain the low, medium and high-resolution signals, respectively. The definition of V_j and the scaling condition given by Eq. (6) ensure that elements in the two consecutive subspaces V_j and V_{j+1} are inter-related as follows:

$$f(t) \in V_j \quad \Leftrightarrow \quad f\left(\frac{t}{2}\right) \in V_{j+1} \tag{18}$$

where the notation $\Leftrightarrow$ indicates mutual implication. The actual relationship is expressed as follows[2]:

$$\phi[2^{j+1}t] = \frac{1}{\sqrt{2}} \sum_n h[n] \ \phi[2^j t - n], \qquad n \in Z \tag{19}$$

where $h[n]$ is a sequence of real numbers known as the scaling function coefficients or low-pass filter coefficients. ($\frac{1}{\sqrt{2}}$ keeps the norm of the scaling function equal to 1).

Since the wavelet-spanned subspace at scale j+1 is a part of V_j (the subspace spanned by the scaling function with scale j, i.e. $W_j \subset V_{j+1}$) the wavelets at scale j+1 can be represented in terms of scalar-multiples of the translated scaling functions at scale j as follows:

$$\psi(2^{j+1}t) = \frac{1}{\sqrt{2}} \sum_n h_1[n] \ \phi[2^j t - n], \qquad n \in Z \tag{20}$$

where $h_1[n]$ is a set of real numbers known as wavelet function coefficients or high-pass filter coefficients. Due to the orthogonal relationship between the wavelet and scaling functions, the wavelet coefficients are related to scaling coefficients as follows[2]:

$$h_1[n] = (-1)^n h[L-1-n] \tag{21}$$

where L is the length of the filter used. In our traffic incident detection case, we use length-4 Daubechies filter coefficients ($L = 4$) as it is found to be accurate and efficient in the area of digital filtering. For this filter, $h[n]$ values are found by solving the recursion equation (19) by the zero wavelet moment design approach[2]:

$$h[n] = \frac{1}{4\sqrt{2}} \left[1+\sqrt{3},\ 3+\sqrt{3},\ 3-\sqrt{3},\ 1-\sqrt{3}\right] \quad \text{for } n=0,1,2,3 \tag{22}$$

Substituting $L = 4$ in Eq. (21) we obtain the corresponding $h_1[n]$ values as follows:

$$h_1[n] = \frac{1}{4\sqrt{2}} \left[1-\sqrt{3},\ -(3-\sqrt{3}),\ 3+\sqrt{3},\ -(1+\sqrt{3})\right] \quad \text{for } n=0,1,2,3 \tag{23}$$

Now we can write any input data series $f[t]$ in V_0 (or $L^2(Z)$) as a series expansion in terms of the scaling functions and the wavelets[8]:

$$f[t] = \sum_{k=-\infty}^{k=\infty} \beta_I[k]\varphi_k[t] + \sum_{j=0}^{I} \sum_{k=-\infty}^{k=\infty} \lambda_j[k]\, \psi_{j,k}[t] \tag{24}$$

In this work, we use discrete data points and not a continuous signal, consequently we do not have to deal with the scaling functions or wavelets directly. Only the coefficients $h[n]$ and $h_1[n]$ in the defining Eqs. (19) and (20) and $\beta[k]$ and $\lambda_j[k]$ in the expansion Eq. (24) need to be considered. The first two sets of coefficients can be

viewed as digital filters (low and high-pass filters respectively) and the last two sets of coefficients can be viewed as digital signals (low and high-resolution coefficients, respectively). In order to use the wavelet transform coefficients directly, the relationship between the expansion coefficients at two consecutive scale levels should be known. By scaling and translating the basic recursion Eq. (6), the required relationship is found for the scaling and wavelet coefficients as follows[2]:

$$\beta_{j+1}[k] = \sum_{m=0}^{L} h[m-2k]\beta_j[m] \tag{25}$$

$$\lambda_{j+1}[k] = \sum_{m=0}^{L} h_1[m-2k]\beta_j[m] \tag{26}$$

4. FILTERING AND DOWNSAMPLING

Digital filtering of the input signal is carried out by convoluting the signal with another set of numbers known as the filter coefficients or impulse responses[9] and the downsampling process involving decimation of some of the input data. In downsampling, the input signal x(n) is transformed into an output signal y(n) such that y(n) = x(2n). This means the alternate data points are discarded as shown in Figure 1 schematically.

Equations (25) and (26) in fact perform digital filtering and downsampling. These equations show that the scaling and wavelet coefficients at different levels of scale can be obtained by convoluting the expansion coefficients at scale j with the filter coefficients $h_0[n]$ and $h_1[n]$ and then downsampling to obtain the expansion coefficients at the next level $j+1$. In other words, the j scale coefficients are filtered by two so-called FIR (Finite Impulse Response)[9] digital filters with coefficients $h_0[n]$ and $h_1[n]$. After filtering and downsampling are completed the next low-resolution scaling and high-resolution wavelet

coefficients are found. This is shown schematically in Figure 2, where H_0 and H_1 represent the two FIR filters. This splitting (dividing of signal into higher and lower resolution signals), filtering and decimation (downsampling) can be repeated on the scaling coefficients to obtain a two- or three-stage two-scale filter (Figure 3).

Having found the relationship among the four sets of coefficients, we now describe how to obtain the input set of scaling coefficients (β_0) from the input signal. In the traffic incident detection model the traffic data are not continuous. We use the traffic volume and occupancy values at 30 second intervals which means the data are pre-filtered and can be used directly as input coefficients. As an example if we use 8-minute traffic patterns, we will have 16 input values for each of the four input parameters: upstream and downstream occupancy and volume. After two stages of downsampling and filtering we will have 8 coefficients of the finest resolution, 4 coefficients of the medium resolution and 4 coefficients of the coarse resolution for each traffic parameter. All 8 high-resolution coefficients are discarded as they represent the ordinary traffic fluctuations, which may not be related to the traffic incidents. For both low as well as medium resolution coefficients, we will take some or all of them, and find out the best combination. The signal is then re-generated using these medium and low-resolution coefficients, which is called de-noised signal. To enhance the feature extraction, linear discriminant analysis is performed on the coefficients obtained from the wavelet transform and multi-resolution analysis. Linear discriminant analysis is discussed in the next section.

5. LINEAR DISCRIMINANT ANALYSIS

We use a linear discriminant analysis to reduce the dimensionality of the problem as well as to improve the generalization capability of the pattern classifier while at the same time reducing its computational processing requirements. This part of feature extraction can be formulated as a mapping from a d-dimensional input space (R^d) to an m-dimensional feature space (R^m) through a transformation matrix $\mathbf{T}$[6]:

$$\mathbf{T} : R^d \rightarrow R^m, \quad m < d \tag{27}$$

The linear discriminant analysis achieves feature extraction through linear mapping of the input space to the output space. The most popular and commonly used linear discriminant classifiers are Fisher Linear Discriminant Classifier (FLD) and Nearest Mean Classifier (NMC) or Euclidean Distance Classifier. The construction procedure for both classifiers is almost the same with minor differences in the end[10,14].

Let $(\mathbf{X_i})_j$ be a vector representing the i^{th} training sample outputted by the discrete wavelet transform in class j. If we use 6 of the 8 medium and low resolution coefficients, as an example then, $(\mathbf{X_i})_j = (x_{i,j}^1, x_{i,j}^2, ..., x_{i,j}^6)$. In the traffic incident detection case j will be either 1 or 2 where $j = 1$ indicates the incident free samples and $j = 2$, indicates the incident samples. Also, $i = 1, 2, \ldots, n_j$, where n_j = number of training samples in class j, and $n = n_1 + n_2$, the total number of training samples. The within-class co-variance square matrix $\mathbf{C}_w$ of dimension d is defined as

$$\mathbf{C}_w = \frac{1}{n} \sum_{j=1}^{2} \sum_{i=1}^{n_j} ((\mathbf{X_i})_j - \mathbf{m}_j)\ ((\mathbf{X_i})_j - \mathbf{m}_j)^T \tag{28}$$

where $\mathbf{m}_j$ is the mean vector for class *j*. The incident detection is a two-class problem involving classification of data between incident and incident-free regions. For this two-class problem, the between-class covariance square matrix, $\mathbf{C}_B$, of dimension d is defined as[6]:

$$\mathbf{C}_B = \frac{n_1}{n}(\mathbf{m}_1 - \mathbf{m})(\mathbf{m}_1 - \mathbf{m})^T + \frac{n_2}{n}(\mathbf{m}_2 - \mathbf{m})(\mathbf{m}_2 - \mathbf{m})^T \qquad (29)$$

where $\mathbf{m}$ is the mean vector of all the data. The goal of linear discriminant analysis is to find a dxm transformation matrix $\mathbf{T}$ such that the within-class scatter is minimized and the between class scatter is maximized. This can be achieved by maximizing the sum of the eigenvalues (J) of the multiplication matrix $\mathbf{C}_W^{-1}\mathbf{C}_B$ [6]. Simplifying Eq. (29) we obtain

$$\mathbf{C}_B = \frac{n_1 n_2}{n^2}(\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T \qquad (30)$$

Since $\mathbf{C}_B$ is a function of only one vector $(\mathbf{m}_2 - \mathbf{m}_1)$, its rank (number of independent rows or columns in the matrix) is one. And since $\mathbf{C}_W$ has a full rank its inverse exists and the rank of $\mathbf{C}_W^{-1}\mathbf{C}_B$ is also equal to one. That is, it has only one non-zero eigenvalue. The corresponding eigenvector of this non-zero eigenvalue is[6,10]

$$\mathbf{E}_1 = \frac{\mathbf{C}_w^{-1}(\mathbf{m}_2 - \mathbf{m}_1)}{\left\|\mathbf{C}_w^{-1}(\mathbf{m}_2 - \mathbf{m}_1)\right\|} \qquad (31)$$

Where the constant denominator is chosen to make the norm of the eigenvector unity, i.e. $\|\mathbf{E}_1\| = 1$. For our two-class incident detection problem the eigenvector is a function of one vector $(\mathbf{m}_2 - \mathbf{m}_1)$ only, requiring one discriminating feature, and the mapping function yielding the output vector $\mathbf{Y}$ is:

$$\mathbf{Y} = \mathbf{E}_1^T\mathbf{X} = c(\mathbf{m}_2 - \mathbf{m}_1)^T\mathbf{C}_w^{-1}\mathbf{X} \qquad (32)$$

where c is a constant.

In the incident detection problem, the value of d is varied from 3 to 6, when we apply LDA to a single data series at a time. If the LDA is carried out using all the data series (upstream and downstream traffic volume and occupancy) together then d will be equal to the number of data series considered (4 in this particular case). On the other hand m = 1 (Figure 4) represents a single value of effective traffic occupancy or volume for a given time period of 8 minutes. Since the incident detection problem is a two-class classification problem only one feature is sufficient to differentiate between the two classes. Consequently, number of input nodes of the neural network is reduced to 4.

Equation (35) represents the general FLD function. The NMC function does not consider the covariance part $(\mathbf{C}_W^{-1})$ and is represented by

$$\mathbf{Y} = \mathbf{E}_1^T \mathbf{X} = c(\mathbf{m}_2 - \mathbf{m}_1)^T \mathbf{X} \tag{33}$$

As the name suggests the Nearest Mean Classifier (NMC) classifies the data on the basis of distance from the class means. Thus, in the two-class incident detection problem it generates the perpendicular bisector between the class means. This type of linear classification is ideal for classes with identical distribution of data around the class means. But in the incident detection case, the incident and incident-free data may not have identical distribution around their class means. Consequently, the covariance part has to be considered for optimal linear classification. The standard FLD takes into account the covariance part, but linear classification using FLD involves inversion of within-class covariance matrix $(\mathbf{C}_W)$, which is often an ill-conditioned matrix. This problem can be overcome by adding some constant value (δ) to the diagonal elements of the covariance matrix as follows[14]:

:

$$\mathbf{C}_{WR} = \mathbf{C}_w + \delta \mathbf{I} \tag{34}$$

where $\mathbf{I}$ is an identity matrix. In this case, the classifier is known as Regularized FLD.

6. DATA ACQUISITION

Traffic incident detection is a real-life problem. Therefore it is quite essential to test the model with realistic traffic data. At present, different types of detection techniques are used to measure traffic flow properties, such as vehicle velocity, traffic volume and occupancy. Different incident detection algorithms use different combinations of these traffic data types. In the companion paper[1], we will consider various combinations of traffic volume, occupancy, and average speeds and investigate their effects on the incident detection algorithm.

A large number (up to a few hundreds) of traffic patterns with and without incidents are normally needed to train a neural network model for incident detection effectively. Data for many traffic patterns with incident cases for a particular location or similar locations are not readily available. Traffic incident data are collected in a variety of ways. The first data source often is the information logged in by the central operator monitoring the traffic conditions on freeways. Methods of surveillance and detection vary from actual observation of the incident by traveling motorists, highway patrol, or traffic reporting units, to sightings of an incident through the use of closed-circuit televisions and cameras, to detection through computerized electronic surveillance and control systems.

The last method is based on using sensors placed along freeways at intervals of a few hundred meters to a few kilometers and computers to process the traffic flow data. In this approach, sensors detect the effects of incident occurred within two neighboring sets of detectors rather than the incidents themselves. In some incident detection algorithms an incident is detected after a few minutes, which is relatively considerable. One of our goals in this research is to minimize the detection time.

In an automated freeway incident detection system (AFIDS) an entire freeway system can be monitored continuously in a central office through the use of a network of sensors without actually anyone observing the incidents. But, first the AFIDS has to be trained using the data obtained from sensors. At present, most of the time an operator creates an incident log manually by examining the incident data obtained from various sources including sensors and specifying and recording the location, time, duration, and cause of the incident, upstream/downstream sensor IDs, number of casualties and injuries, and number of lanes blocked. Unfortunately, some of the important information such as the sensor ID or the time of detection is frequently left out due to human error or other reasons. Consequently, incident logs obtained from departments of transportation can not be used directly to train the incident detection algorithm. Also some incidents, called isolated incidents, may not have any impact on the traffic flow and therefore are of no consequence to the traffic incident detection algorithm and should not be included in the training set.

As such, we found the traffic incident logs obtained from several departments of transportation (DOT) including Arizona DOT and Minnesota DOT not to be helpful in training the IDA. An alternative to the use of the actual incident data with the

aforementioned drawbacks is simulation of the freeway traffic using a simulation package. Traffic simulation further provides a means to investigate various traffic conditions.

In this work we use the traffic simulation package TSIS/CORSIM developed by ITT Systems and Sciences Corporation (http://www.fhwa-tsis.com) to simulate the freeway traffic needed to train the IDA. This simulation package allows you to simulate the road conditions for a given grade, curvature, or maximum allowable speed as well as the traffic conditions such as traffic flow, incident location, percentage lane blockage, and duration of the incident. The simulated data can be displayed graphically on the computer screen. An example is shown in Figure 5 displaying a straight four-lane freeway segment with two sets of entry and exit ramps. TSIS/CORSIM provides a comprehensive freeway incident simulation module called FRESIM (Freeway Simulation Package). An example of a simulation instance for the freeway of Figure 5 is shown in Figure 6, displaying the location of the accident and the traffic congestion after the incident.

We can specify either blockages in one or both lanes or *rubbernecking* which is a reduction in the capacity of a lane without a blockage (defined as a percentage reduction in the capacity) due to blockage in a neighboring lane or an incident on the shoulder. The user can specify the following for an incident: the longitudinal location on a freeway link, the length of the blockage, and the duration of the incident. The characteristics of an incident can be changed during the incident duration. For example, it is possible to specify a two-lane blockage turning into a one-lane blockage after a specified duration. The lane from which the blockage is removed can then become unrestricted or subjected

to rubbernecking. The simulation parameters to be chosen are the percentage reduction in the capacity of the freeway.

7. RESULTS

Figures 7(a) to 7(d) show the results of the filtered data after applying the wavelet transform for upstream and downstream traffic volume and occupancy using the simulated data for an 8-minute traffic pattern. Similarly, Figures 8(a) to 8(d) show similar results using actual data obtained from the Minnesota DOT over a period of 150 minutes. The smoothening effects of the traffic data are clearly noted in these figures. In a companion paper, the two-stage feature extraction algorithm presented in this article is used as a preprocessor for a robust neural network model for automatic detection of traffic incidents[1].

ACKNOWLEDGMENT

This article is based on the research sponsored by Ohio Department of Transportation, Federal Highway Administration, and The Ohio State University Center for Intelligent Transportation Research. The authors are also grateful to Mr. George Saylor of Ohio Department of Transportation (DOT), Mr. Leonard Palek of Minnesota DOT, and Phil Carter of Arizona DOT for their help in obtaining actual traffic data and Dr. Henry Lieu of Federal Highway Administration for providing access to the traffic simulation package (TSIS/CORSIM) to test the computational models developed in this research.

REFERENCES

1. Adeli, H. and Samant, A.M., "An adaptive conjugate gradient neural network-wavelet model for traffic incident detection", submitted, 2000.
2. Burrus, C.S., Gopinath, R.A. and, Guo, H., *Introduction to Wavelets and Wavelet Transforms: A Primer,* Prentice Hall, Upper Saddle River, NJ, 1998.
3. Cheu, R.L. and Ritchie, S.G., "Automated detection of lane-blocking freeway incidents using artificial neural networks", *Transportation Research* 3C(6), 1995, pp. 371-388.
4. Daubechies, I., *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.
5. Farge, M., Hunt, J.C.R., and Vassilicos, J.C., Ed., *Wavelets, Fractals, and Fourier Transform*, The Institute of Mathematics and its Applications Conference series, Clarendon Press, Oxford, 1993.
6. Fukunaga, K., *Introduction to Statistical Pattern Recognition*, 2nd Ed., Academic Press, NY, 1990.
7. Hsiao, C.H., Lin, C.T., and Cassidy, M., "Application of fuzzy logic and neural networks to automatically detect freeway traffic incidents", *Journal of Transportation Engineering*, Vol. 120, No. 5, 1994, pp. 753-771.
8. Jameson, L.M., Hussaini, M.Y., Erlebacher, M., Eds., *Wavelets: Theory and Applications*, ICASE/LaRC Series in Computational Science and Engineering, Oxford University Press, NY, 1996.
9. Mallat, S.G., *Wavelet Tour of Signal Processing*, Academic Press, London, 1998.
10. McLachlan, G.J., *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley, NY, 1992.
11. Payne, H.J., and Tignor, S.C., "Freeway incident-detection algorithms based on decision trees with states", *Transportation Research Records* 682, TRB, National Research Council, Washington, DC, 1978, pp.30-37.

12. Persaud, B.N., Hall, F.L., and Hall, L.M., "Congestion identification aspects of the McMaster incident detection algorithm", *Transportation Research Records* 1287, TRB, National Research Council, Washington, DC, 1990, pp.167-175.

13. Ruskai, M.B., Beylkin, G., Coifman, R., Daubechies, I., Mallat, S., Meyer, Y., and Raphael L., *Wavelets and their Applications*, Jones and Bartlett Publishers, Boston, 1992.

14. Skurichina, M. and Duin, R., "Bagging for linear classifiers", *Pattern Recognition,* Vol. 31, No. 7, 1998, pp. 909-930.

15. Stephanedes, Y.J. and Chassiakos, A.P., "Application of filtering techniques for incident detection", *Journal of Transportation Engineering*, Vol. 119, 1993, pp.13-20.

16. Stephanedes, Y.J. and Liu, X., "Artificial neural networks for freeway incident detection", *Transportation Research Records* 1494, TRB, National Research Council, Washington, DC, 1995, pp. 91-97.

17. Vetterli, M. and Kovacevic, J., *Wavelets and Subband Coding*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1995.

18. Wickerhauser, M.V., *Adapted Wavelet Analysis from Theory to Software,* A K Peters, Ltd., Wellesley, MA, 1994.

CAPTIONS FOR FIGURES

Figure 1 Downsampler

Figure 2 Two-band single stage filter

Figure 3 Two-band two-stage filter

Figure 4 Pre-processing (Filtering and Downsampling of traffic data)

Figure 5 Simulated freeway geometry

Figure 6 Simulation of a traffic incident

Figure 7 Results of filtered data after applying DWT and LDA using simulated data for an 8-minute traffic pattern

Figure 8 Results of filtered data after applying DWT and LDA using actual data from the Minnesota DOT for a period of 150 minutes

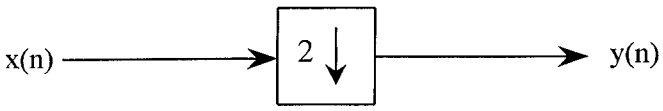


Figure 1

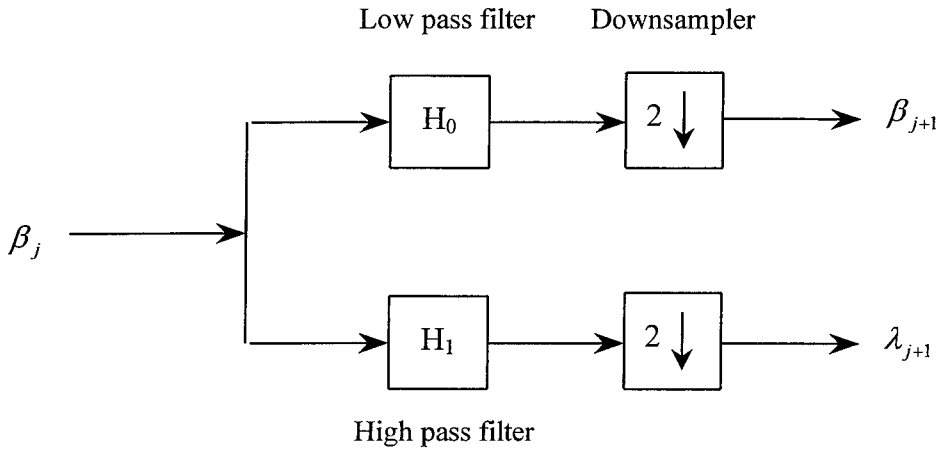

Low pass filter
Downsampler
H_0
2
β_{j+1}
β_j
H_1
2
λ_{j+1}
High pass filter

Figure 2

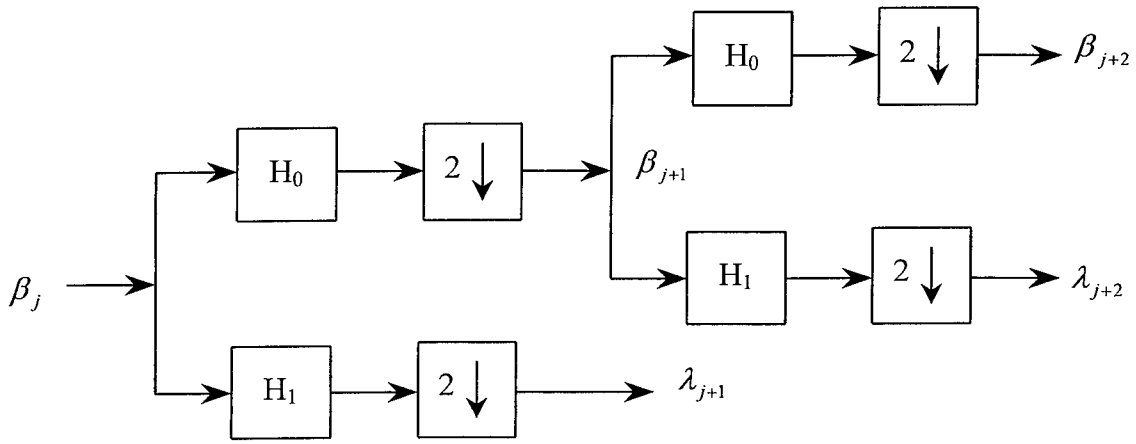


Figure 3

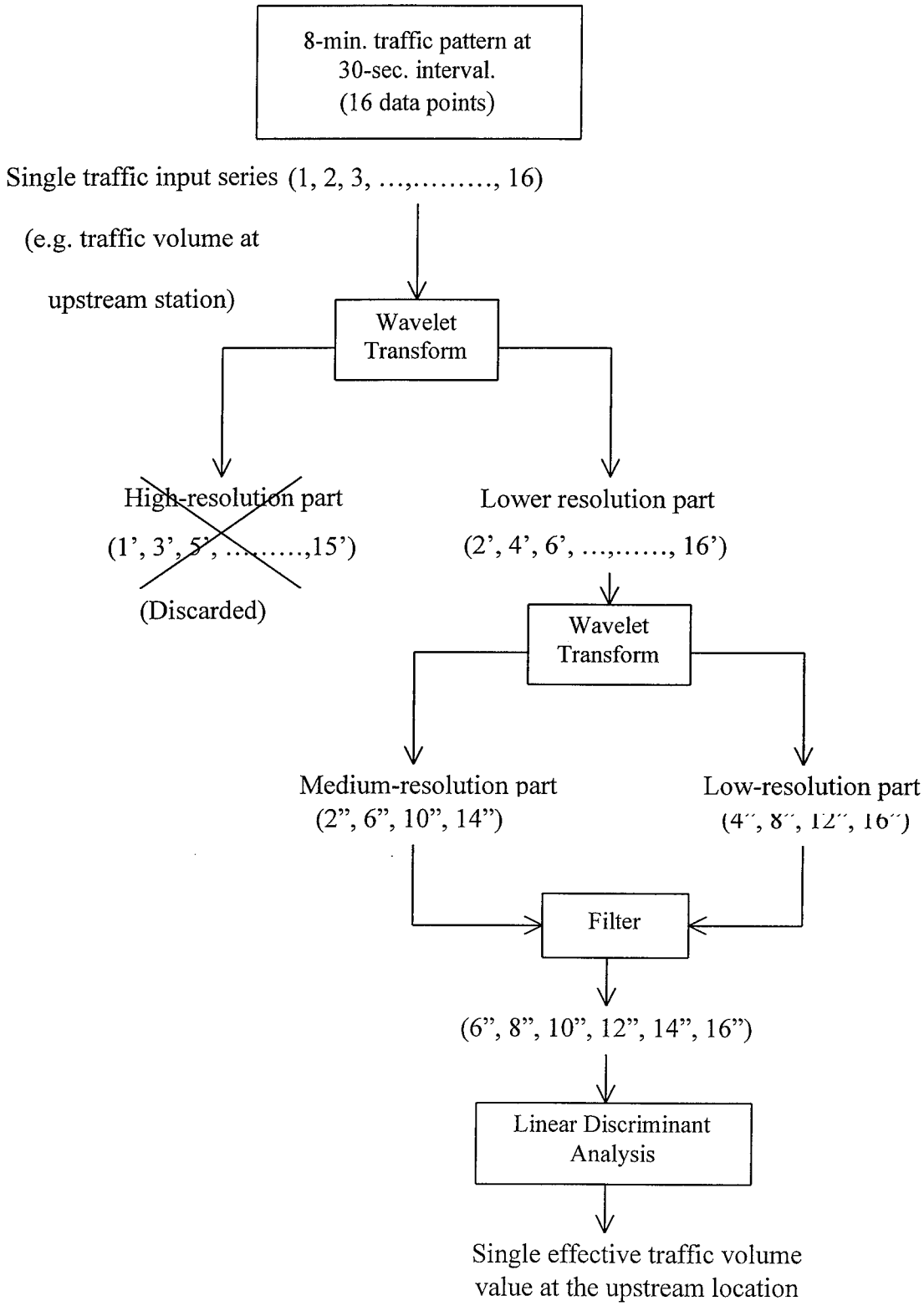

8-min. traffic pattern at
30-sec. interval.
(16 data points)
Single traffic input series (1, 2, 3, ...,....., 16)
(e.g. traffic volume at
upstream station)
Wavelet
Transform
High-resolution part
(1', 3', 5', ...,15')
(Discarded)
Lower resolution part
(2', 4', 6', ...,....., 16')
Wavelet
Transform
Medium-resolution part
(2", 6", 10", 14")
Low-resolution part
(4", 8", 12", 16")
Filter
(6", 8", 10", 12", 14", 16")
Linear Discriminant
Analysis
Single effective traffic volume
value at the upstream location

Figure 4

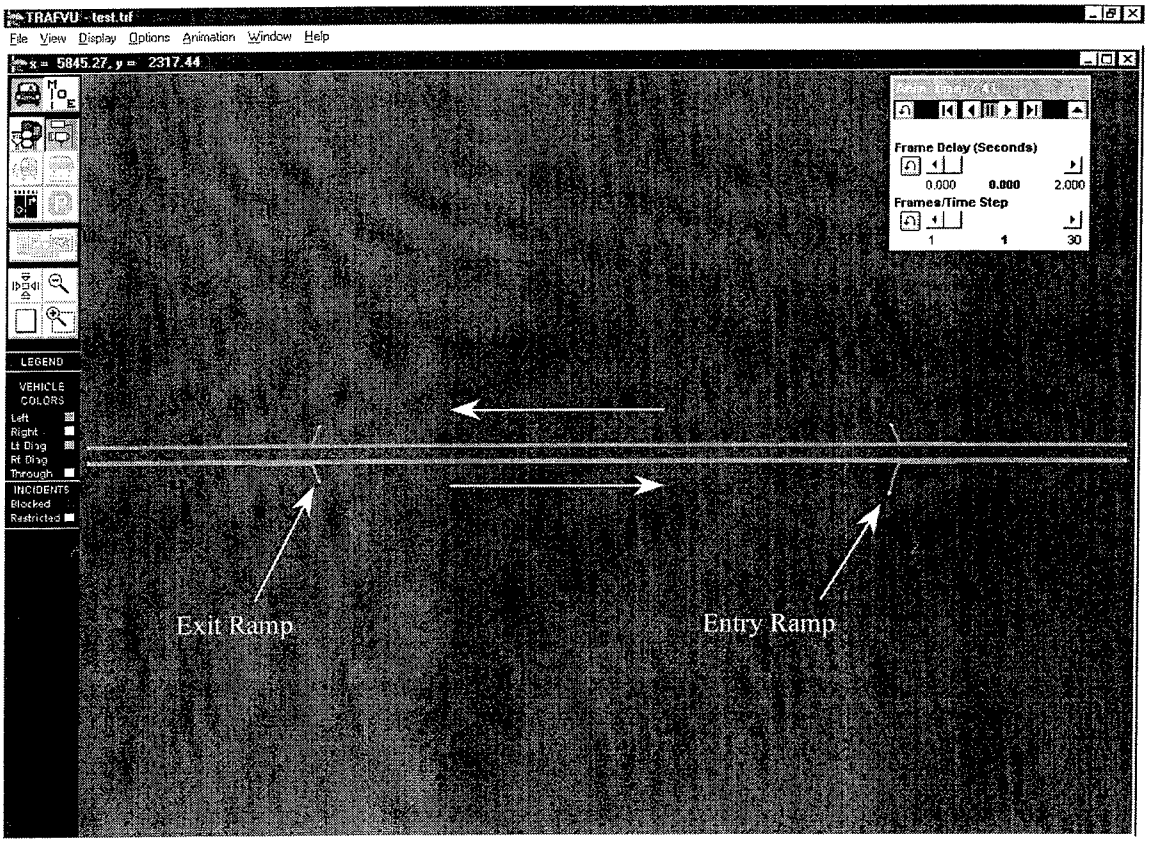

TRAFVU - test.tif
File View Display Options Animation Window Help
x = 5845.27, y = 2317.44
Frame Delay (Seconds)
0.000 0.000 2.000
Frames/Time Step
1 1 30
LEGEND
VEHICLE COLORS
Left
Right
Lt Diag
Rt Diag
Through
INCIDENTS
Blocked
Restricted
Exit Ramp
Entry Ramp

Figure 5

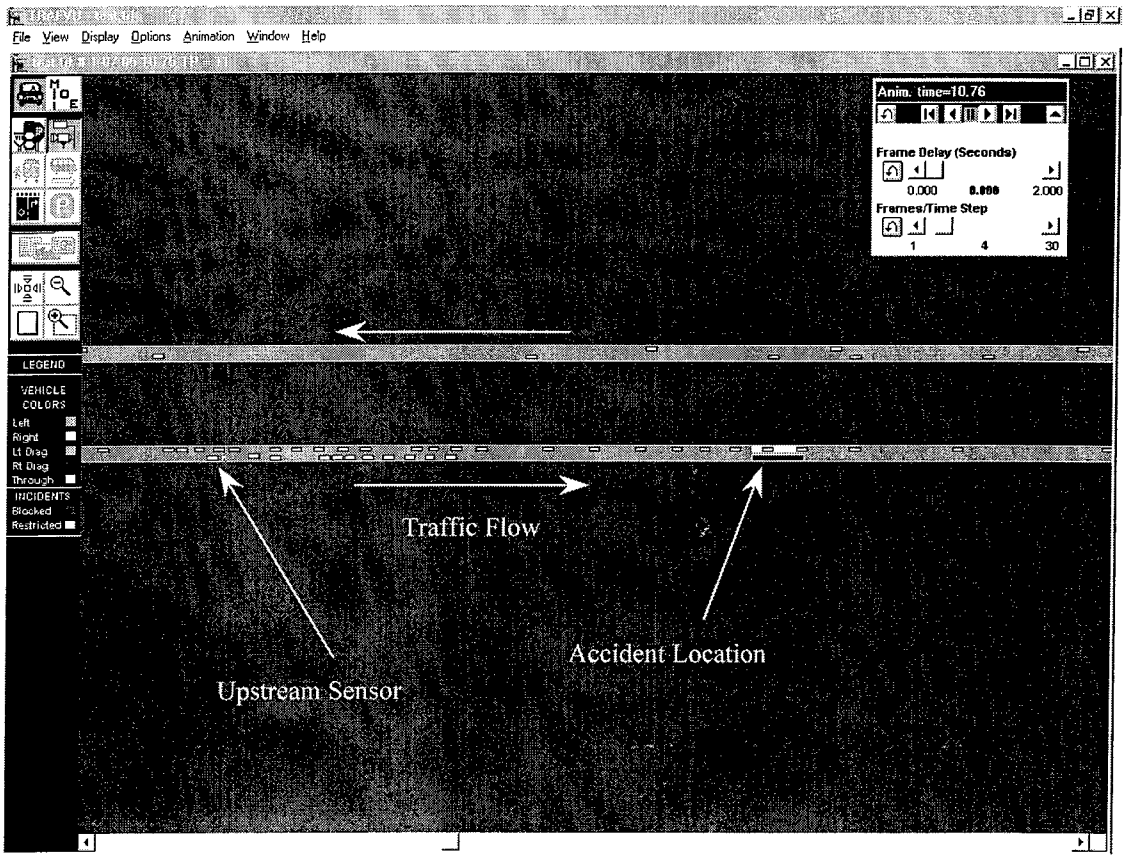

File View Display Options Animation Window Help
Anim. time=10.76
Frame Delay (Seconds)
0.000
0.000
2.000
Frames/Time Step
1
4
30
LEGEND
VEHICLE COLORS
Left
Right
Lt Diag
Rt Diag
Through
INCIDENTS
Blocked
Restricted
Traffic Flow
Accident Location
Upstream Sensor

Figure 6

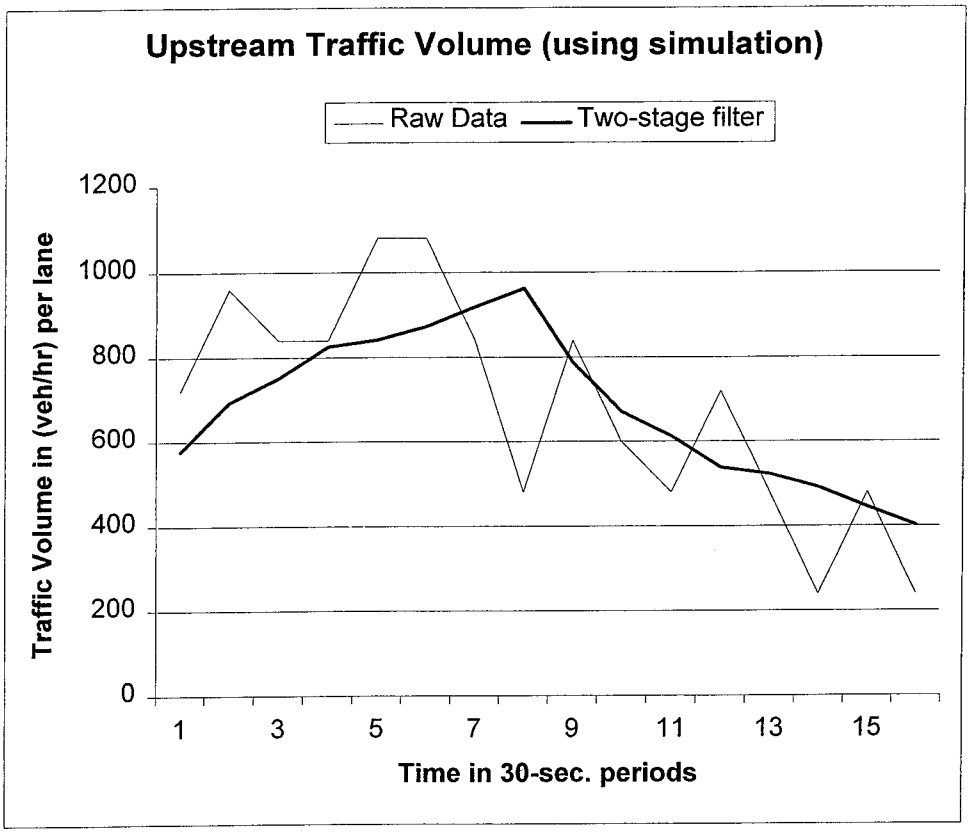

Upstream Traffic Volume (using simulation)
Raw Data
Two-stage filter
Traffic Volume in (veh/hr) per lane
1200
1000
800
600
400
200
0
1
3
5
7
9
11
13
15
Time in 30-sec. periods

Figure 7(a)

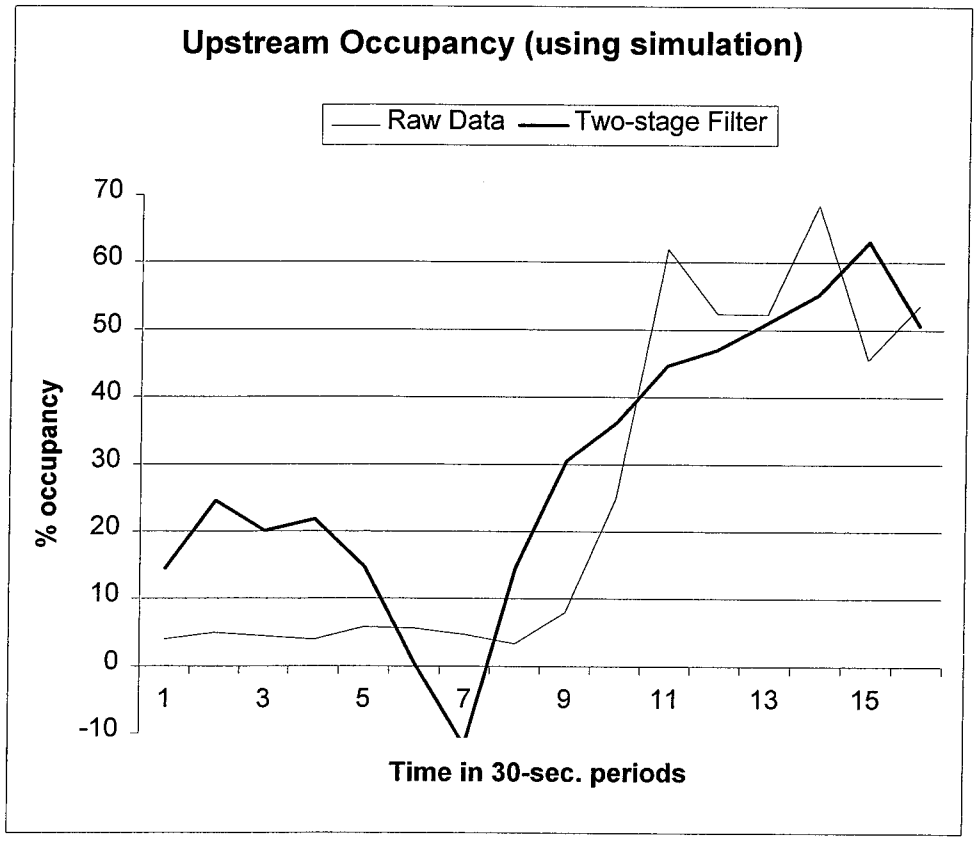

Upstream Occupancy (using simulation)
Raw Data
Two-stage Filter
70
60
50
40
30
20
10
0
-10
% occupancy
1
3
5
7
9
11
13
15
Time in 30-sec. periods

Figure 7(b)

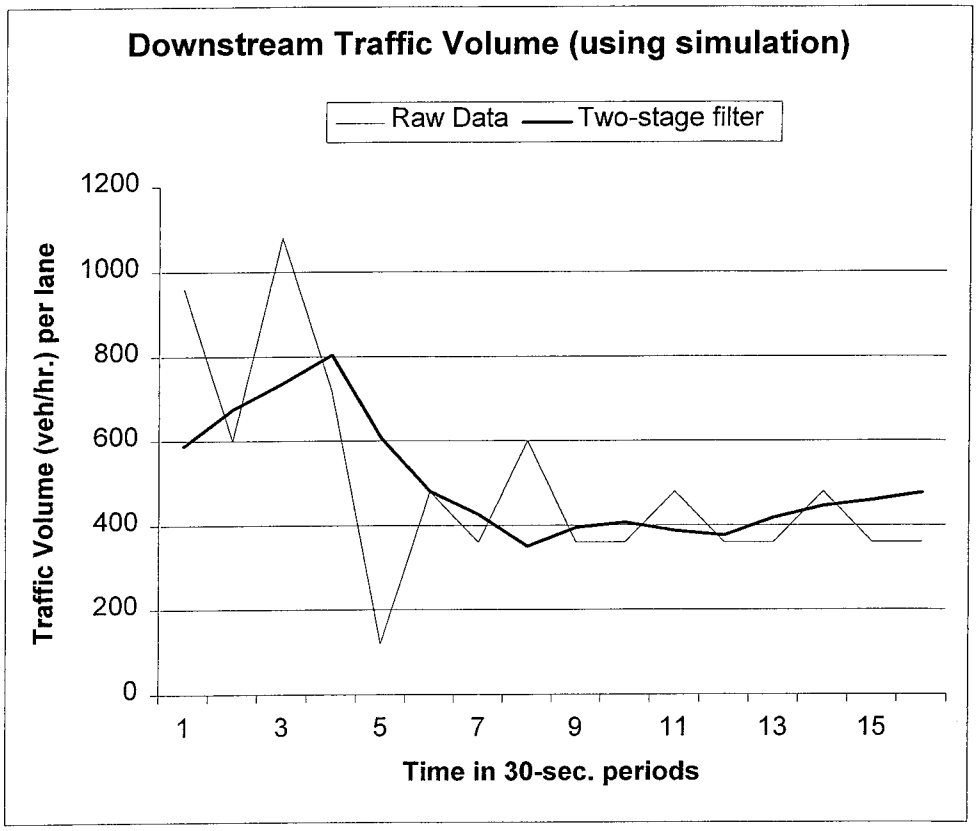

Downstream Traffic Volume (using simulation)
Raw Data
Two-stage filter
Traffic Volume (veh/hr.) per lane
1200
1000
800
600
400
200
0
1
3
5
7
9
11
13
15
Time in 30-sec. periods

Figure 7(c)

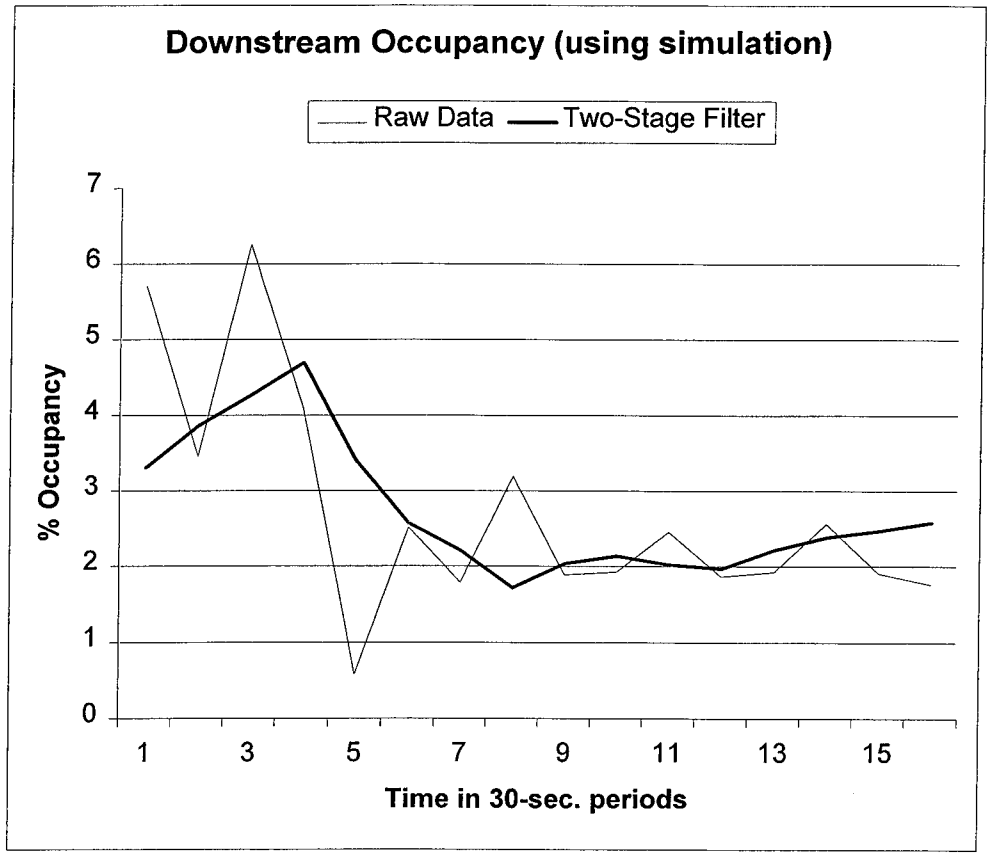

Downstream Occupancy (using simulation)
Raw Data
Two-Stage Filter
% Occupancy
0
1
2
3
4
5
6
7
1
3
5
7
9
11
13
15
Time in 30-sec. periods

Figure 7(d)

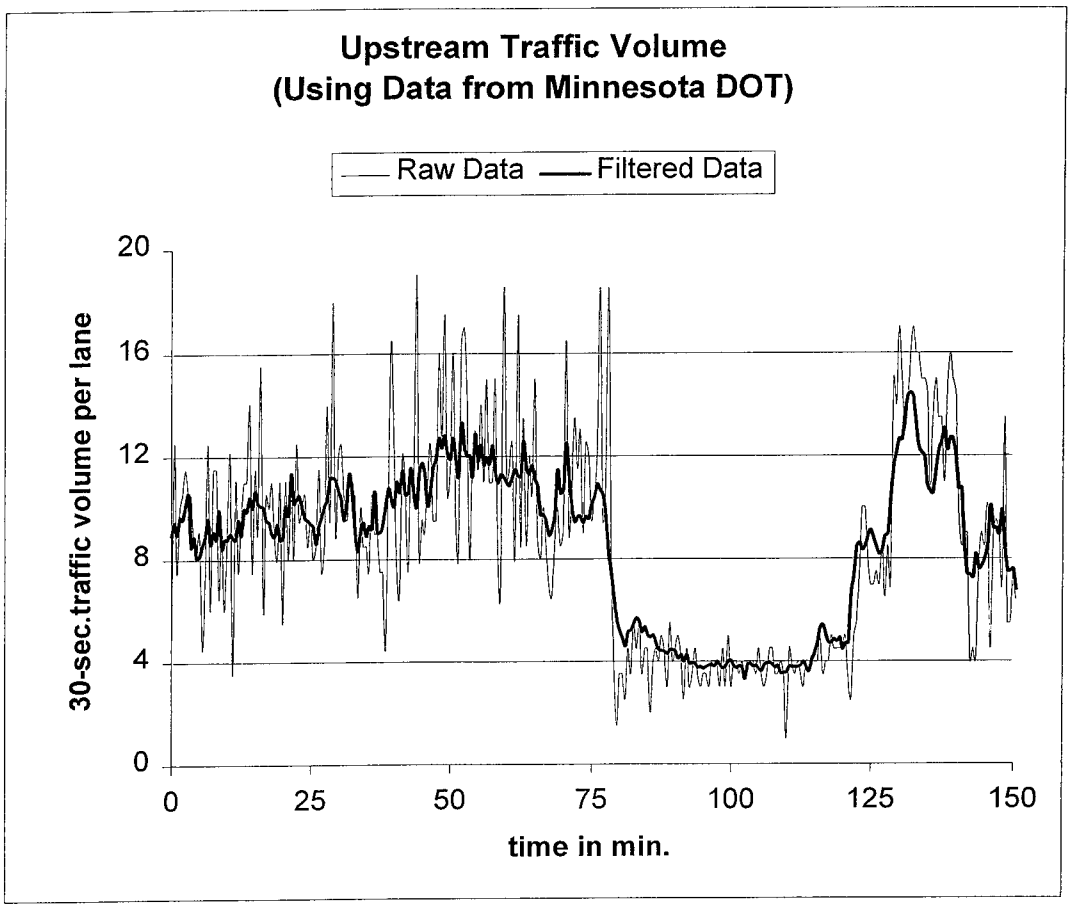

Upstream Traffic Volume
(Using Data from Minnesota DOT)
Raw Data
Filtered Data
30-sec.traffic volume per lane
20
16
12
8
4
0
0
25
50
75
100
125
150
time in min.

Figure 8(a)

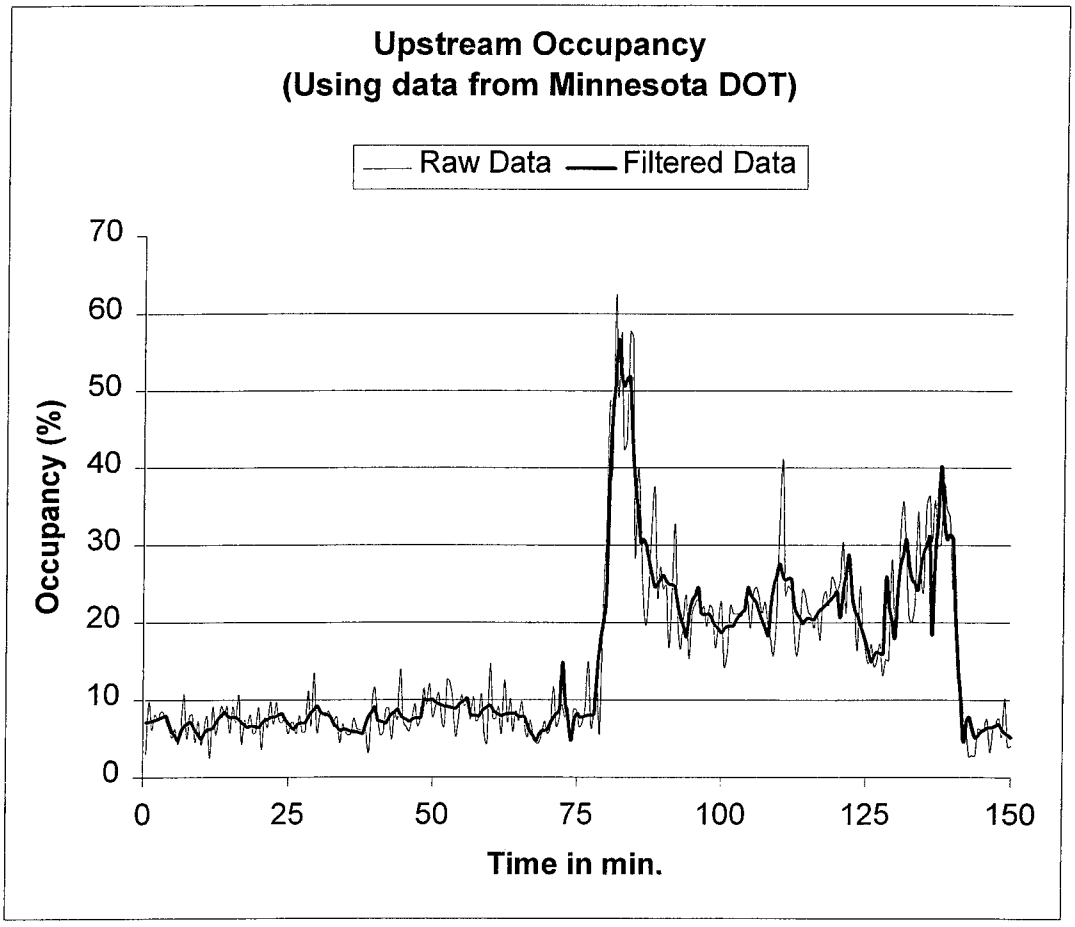

Upstream Occupancy
(Using data from Minnesota DOT)
Raw Data
Filtered Data
Occupancy (%)
70
60
50
40
30
20
10
0
0
25
50
75
100
125
150
Time in min.

Figure 8(b)

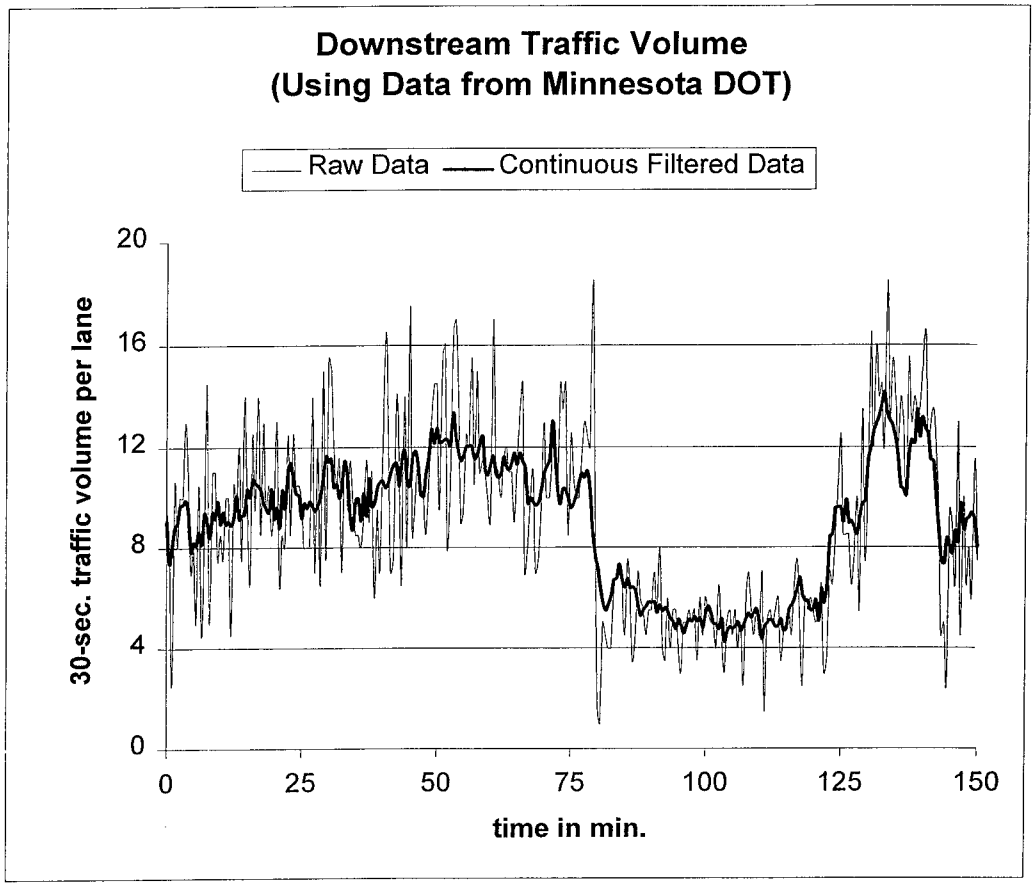

Downstream Traffic Volume
(Using Data from Minnesota DOT)
Raw Data
Continuous Filtered Data
30-sec. traffic volume per lane
20
16
12
8
4
0
0
25
50
75
100
125
150
time in min.

Figure 8(c)

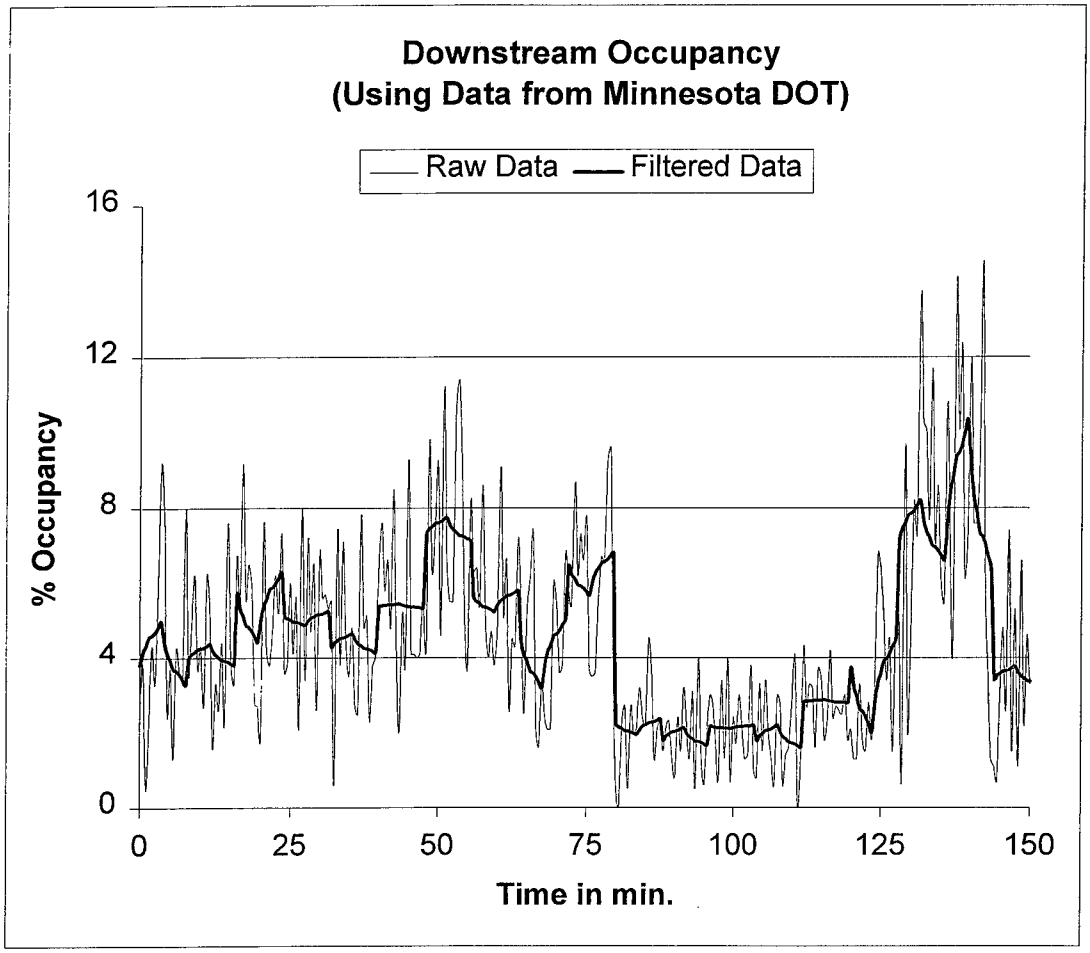

Downstream Occupancy
(Using Data from Minnesota DOT)
Raw Data
Filtered Data
% Occupancy
16
12
8
4
0
0
25
50
75
100
125
150
Time in min.

Figure 8(d)

Part 2

AN ADAPTIVE CONJUGATE GRADIENT NEURAL NETWORK-WAVELET MODEL FOR TRAFFIC INCIDENT DETECTION

H. Adeli[1] and A. Samant[2]

[1]Professor, [2]Graduate Research Associate

Dept. of Civil and Environmental Engineering and Geodetic Science

The Ohio State University, 470 Hitchcock Hall

2070 Neil Avenue, Columbus, OH 43210.

ABSTRACT: Artificial neural networks are known to be effective in solving problems involving pattern recognition and classification. The traffic incident detection problem can be viewed as recognizing incident patterns from the incident-free patterns. A neural network classifier has to be trained first using incident and incident-free traffic data. The dimensionality of the training input data is high and the embedded incident characteristics are not easily detectable. In this article we present a computational model for automatic traffic incident detection using discrete wavelet transform, linear discriminant analysis, and neural networks. Wavelet transform and linear discriminant analysis are used for feature extraction, de-noising, and effective preprocessing of data before an adaptive neural network model is used to make the traffic incident detection. Simulated as well as actual traffic data are used to test the model. For incidents with duration of more than five minutes, the incident detection model yields a detection rate of nearly 100% and false alarm rate of about 1% for two- or three-lane freeways.

1. INTRODUCTION

Stephanedes et al.[15] used a moving average method to reduce the effect of random fluctuations in the traffic on the incident detection algorithm. They average the differences in the occupancies at upstream and downstream locations over 3-minute periods using data recorded at 30-second intervals. Their comparison with other existing approaches showed improvement in reducing the false alarm rates. They report a detection rate of around 90% for a false alarm rate of about 1%. They also note that "the algorithm performance may exhibit varying degree of transferability across test locations". To take into account the uncertainty and imprecision inherent in the incident detection[9], researchers have recently explored the use of new computing approaches such as fuzzy logic[9,17] and neural networks[2,3] to improve the incident detection rate with simultaneous reduction in false alarms. Neural networks are known as a powerful method for pattern recognition and classification[2]. The price to pay for their adaptive learning capability is often the need for large computational resources when the problem is complicated requiring a large network and a large number of training instances. As an example, if we use an 8-minute traffic pattern with 30-sec. intervals and upstream and downstream traffic volumes and occupancies as input, then the number of input nodes for the neural network model will be 4x8x2 = 64. If we use one hidden layer with the same number of nodes as the input layer then the number of links connecting the input layer to the hidden layer would be 64x64 = 4096. This means we have to solve a large optimization problem with 4096 + 64 = 4160 variables (assuming one output node) in order to find the 4160 weights of the network. Further, a few hundreds training instances are needed to train such a large network.

In order to reduce the high dimensionality of the network and improve its computational efficiency, we first employ a two-stage feature extraction model using the discrete wavelet transform and linear discriminant analysis, as described in the companion paper[14]. This will reduce the number of nodes in the input and hidden layers for the aforementioned example to 4, thus reducing the size of the network substantially and resulting in significant computational efficiency (Figure 1).

A robust feature extraction algorithm also helps reduce the dimension of the input space for a neural network model without any significant loss of related traffic information, resulting in a substantial reduction in

- the network size (i.e., the number of nodes in the input and hidden layers),
- the effect of random traffic fluctuation on the learning curve of the neural network, (Learning curve for any neural network is defined as the relation between the mean squared error of the output and the number of iterations required for the training. As the random traffic fluctuations are reduced the total number of iterations required for convergence reduces too.)
- the computational resources required to train the network, and
- the required number of training samples (that means more accurate generalization).

Backpropagation neural network[13] has been used to solve the traffic incident detection problem[5,16]. The attraction of backpropagation is its simplicity. But, it suffers from a number of shortcomings[1,10]:

1. It often requires a very large number of iterations for convergence,
2. Its convergence depends heavily on the selection of two problem-dependent parameters, *learning and momentum ratios*, that have to be selected by trial and error,

3. It suffers from the hill-climbing problem, that is entrapment in a local minimum.

In this work we use the adaptive conjugate gradient neural network learning algorithm of Adeli and Hung[1], which combines the conjugate gradient method originally proposed by Fletcher and Reeves[7] and modified by Powell[12] with an inexact line search with three criteria for finding the optimal search direction.

2. ADAPTIVE CONJUGATE GRADIENT NEURAL NETWORK LEARNING MODEL

The conjugate gradient method is based on the steepest descent method where weight changes are made along the direction resulting in the maximum decrease in the system error. Determination of the step length of a gradient-based optimization algorithm has a significant impact on its efficiency[1]. A very accurate or "exact" line search requires many function evaluations thus making the algorithm prohibitively and unnecessarily expensive. An appropriate inexact line search algorithm can determine the step length, within a small percentage of that found based on an exact search. Adeli and Hung[1] use the backtracking inexact line search algorithm of Dennis and Schnable[6], the step length selection terminating criterion of Armijo[4] to ensure the step length is not too large, the terminating criterion of Goldstein[8] to ensure the step length is not too small, and the direction convergence criterion of Nocedal[11] to ensure that the descent direction is always generated.

The steps of the adaptive conjugate gradient algorithm for training of neural networks are presented here briefly. For a classification problem involving T decision variables, the training of the network is started using a randomly generated initial weight

vector of ($W^0 \in R^T$). Two stopping criteria are provided for convergence, one for the gradient vector ($\varepsilon = 10^{-5}$ to 10^{-6}) and one for the minimum system error (0.01 or 0.001). The minimum (*minlen*) and maximum (*maxlen*) step length is set to 0.0001 and 100, respectively. The initial search direction is set to 0. The parameters θ and β are chosen equal to 0.9 and 0.01, respectively, per Adeli and Hung[2]. The outer iteration number, n, is set to 1. The decision variable counter, t, is set to 0.

1. Steps (a) through (e) are carried out for p training samples (k = 1 to p).

   a) Feed-forward procedure is performed on the neural network. The output of any node k in layer i+1 is calculated by:

$$O_k^{i+1} = A(P) \tag{1}$$

   where

$$P = \sum_{i=1}^{p} w_i^k o_i^k \quad and \quad A = \frac{1}{1+e^{-P}} \tag{2}$$

   P is called the pre-processing function and A is called an activation function.

   b) The system error is calculated for the kth training instance. In the traffic incident detection case, there is only one output node so the error will be just the square of the difference between actual (y_k) and actual output (o_k).

$$E_k = \frac{1}{2}(y_k - o_k)^2 \tag{3}$$

   c) The deltas in the output layer for the kth training instance are calculated as follows:

$$\delta_k^o = (y_k - o_k)(1 - o_k)\, o_k \tag{4}$$

d) Deltas for the hidden layers are then calculated back, propagating the error:

$$\delta_{kq}^{h} = o_{kq}(1 - o_{kq})\sum_{j=1}^{5}(\delta_{k}^{o} w_{qj}) \tag{5}$$

e) The gradient vector for the k^{th} training instance is calculated as:

$$\nabla E_k = \frac{\partial E}{\partial w_{qr}} = \delta_{kq}^{(i+1)} o_{kr}^{i} \tag{6}$$

2. The total system error is then calculated by adding-up the individual errors from step 1(b). If the total system error satisfies the minimum error convergence criterion, the training is completed. Otherwise, the gradient vector for the total system error is calculated. A new search direction is assigned as negative of the gradient vector as follows:

$$\mathbf{d}^{(n)} = -\nabla \mathbf{E}\left(\mathbf{W}^{(n)}\right) \tag{7}$$

If the gradient vector satisfies the convergence criterion $\left|\nabla \mathbf{E}(\mathbf{W}^{(n)})\right| < \varepsilon$ then the training is stopped and the weight vector obtained is the final solution. Otherwise, following steps are performed.

3. The decision variable counter (t) is increased: t = t + 1. If t ≥ T, that is if t exceeds the number of decision variables, then it is set to 0 (t = 0). If t = 1, α_n is set to 0. Otherwise, a new conjugate direction is calculated as follows:

$$\mathbf{d}^{(n)} = -\nabla \mathbf{E}\left(\mathbf{W}^{(n)}\right) + \alpha_n \mathbf{d}^{(n-1)} \tag{8}$$

where

$$\alpha_n = \max\left\{0, \frac{\nabla \mathbf{E}\left(\mathbf{W}^{(n)}\right)^{\mathrm{T}} \mathbf{v}^{(n-1)}}{\left|\nabla \mathbf{E}\left(\mathbf{W}^{(n-1)}\right)\right|^2}\right\} \tag{9}$$

and

$$\mathbf{v}^{(n-1)} = \nabla \mathbf{E}\left(\mathbf{W}^{(n)}\right) - \nabla \mathbf{E}\left(\mathbf{W}^{(n-1)}\right) \tag{10}$$

4. The inexact line search algorithm is performed to calculate the step length λ. First, λ is initialized equal to one. Then, the Armijo[4] criterion is applied to ensure the step length is not too large. If the step length is too large, step 10 is carried out. Otherwise, the Goldstein[8] criterion is applied to ensure the step length is not too small. If this criterion is satisfied then step 8 is carried out where a new search direction is calculated using the new value of λ. If the Goldstein[8] criterion is not satisfied then value of λ is checked. If its value changes (that is $\lambda \neq 1$), then step 6 is carried out. Otherwise, next step is performed.

5. A new λ value is set as follows:

$$\lambda = \min(2\lambda_n, \mathit{maxlen}) \tag{11}$$

A new search direction $\mathbf{d}^{(n+)}$ is calculated (Eq. 8). Using this new search direction Nocedal[11] direction convergence criterion is checked. If the direction convergence criterion is not satisfied then step 6 is carried out. If the direction convergence criterion is satisfied then Goldstein[8] criterion is checked. If Goldstein criterion is not satisfied or if λ value becomes greater than *maxlen*, step 6 is carried out. Otherwise this step is repeated.

6. If $\lambda<1$, or, if $\lambda>1$ and direction convergence criterion of Nocedal[11] is not satisfied then step 7 is carried out. Otherwise, step 12 is carried out directly.

7. A new value of λ is calculated using backtracking and parabolic interpolation. A new search direction is calculated using Eq. (8). This is repeated until both Nocedal[11] and Goldstein[8] criteria are satisfied simultaneously. Then, step 12 is carried out.

8. A new search direction is found and checked for the descent condition criterion of Nocedal[11]. If it is satisfied then step 12 is carried out directly. Otherwise, the next step is performed.

9. A new λ is found by backtracking and a new search direction is computed (Eq. 8). This step is repeated until the gradient descent condition of Nocedal[11] is satisfied. Then step 12 is carried out directly.

10. If $\lambda <$ *minlen*, λ is set to 0 and step 12 is performed.

11. If $\lambda = 1.0$, backtracking is performed using parabolic interpolation to find a new λ. Otherwise, cubic interpolation is used to find a new λ.

12. If this step is executed directly after step 7, 8, 9 or 10 then the inexact line search algorithm is stopped and step 13 is performed. Otherwise, step 4 is carried out using a new value of λ.

13. Weight vector is updated along with the iteration counter as follows:

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} + \lambda_n \mathbf{d}^{(n)} \tag{12}$$

$$n = n + 1 \tag{13}$$

If n exceeds the specified maximum number of iterations the training is stopped.

14. If step 14 is executed directly after step 2 or step 3, then stop. In this case weight vector obtained is the optimum weight vector.

This algorithm is repeated after every T iterations (for T decision variables), and α_n is set to zero for the t = 1.

3. INCIDENT DETECTION RESULTS USING VARIOUS APPROACHES

As discussed in the companion paper[14] the actual data obtained from several state departments of transportation including Minnesota DOT were not sufficient to train the classifiers and the neural network. Consequently, the results presented in this section are based on simulated data using TSIS/CORSIM developed by ITT Systems and Sciences Corporation (http://www.fhwa-tsis.com). Three types of traffic data are used and investigated: traffic volume, traffic occupancy, and average vehicle speed.

Deciding on the data polling frequency, that is the data-recording interval, is crucial in developing an automated freeway incident detection and management system. If the interval is very small, say 5 sec., then the change in the traffic data per interval may not be noticeable and the hardware and computational cost can become prohibitively high. The increase in the computational cost will be due to an increase in the size of the network as well as the required number of training instances. On the other hand, if this interval is made large, say 5 minutes, then it will take a relatively long time to detect the incident and take appropriate recovery measures such as re-routing the traffic or providing emergency medical assistance. A data polling period of 20-40 sec is commonly used in automatic traffic incident detection models. We have used 30-sec intervals for the simulated traffic data.

The distance between the sensors also affects the incident detection rate and specially the time to detect the incident. If the distance is too small, say a couple of hundred meters, the number and cost of sensors needed to cover the same segment of the freeway will increase. On the other hand, if this distance is too large, say a few kilometers, the sensors will take a long time to detect the incident and may not detect small incidents at all. The appropriate distance appears to be in the range of 2000-3000 ft (600-900 m). The lower end of the range can be used for the critical sections of the freeway where the probability of incident occurrence is high, such as before the exit ramp and after the entry ramp, or where there is a reduction in the number of lanes. These are considered critical sections because of a large number of lane changes which is one of the main factors causing incidents.

The incident detection rate is mainly governed by the upstream traffic patterns and the incident detection time is mostly governed by the downstream traffic data, because an incident has a major impact on the upstream traffic and a relatively minor impact on the downstream traffic flow (especially the occupancy). However, these changes in the downstream traffic flow are immediate after the traffic incident as compared to those of the upstream traffic flow, which explains their impact on the detection time.

The traffic incident detection results presented in this section for various approaches are for a straight two-lane freeway segment (in one direction). The simulated data used for testing include 45 incidents with traffic volume varied from 300 to 2000 vehicles/hour per lane. In the subsequent section, we will consider the effects of

geometry, such as curvature and the number of lanes on the performance of the incident detection algorithm.

3.1 LDA

We will investigate the application of linear discriminant analysis in two different ways, as a linear classifier and as a feature enhancer. As a linear classifier, it is applied to all the data series simultaneously using a single data point from each data series, without using neural networks. As a feature enhancer, it is applied to each data series separately and the resulting traffic parameter values are used as input to the neural network model. Table 1 presents the classification results using three different types of LDA classifiers described in the companion paper[14]. It includes the incident detection rate, the false alarm rate, and the mean time for detection. These results show that LDA by itself is a poor classifier for the problem at hand.

3.2 DWT and LDA

Table 2 presents the classification results when the wavelet transform is preceded by the LDA. The results show improvement over LDA, but still not acceptable. However, the results of the two-stage feature extraction model presented in Samant and Adeli[14] can be used as input to a neural network model described in the previous section to obtain an accurate incident detection model, as presented subsequently.

3.3 ACGNN

In this work we will investigate various combinations of different traffic data series such as traffic volume, occupancy and average vehicle speed at upstream and downstream stations. Parametric studies will be performed to find out the most effective combination of the traffic data series.

Table 3 shows the results of traffic incident detection using three different types of traffic data and their combinations employing the ACGNN learning model. It is observed that the combination of all three parameters yields the best incident detection rate of 91.1% and the lowest false alarm rate of 5.1%. But, the results are only slightly better than those obtained from the combination of the traffic volume and occupancy with the corresponding numbers of 88.9% and 5.1%. Considering the fact that the three-parameter traffic data input increases the number of nodes in the input and hidden layers by a factor of 1.5 and the number of links (and the unknown weights) connecting the hidden layer to input and output layers by a factor of 1.5^2=2.25, we will choose the traffic volume and occupancy as the input parameters for the final incident detection algorithm.

The results presented in Table 3 show that the ACGNN is superior to the combination of DWT and LDA (Table 2). However, the 5.1% rate of false alarm is still too high. This can be explained by the fact that the incident and incident-free domains are not easily separable using the original unfiltered data.

3.4 DWT, LDA, and ACGNN

Table 4 shows the incident detection results employing the ACGNN algorithm after the filtering and preprocessing of data by DWT and LDA using the traffic volume and occupancy as input data. As explained in Samant and Adeli[14] the traffic data are first filtered using DWT and multi-resolution analysis and the high-resolution components are discarded. The low and medium resolution components are found to be sufficient for representing the traffic flow.

After the wavelet transform is performed, the resulting data can be applied to LDA or ACGNN in two different ways. Wavelet transform coefficients can be used

directly as the input to LDA or ACGNN. Alternatively, the traffic signal can be re-generated using an inverse of the DWT and setting the high-resolution coefficients equal to zero. The results for both cases are shown in Table 4. The two methods yield comparable results. Re-generating the traffic signals is an additional and unnecessary computational burden. While the wavelet transform coefficients have no physical significance their use is adequate and therefore recommended for computational efficiency.

It is observed that the new computational model for traffic incident detection based on preprocessing of the traffic data by DWT and LDA followed by application of the ACGNN yields a high incident detection rate of 97.8% and a low false alarm rate of around 1%. Further, the mean time for detection is about 38 seconds.

The traffic data obtained from Minnesota DOT included only two incidents over a 150-min. period. We used these data to test the new incident detection model trained using the simulated data. The model detected both incidents with time to incident detection of less than a minute.

4. EFFECT OF DATA FILTERING USING DWT

In order to see the effect of DWT on improving the performance, the raw upstream and downstream traffic volume data obtained from Minnesota DOT as well as the data filtered by DWT are shown in Figures 2 and 3, respectively. These figures show the incident and incident-free regions are more distinct after the data are filtered using DWT. This helps the neural network model classify the incident and incident regions more effectively resulting in better incident detection and low false alarm rates. Further,

this helps improve the convergence of the ACGNN learning model substantially, as shown in Figure 4.

5. RELATIVE CONTRIBUTION OF DWT AND LDA FOR FEATURE EXTRACTION

Our feature extraction model is a two-step algorithm consisting of DWT and LDA. In order to investigate their relative contribution in feature extraction, we also used DWT as the sole feature extractor. The results are shown in Table 5. A comparison of the data in Tables 4 and 5 indicate that most of the feature extraction capability is due to DWT. LDA has a smaller contribution toward improving the incident detection. One can say it has a fine tuning effect for reducing the false alarm rates.

6. EFFECTS OF FREEWAY GEOMETRY ON THE INCIDENT DETECTION

In order to show the efficacy and robustness of the new incident detection algorithm in various situations we performed a parametric study. To investigate the effect of various geometric changes on the incident detection algorithm, we used 65 incident test runs with minimum incident duration of 5 minutes and minimum traffic flow of 50% of the freeway capacity. Selected results of this study are presented here.

6.1 Effect of Curvature

Freeway geometric features such as grade, super-elevation, curvature, and pavement conditions do not affect the incident detection algorithm directly. They may have an indirect effect. For example, an incident on a curved freeway often causes more congestion than a similar incident on a straight segment. As a result, smaller duration incidents can cause sufficient congestion to get detected by the incident detection

algorithm. As an example, Figure 5 displays an instance 45 seconds after a simulated incident on a curved freeway. Comparing the results obtained for a curved freeway segment with those obtained for the straight freeway segment in Table 6, it is concluded that the curvature does not have an appreciable effect on the incident detection and false alarm rates of the incident detection model. However, the detection time for the curved segment is lower than that for the straight segment, because freeway gets congested faster.

6.2 Effect of Number of Lanes

The number of lanes in a freeway also affects the incident detection time and the detection rate of the incident detection algorithm. For similar incidents, having similar blockage characteristics as well as duration, the percentage changes in the traffic parameters are smaller for a larger freeway. Consequently, it takes more time to detect an incident as number of lanes increases. An example of an incident on a five-lane freeway is shown in Figure 6(a) to 6(c). Figure 6(a) shows the traffic pattern 45 seconds after the incident. Normally, this type of incident involving a lane blockage on a two-lane freeway (in one direction) gets detected within this time range. But for an incident on five-lane freeway (in one direction) two to four minutes may be required to detect the same. Figure 6(b) shows the traffic pattern 3 minutes after an incident. Figure 6(c) displays the incident characteristics.

For a small-duration incident the incident may not get detected. Thus, it affects the detection rate of an incident detection algorithm. The detection rate computed for a five-lane freeway is about 94% and the average detection time is 2 minutes and 47 seconds. The false alarm rate remains practically the same. Figure 7(a) and 7(b) show the

effect of the size of the freeway (number of lanes) on the incident detection rate and time for detection, respectively. It is observed that change in the detection rate and time is much higher for ACGNN using raw data than for the ACGNN using data filtered by DWT and LDA.

7. CONCLUSION

In this and the companion papers, we presented a robust incident detection computational model and algorithm through adroit integration of three different computational approaches/disciplines: signal processing and wavelet transform, statistical linear discriminant analysis, and artificial neural networks. For incidents with duration of more than five minutes, the algorithm yields a detection rate of nearly 100% and false alarm rate of about 1% for two- or three-lane and freeways. For incidents with duration of less than 5 minutes, the incident detection rate for two- or three-lane freeways is about 98% with a false alarm rate of about 1%.

For four-lane and five-lane freeways, the detection rate is reduced to 96% and 94%, respectively, but the false alarm rate remains around 1%. It is also observed that the freeway curvature does not affect the performance of the algorithm.

There is one type of incidents that the new algorithm cannot detect, that is the so-called *isolated* incident where there are no appreciable traffic volume and occupancy changes. Because the incident detection model considers the variations in traffic parameters obtained from loop detectors to detect the incidents. To detect the isolated incidents visual sensor input is needed.

ACKNOWLEDGMENT

This article is based on the research sponsored by Ohio Department of Transportation, Federal Highway Administration, and The Ohio State University Center for Intelligent Transportation Research. The authors are also grateful to Mr. George Saylor of Ohio Department of Transportation (DOT), Mr. Leonard Palek of Minnesota DOT, and Phil Carter of Arizona DOT for their help in obtaining actual traffic data and Dr. Henry Lieu of Federal Highway Administration for providing access to the traffic simulation package (TSIS/CORSIM) to test the computational models developed in this research.

REFERENCES

1. Adeli, H. and Hung, S.L., "An adaptive conjugate gradient algorithm for efficient training of neural networks", *Applied Mathematics and Computation*, Vol. 62, 1994, pp. 81-102.

2. Adeli, H. and Hung, S.L., *Machine Learning - Neural Networks, Genetic Algorithms, and Fuzzy System*, John Wiley, New York, 1995.

3. Adeli, H. and Park, H.S., *Neurocomputing in Design Automation*, CRC Press, Boca Raton, Florida, 1998.

4. Armijo, L., "Minimization of functions having Lipshitz-continuous first partial derivatives", *Pacific Journal of Mathematics*, Vol. 16, No. 1, 1966, pp. 1-3.

5. Cheu, R.L. and Ritchie, S.G., "Automated detection of lane-blocking freeway incidents using artificial neural networks", *Transportation Research* 3C(6), 1995, pp. 371-388.

6. Dennis, J.E., Jr., and Schnable, R.B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, NJ, 1983.

7. Fletcher, R. and Reeves, R.M., "Function minimization by conjugate gradients", *The Computer Journal*, Vol. 7, No. 2, 1964, pp. 149-160.

8. Goldstein, A.A., *Constructive Real Analysis*, Harper & Row, NY, 1967.

9. Hsiao, C.H., Lin, C.T., and Cassidy, M., "Application of fuzzy logic and neural networks to automatically detect freeway traffic incidents", *Journal of Transportation Engineering*, Vol. 120, No. 5, 1994, pp. 753-771.

10. Kollias, S. and Anastassiou, D., "An adaptive least-squares algorithm for the efficient training of artificial neural networks", *IEEE Transactions on Circuits and Systems*, Vol. 36, No. 8, 1989, pp. 1092-1101.

11. Nocedal, J., "The performance of several algorithms for a large scale unconstrained optimization", in Coleman T.F., and Li, Y., Eds., *Large-Scale Numerical Optimization*, Society for Industrial and Applied Mathematics, Philadelphia, 1990, pp. 138-151.

12. Powell, M.J.D., "Convergence properties of algorithms for nonlinear optimization", *SIAM Review*, Vol. 28, No. 4, 1986, pp. 487-500.

13. Rumelhart, D.E., Hinton, G.E., and Williams, R.J., "Learning internal representation by error propagation", in Rumelhart, D.E. et al., Eds. *Parallel Distributed Processing*, MIT Press, Cambridge, MA, 1986, pp. 318-362.

14. Samant, A. and Adeli, H., "Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis", submitted, 2000.

15. Stephanedes, Y.J., Chassiakos, A.P., and Michalopoulos, P.G., "Comparative performance evaluation of incident detection algorithms", *Transportation Research Records* 1360, TRB, National Research Council, Washington, DC, 1992, pp. 50-57.

16. Stephanedes, Y.J. and Liu, X., "Artificial neural networks for freeway incident detection", *Transportation Research Records* 1494, TRB, National Research Council, Washington, DC, 1995, pp. 91-97.

17. Zadeh, L., "Fuzzy set as a basis for a theory of possibility", *Fuzzy Sets and Systems*, Vol. 1, No. 1, NY, 1978, pp.3-28.

Table 1 Incident detection results using LDA only

Classifier	Incident detection rate (%)	False alarm Rate (%)	Mean time for detection
NMC (Nearest Mean)	62.2 (28/45)	14.9 (107/720)	73.2 sec.
FLD (Fisher LD)	68.9 (31/45)	13.8 (99/720)	69.5 sec.
Regularized FLD ($\delta = 20$)	71.1 (32/45)	13.1 (94/720)	69.1 sec.

Table 2 Incident detection results using DWT and LDA

Classifier	Incident detection rate (%)	False alarm rate (%)	Mean time for detection
NMC (Nearest Mean)	71.1 (32/45)	10.0 (72/720)	70.7 sec.
FLD (Fisher LD)	71.1 (32/45)	9.4 (68/720)	68.9 sec.
Regularized FLD ($\delta = 20$)	73.3 (33/45)	8.5 (61/720)	67.2 sec.

Table 3 Study of traffic parameters using ACGNN

Traffic Data	Incident detection rate (%)	False alarm rate (%)	Mean time for detection
Volume (vehicles/hr)	75.5 (34/45)	9.6 (69/720)	63.8 sec.
Occupancy (%)	71.1 (32/45)	9.3 (67/720)	59.7 sec.
Avg. Speed (miles/hr)	68.9 (31/45)	8.9 (64/720)	68.9 sec.
Vol. + Occupancy	88.9 (40/45)	5.1 (37/720)	51.4 sec.
Vol. + Avg. Speed	84.4 (38/45)	6.0 (43/720)	52.5 sec.
Vol. + Occupancy + Speed	91.1 (41/45)	5.1 (37/720)	47.6 sec.

Table 4 Incident detection results for a straight two-lane freeway segment using DWT, LDA and ACGNN

Input Data	Incident detection rate (%)	False alarm rate (%)	Mean time for detection
Using wavelet coefficients	97.8 (44/45)	1.0 (7/720)	38.9 sec.
Using re-generated traffic signals	97.8 (44/45)	1.1 (8/720)	38.1 sec.

Table 5 Incident detection results using DWT and ACGNN

Input Data	Incident detection rate (%)	False alarm rate (%)	Mean time for detection
Using wavelet coefficients	97.8 (44/45)	1.8 (13/720)	40.1 sec.
Using re-generated traffic signals	97.8 (44/45)	2.1 (15/720)	39.8 sec.

Table 6 Effect of curvature using DWT, LDA, and ACGNN

Input Data	Incident detection rate (%)	False alarm rate (%)	Mean time for detection
Two-lane straight freeway	100.0 (65/65)	1.2 (16/1300)	47.8 sec.
Two-lane curved freeway	100.0 (65/65)	1.4 (18/1300)	40.2 sec.

CAPTIONS FOR FIGURES

Figure 1 Artificial neural network for traffic incident detection problem

Figure 2 Comparison of raw and filtered data for upstream traffic (from Minnesota DOT)

Figure 3 Comparison of raw and filtered data for downstream traffic (from Minnesota DOT)

Figure 4 Convergence curve for ACGNN learning model for a two-lane straight freeway segment

Figure 5 An incident on a curved freeway

Figure 6 An incident on a five-lane freeway

Figure 7 Effect of size of a freeway (number of lanes) on the incident detection rate

Figure 8 Effect of size of a freeway (number of lanes) on the incident detection time

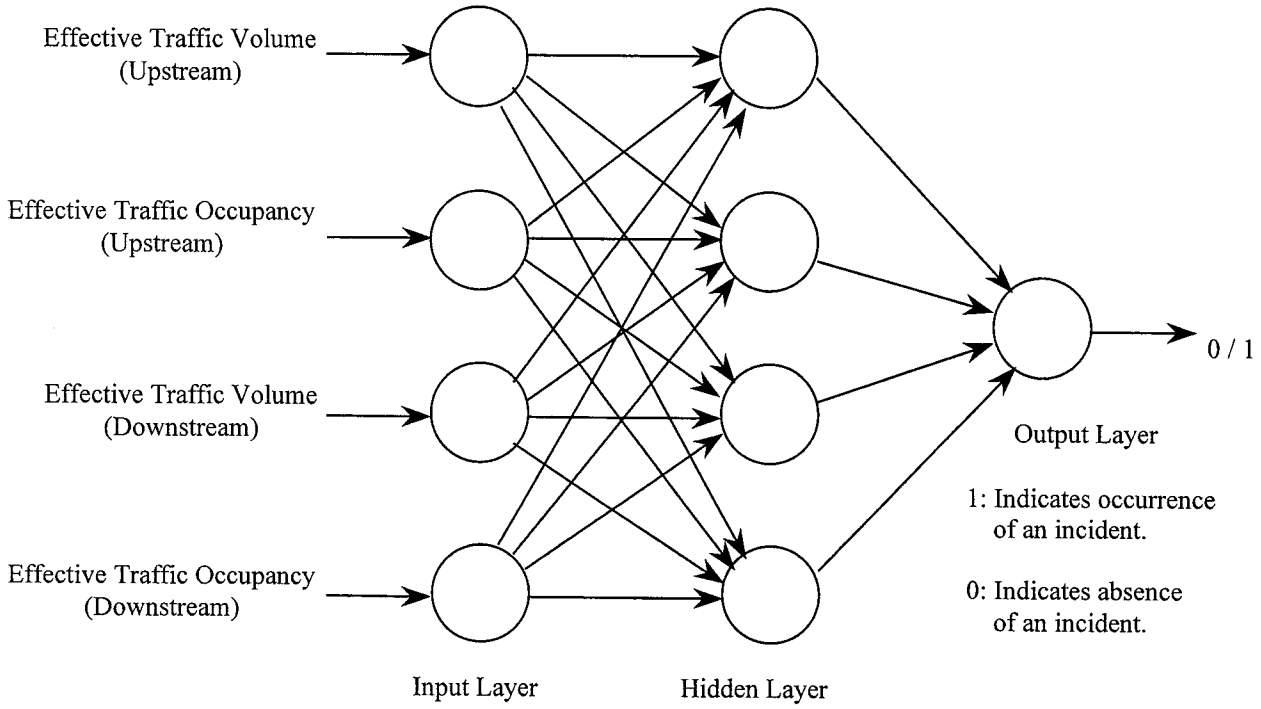

Effective Traffic Volume
(Upstream)
Effective Traffic Occupancy
(Upstream)
Effective Traffic Volume
(Downstream)
Effective Traffic Occupancy
(Downstream)
Input Layer
Hidden Layer
Output Layer
0 / 1
1: Indicates occurrence
of an incident.
0: Indicates absence
of an incident.

Figure 1

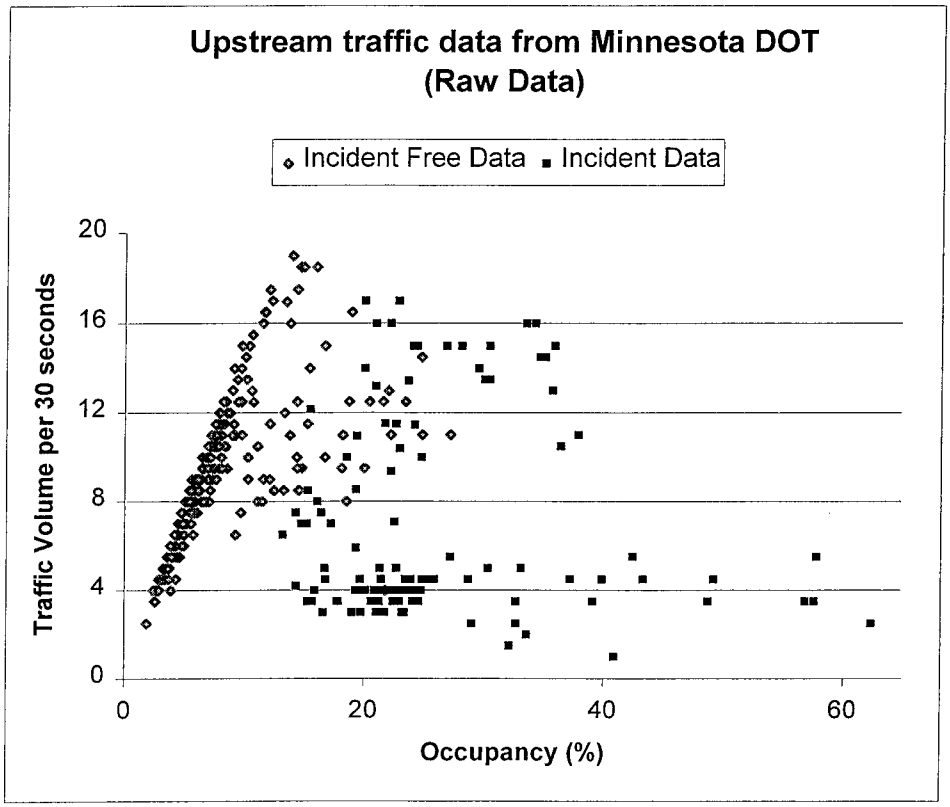

Upstream traffic data from Minnesota DOT
(Raw Data)
Incident Free Data
Incident Data
Traffic Volume per 30 seconds
20
16
12
8
4
0
0
20
40
60
Occupancy (%)

Figure 2(a)

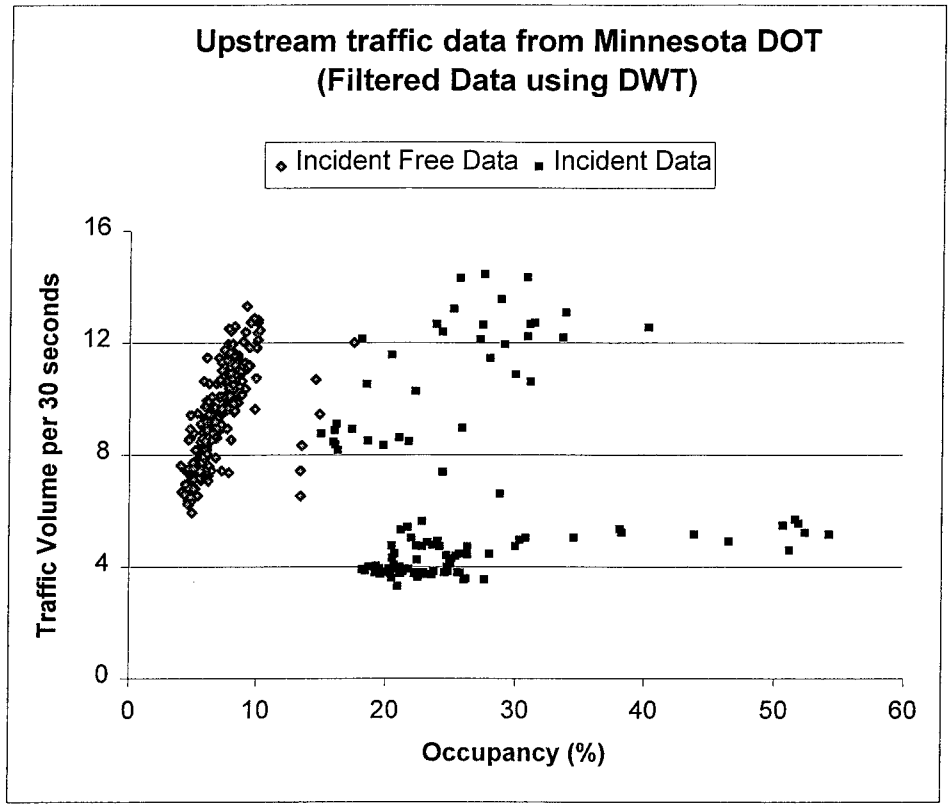

Upstream traffic data from Minnesota DOT
(Filtered Data using DWT)
Incident Free Data
Incident Data
Traffic Volume per 30 seconds
16
12
8
4
0
0
10
20
30
40
50
60
Occupancy (%)

Figure 2(b)

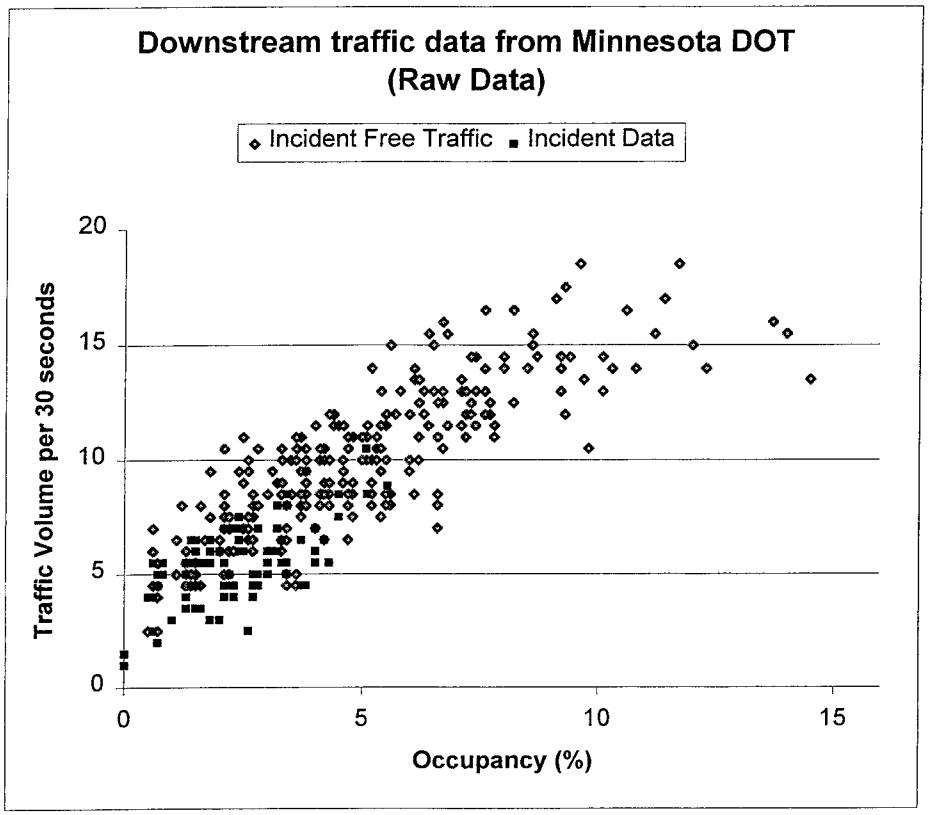

Downstream traffic data from Minnesota DOT
(Raw Data)
Incident Free Traffic
Incident Data
Traffic Volume per 30 seconds
20
15
10
5
0
0
5
10
15
Occupancy (%)

Figure 3(a)

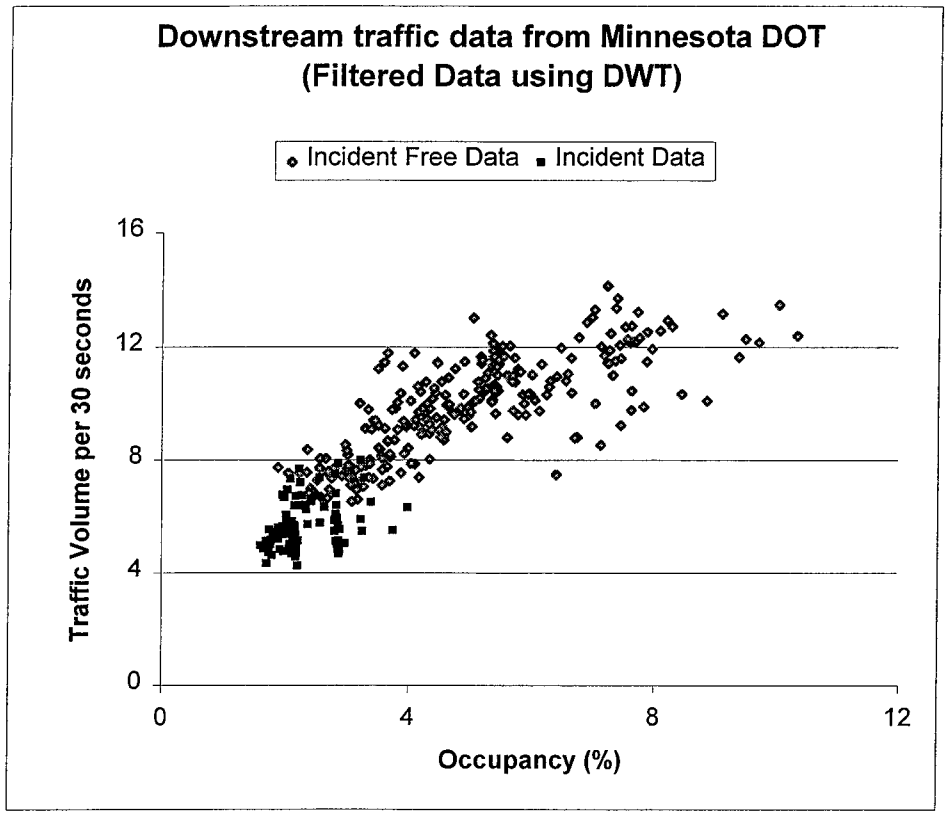

Downstream traffic data from Minnesota DOT
(Filtered Data using DWT)
Incident Free Data
Incident Data
Traffic Volume per 30 seconds
16
12
8
4
0
0
4
8
12
Occupancy (%)

Figure 3(b)

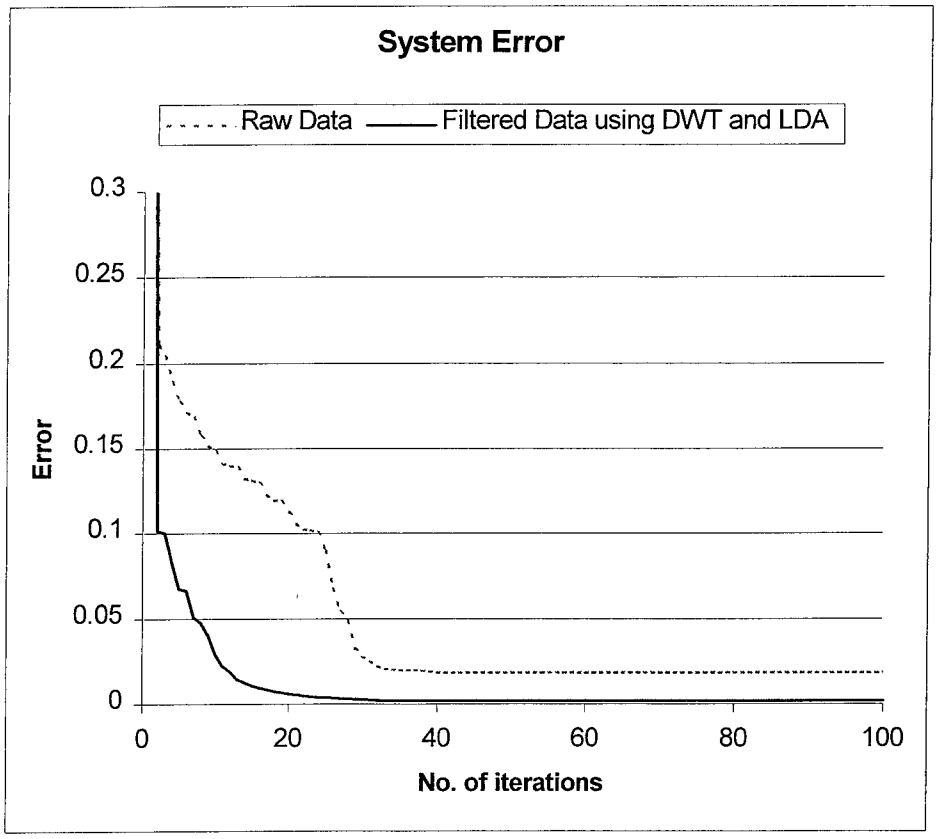

System Error
Raw Data
Filtered Data using DWT and LDA
Error
0.3
0.25
0.2
0.15
0.1
0.05
0
0
20
40
60
80
100
No. of iterations

Figure 4

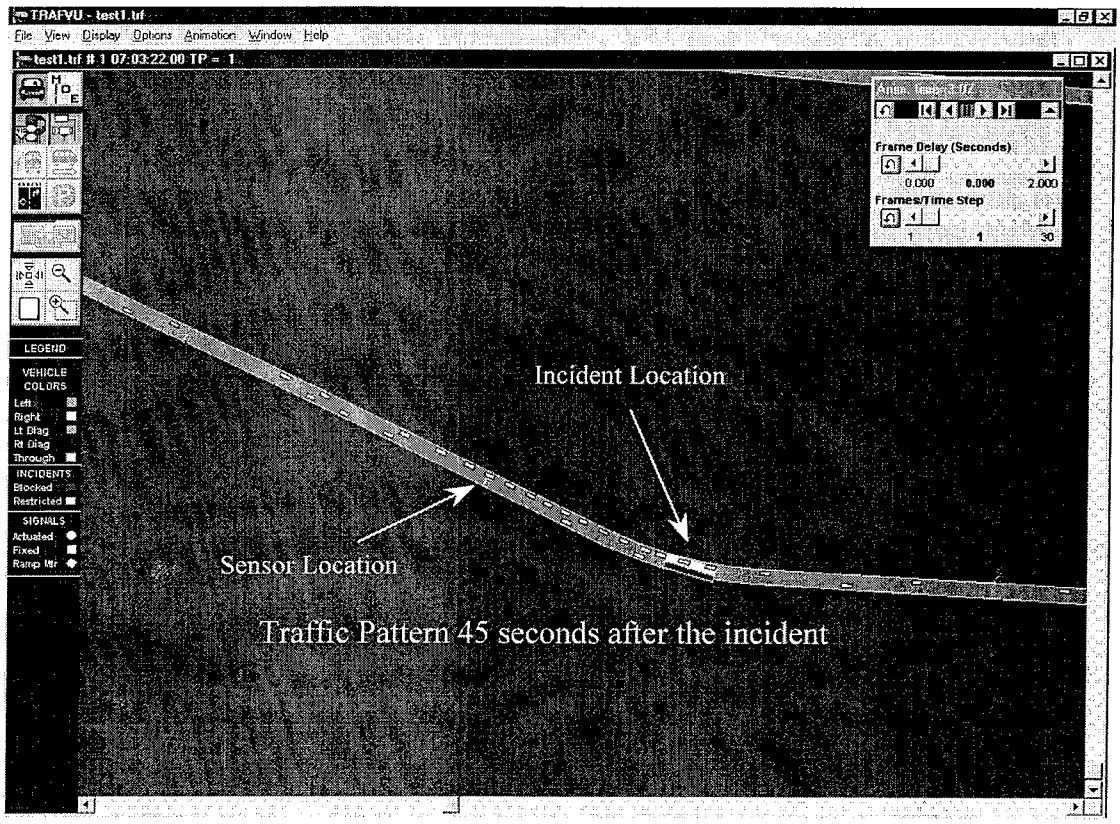

TRAFVU - test1.trf
File View Display Options Animation Window Help
test1.trf # 1 07:03:22.00 TP = 1
Frame Delay (Seconds)
0.000 0.000 2.000
Frames/Time Step
1 1 30
LEGEND
VEHICLE COLORS
Left
Right
Lt Diag
Rt Diag
Through
INCIDENTS
Blocked
Restricted
SIGNALS
Actuated
Fixed
Ramp Mtr
Incident Location
Sensor Location
Traffic Pattern 45 seconds after the incident

Figure 5

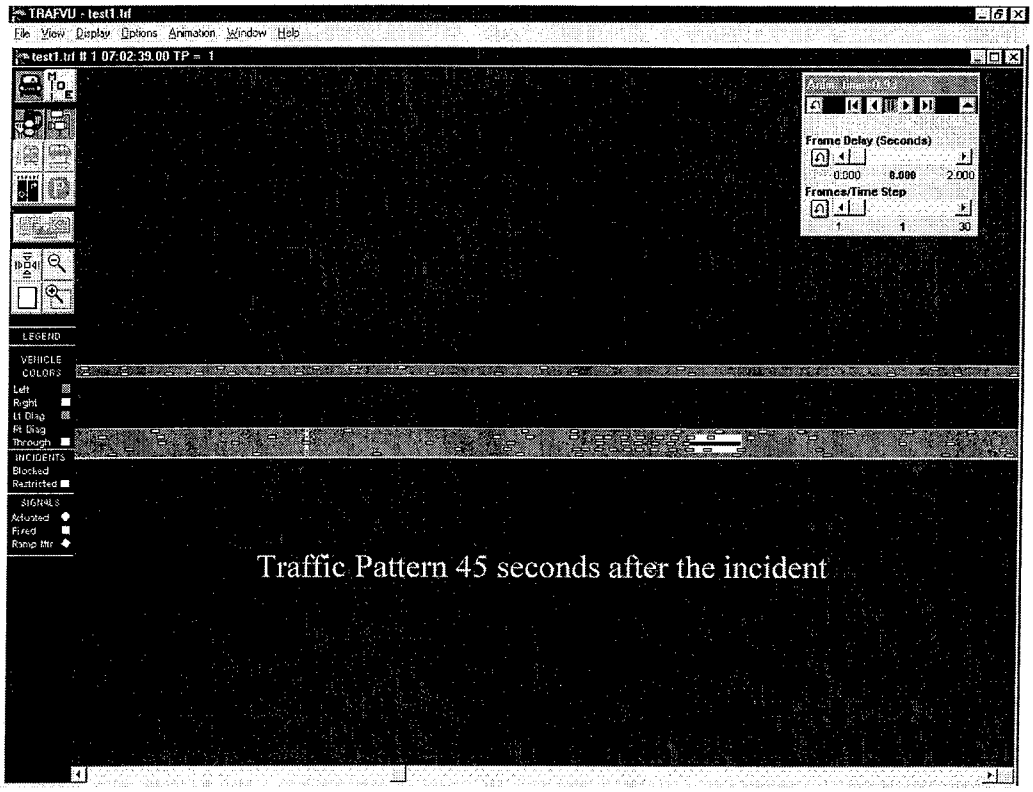

TRAFVU - test1.tif
File View Display Options Animation Window Help
test1.tif # 1 07:02:39.00 TP = 1
Frame Delay (Seconds)
0.000 0.000 2.000
Frames/Time Step
1 1 30
LEGEND
VEHICLE COLORS
Left
Right
Lt Diag
Rt Diag
Through
INCIDENTS
Blocked
Restricted
SIGNALS
Actuated
Fixed
Ramp Mtr
Traffic Pattern 45 seconds after the incident

Figure 6(a)

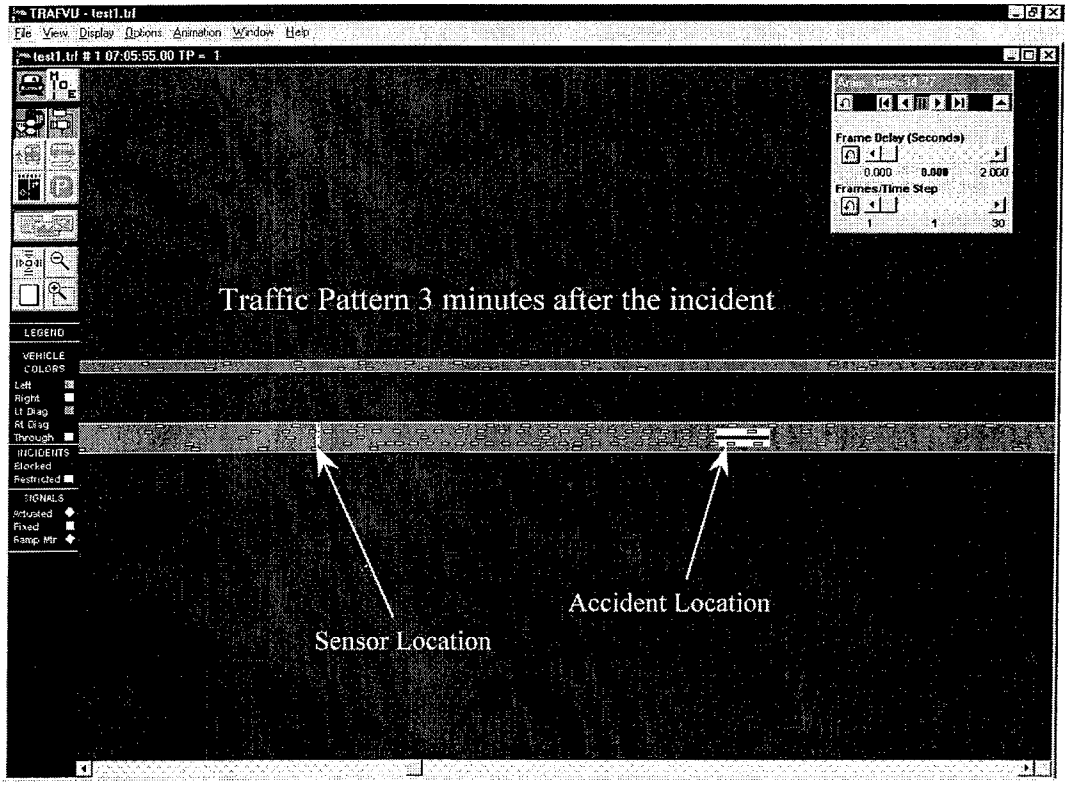

TRAFVU - test1.tsf
File View Display Options Animation Window Help
test1.tsf # 1 07:05:55.00 TP = 1
Frame Delay (Seconds)
0.000 0.000 2.000
Frames/Time Step
1 1 30
Traffic Pattern 3 minutes after the incident
LEGEND
VEHICLE COLORS
Left
Right
Lt Diag
Rt Diag
Through
INCIDENTS
Blocked
Restricted
SIGNALS
Actuated
Fixed
Ramp Mtr
Accident Location
Sensor Location

Figure 6(b)

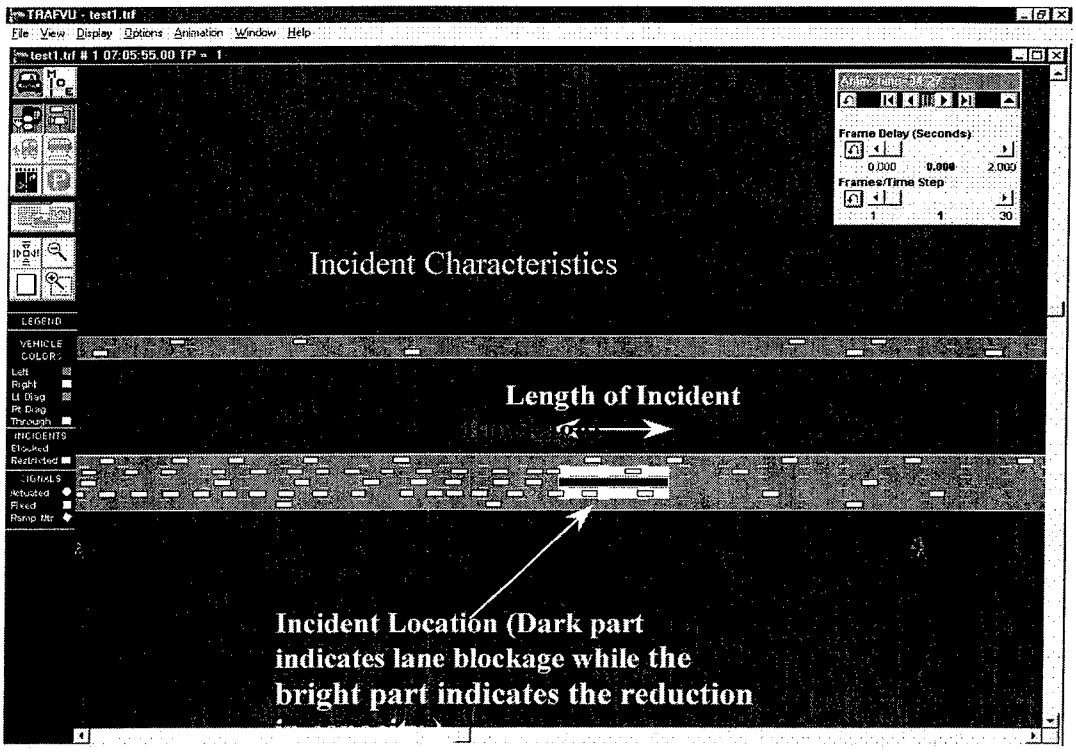

TRAFVU - test1.trf
File View Display Options Animation Window Help
test1.trf # 1 07:05:55.00 TP = 1
Frame Delay (Seconds)
0.000 0.000 2.000
Frames/Time Step
1 1 30
Incident Characteristics
LEGEND
VEHICLE COLORS
Left
Right
Lt Diag
Rt Diag
Through
INCIDENTS
Blocked
Restricted
SIGNALS
Actuated
Fixed
Ramp Mtr
Length of Incident
Incident Location (Dark part indicates lane blockage while the bright part indicates the reduction

Figure 6(c)

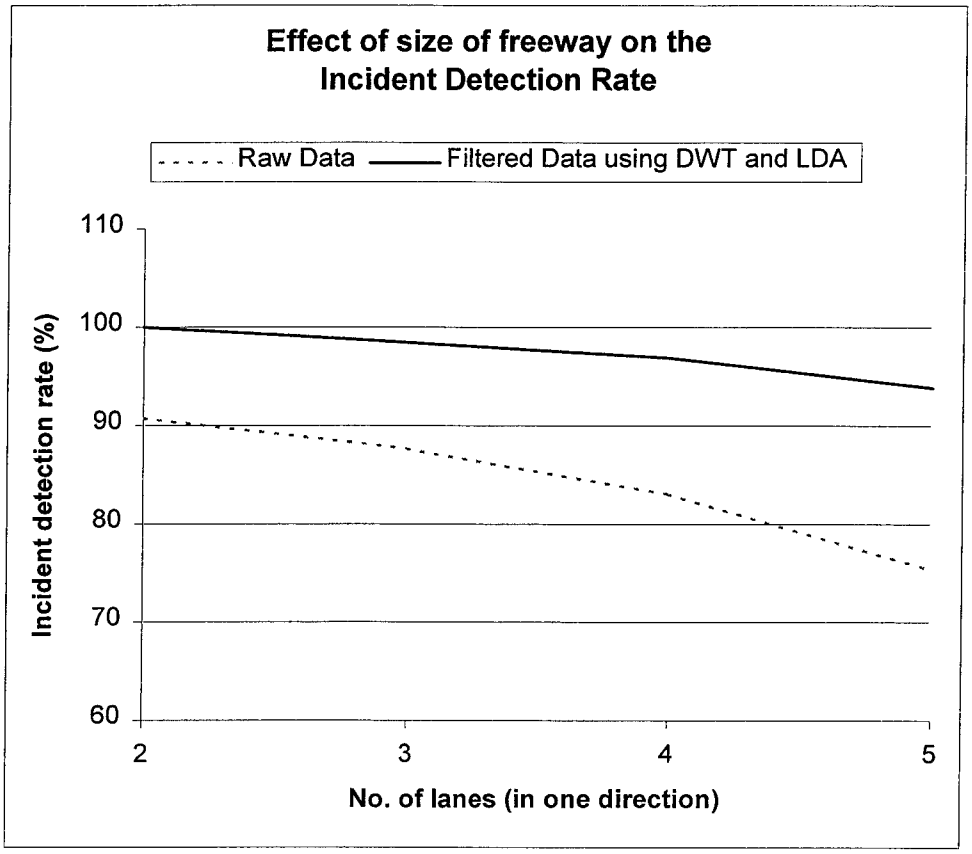

Effect of size of freeway on the Incident Detection Rate
Raw Data
Filtered Data using DWT and LDA
Incident detection rate (%)
110
100
90
80
70
60
2
3
4
5
No. of lanes (in one direction)

Figure 7(a)

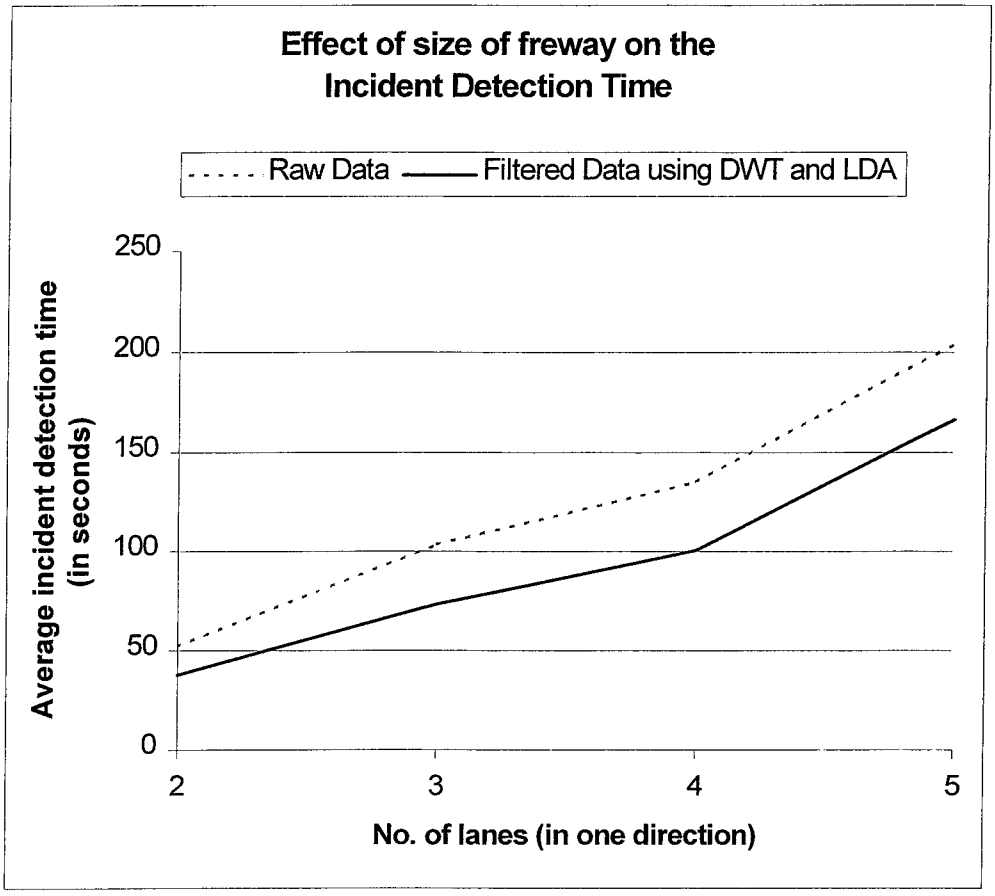

Effect of size of freway on the Incident Detection Time
Raw Data
Filtered Data using DWT and LDA
Average incident detection time (in seconds)
250
200
150
100
50
0
2
3
4
5
No. of lanes (in one direction)

Figure 7(b)

Part 3

ENHANCING NEURAL NETWORK TRAFFIC INCIDENT DETECTION ALGORITHMS USING WAVELETS

A. Samant[1] and H. Adeli[2]

ABSTRACT: Researchers have presented freeway traffic incident detection algorithms by combining the adaptive learning capability of neural networks with imprecision modeling capability of fuzzy logic. In this article it is shown that the performance of a fuzzy neural network algorithm can be improved through preprocessing of data using a wavelet based feature extraction model. In particular, the discrete wavelet transform (DWT) de-noising and feature extraction model proposed by Samant and Adeli (2000) is combined with the fuzzy-neural network approach presented by Hsiao et al. (1994). It is shown that substantial improvement can be achieved using the data filtered by DWT. Use of the wavelet theory to de-noise the traffic data increases the incident detection rate, reduces the false alarm rate and the incident detection time, and improves the convergence of the neural network training algorithm substantially.

---

[1]Graduate Research Associate, [2]Professor, Dept. of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, 470, Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210.

1. INTRODUCTION

Fuzzy logic was created by Zadeh (1978) to model the imprecision or uncertainty involved in the human decision-making process. For modeling imprecision a fuzzy logic system usually consists of three components:

1. *Input fuzzification.* It transforms the input values to the *degree of membership* values using linguistic rules and the concept of membership function.
2. *Knowledge or inference rule base.* It is a collection of inference or heuristic rules. In the freeway traffic incident detection problem, an example of an inference rule is

   If (traffic volume is high AND occupancy is low AND average vehicle speed is high) THEN occurrence of incident is impossible.

   Based on the degree of membership values for the input variables (traffic volume, occupancy, and the vehicle speed) in the antecedents of all the applicable rules, degrees of membership are computed for the consequences of the rules (incident or no-incident) using fuzzy operations such as fuzzy-AND and fuzzy-OR.
3. *Output defuzzification.* It takes the output from the inference rule base and defuzzify it to produce the output variable for decision making.

A few researchers have presented freeway traffic incident detection algorithms by combining the adaptive learning capability of neural networks with fuzzy logic. Hsiao et al. (1994) present a Fuzzy Logic Incident Patrol System (FLIPS) for the freeway traffic incident problem by treating threshold as variable and finding its values using fuzzy logic rules and membership functions. They use the simple backpropagation (BP) neural network learning rule (Rumelhart et al., 1986, Adeli and Hung, 1995) and bell-shaped

membership functions. They test the model using an "empirical data base collected in Toronto, Canada". They report detection rates in the range of 54% (with a false alarm rate of 0%) to 90% (with a false alarm rate of 7.9 %).

A main reason for unreliability of the traffic incident detection algorithms is the noise in the traffic data. In other words, the traffic data are often corrupted as they are collected by sensors and then transmitted to a central processing station. To eliminate false alarms an effective traffic incident detection algorithm must be able to extract features from the traffic patterns, which are related to the incident. A robust feature extraction algorithm also helps reduce the dimension of the input space for a neural network model without any significant loss of related traffic information, resulting in a substantial reduction in the network size, effect of random traffic fluctuations, number of required training samples, and computational resources required to train the neural network.

Samant and Adeli (2000) present an effective traffic data de-noising and feature extraction model using discrete wavelet transform (DWT) and linear discriminant analysis. The DWT is first applied to raw traffic data and the finest resolution coefficients representing the random fluctuations of traffic are discarded. Next, LDA is employed to the filtered signal for further feature extraction and reducing the dimensionality of the problem. The results of LDA are used as input to a neural network model for traffic incident detection.

In this article it is shown that the performance of a fuzzy neural network algorithm can be improved through preprocessing of data using a wavelet-based feature extraction model. In particular, the DWT de-noising and feature extraction model

proposed by Samant and Adeli (2000) is combined with the fuzzy-neural network approach presented by Hsiao et al. (1994). It is shown that substantial improvement can be achieved using the data filtered by DWT.

2. DISCRETE WAVELET TRANSFORM

The wavelet transform is found to be an effective tool in signal and image processing due to its attractive properties such as time-frequency and multi-resolution analysis (Daubechies, 1992; Jameson et al., 1996; Mallat, 1998). Using these properties one can extract the desired features from an input signal characterized by certain local properties in time and space. A feature extraction approach using wavelet transform is used to achieve higher level of accuracy in the decision making process by the fuzzy neural network algorithm. The details of the feature extraction model for the traffic incident detection problem are presented in Samant and Adeli (2000). The basic idea is briefly described here in non-mathematical terms.

We view the traffic flow as a signal, with traffic incidents as well as other traffic patterns such as traffic bottleneck or compression wave having different time-space properties. We use the wavelet transform to extract the specific features distinguishing such traffic patterns as it can extract features from different time scales having different resolutions quite effectively.

For the traffic incident detection problem, we consider various traffic data recorded at a fixed time interval as input for DWT. Each of these data series can be represented by x[j], where $j \in Z$ and Z is a set of integers (square brackets represent a

series or a sequence and circular brackets represent functions). The output of the DWT consists of the coordinates $\beta_l[l]$ and $\lambda_j[l]$ of the orthonormal wavelet bases

$$\beta_l[l] = \left\langle x[k], \phi_{l,l}[k] \right\rangle \quad \text{and} \quad l = \frac{k}{2^l} \tag{1}$$

$$\lambda_j[l] = \left\langle x[k], \psi_{j,l}[k] \right\rangle \; ; \; l = \frac{k}{2^j} \text{ and } j = 1, 2, \ldots, I \tag{2}$$

where $\langle \; \rangle$ denotes the inner product of the two sequences, k represents the total number of input data points, l represents the number of coefficients of each data series such as traffic volume or occupancy, $\{\phi_{l,l}\}_{l \in Z}$ and $\{\psi_{j,l}\}_{l \in Z}$ (the brackets $\{\}$ denote a set of series) represent scaling and wavelet functions (Daubechies, 1992; Farge et al., 1993), respectively, and I is a positive integer. We use Daubechies wavelet function as it is found to be quite effective in digital signal processing. The value of I is chosen such that the desired level of resolution is obtained and $j = 1, 2, \ldots, I$. The coordinates β and λ are in fact, low and high-resolution coefficients of the given data series $x[k]$, respectively. The coordinates of the wavelet bases (βs and λs) are computed using a concept called the quadrature mirror filters (Wickerhauser, 1994).

To extract the traffic incident pattern from the traffic data we perform multi-resolution analysis of the wavelet transforms of traffic patterns. Multi-resolution analysis involves dividing the original *signal* (e.g. traffic volume or occupancy) into *signals* having different frequencies and time localizations and analyzing the signal in different scales.

3. ARCHITECTURE

The architecture of the enhanced incident detection model is represented schematically in Figure 1. The learning block shown in this figure makes use of the fuzzy logic rules and guides the fuzzification and the defuzzification blocks to learn the membership functions. Different types of membership functions such as triangular, trapezoidal, and bell-shaped have been used to solve various problems. We employ the same bell-shaped membership function used by Hsiao et al. (1994) in the following form:

$$v_j = e^{\left(-\frac{(u_i - m_{ij})^2}{\sigma_{ij}^2}\right)} \tag{3}$$

where u_i is the i^{th} input variable, m_{ij} and σ_{ij} are the mean and the variance of the j^{th} fuzzy set of the i^{th} input variable, and v_j is the output of the input membership function providing degree of membership for the j^{th} fuzzy set. Three fuzzy sets are defined for each one of the three traffic variables as: low, medium, and high.

The fuzzy wavelet neural network for the incident detection problem consists of an input layer with three nodes representing traffic volume, occupancy, and vehicle speed and an output layer with a single node with two output states representing incident and no-incident (Figure 2). The traffic data are de-noised using DWT. The filtered data from the input layer is transferred to an input membership function layer where each node calculates the degree of membership of input data in each one of the three predefined fuzzy sets (for three linguistic variables low, medium, and high). The output of the input membership function layer varies from 0 to 1, with 0 indicating no membership and 1 indicating full membership.

The degree of membership values are then passed to a knowledge or inference rule base. Each fuzzy rule in the rule base performs a fuzzy-AND operation to produce the feasible range, w_k, for the k^{th} fuzzy rule as follows:

$$w_k = \min(v_i) \qquad i = 1, 2, 3 \qquad k = 1, 2, 3, \ldots, 27 \tag{4}$$

where the input v_i is the degree of membership value for the i^{th} fuzzy set obtained from the input membership function layer. The inference rule base contains twenty-seven fuzzy logic rules given by Hsiao et al. (1994). After obtaining the feasible range for every rule, the rules having the same consequences, for example, possible occurrence of an incident, are combined using a fuzzy-OR operation as follows:

$$x_i = \min(1, \sum_k w_k) \qquad i=1,2 \tag{5}$$

where x_i is the degree of membership value for the i^{th} output membership function. Using these degree of membership values a crisp output is obtained by using a "center of area" defuzzification method (Lee, 1990) as follows:

$$o = \frac{\sum_i (m_i' \sigma_i') x_i}{\sum_i \sigma_i' x_i} \qquad i=1,2 \tag{6}$$

where m_i' and σ_i' are the means and the variances of the i^{th} output membership function for the output variable. The value of the output variable, o, can vary between 0 and 1. A value less than 0.5 indicates the state of no-incident and a value equal or greater than 0.5 indicates the state of incident occurrence.

4. TRAINING OF THE NETWORK

The training of the network requires finding the means ($m_{ij}, m_i^{'}$) and variances ($\sigma_{ij}, \sigma_i^{'}$) of the input and output membership functions. Shortcomings of the BP learning algorithm such as very slow rate of learning and trial-and-error problem-dependent selection of learning and momentum ratios have been discussed in the recent literature (Adeli and Hung, 1994). Since the objective of this article is to demonstrate how a fuzzy neural network incident detection model can be improved through a DWT feature extraction model we use the same feed forward BP learning rule used by Hsiao et al. (1994) to train the neural network.

The training is initialized by providing the desired initial ranges of input and output fuzzy partitions in the form of means and variances of the membership functions. For example, for occupancy initial mean values of 0%, 50%, and 100% are provided for the three linguistic variables low, medium and high with a variance value of 30% for each one. The initialization is done such that the linguistic variable covers the feasible region of the corresponding input/output space uniformly (Hsiao et al., 1994).

After the initialization the mean and variance values are obtained by minimizing an error function in the following form:

$$E = \frac{1}{2}(y - o)^2 \tag{7}$$

where y = desired output and o = computed output. The error is back propagated and the deltas (Δm, $\Delta\sigma$) for the output and input membership function parameters are calculated as follows (Rumelhart, 1986; Hsiao et al., 1994):

$$\Delta m = -\eta \frac{\partial E}{\partial m} \qquad \text{and} \qquad \Delta\sigma = -\eta \frac{\partial E}{\partial \sigma} \tag{8}$$

where η is the so-called learning rate parameter. Using Eqs. (6), (7) and (8) the means of the output membership functions are updated as follows:

$$\Delta m_i^{'(n)} = \eta \left(y^{(n)} - o^{(n)} \right) \frac{\sigma_i' u_i}{\sum \sigma_i' u_i} \tag{9}$$

$$m_i^{'(n+1)} = m_i^{'(n)} + \Delta m^{'(n)} \tag{10}$$

where superscript n is the iteration counter. Similarly the variances of the output membership functions are updated. The error in the evaluation of the output membership functions is back-propagated to the input membership function layer through inference rule base by a rule matching process and the input membership function parameters are updated in a similar way.

5. FILTERING OF TRAFFIC DATA USING DWT

The raw traffic data is obtained through simulation of freeway traffic flow using the TSIS/CORSIM simulation package (http://www.fhwa-tsis.com). The traffic flow parameters (traffic volume, occupancy and vehicle speed) are recorded at 30-second intervals. DWT is then applied to each of the traffic data series separately. Eight-minute traffic patterns yielding 16 data points are used at a time for the filtering process. DWT divides the signal into two parts: high-resolution signal and low-resolution signal. Thus, a

single stage DWT produces 8 high-resolution data points and 8 low-resolution data points. The high-resolution data points are discarded as they mainly represent the random fluctuations in the traffic. DWT is again applied to the remaining 8 low-resolution data points to obtain 4 medium-resolution and 4 low-resolution data points. The traffic signal is then regenerated using these medium and low-resolution data points which carry the incident related information. This process is called multi-resolution analysis (MRA) as it extracts the signals having different resolutions. The new filtered signal is used as a direct input to the fuzzy-neural network. The linear discriminant analysis used in Samant and Adeli (2000) is not needed here for feature extraction as the means and variances of the traffic data are incorporated in the form of membership function parameters of the fuzzy sets.

6. INCIDENT DETECTION RESULTS

The fuzzy wavelet neural network is trained using the data obtained from 32 simulation runs, 25 of which include an incident. The network was then tested using 45 new simulated lane-blocking incidents on freeways with different number of lanes.

Figures 3a to 3c show the learned membership functions for traffic volume, occupancy and vehicle speed, respectively, for a two-lane freeway (in one direction) using the fuzzy wavelet neural network. We obtained similar curves when the data was not filtered by DWT.

Table 1 shows the incident detection results for a two-lane freeway (in one direction) using the fuzzy wavelet neural network model as well the fuzzy neural network model of Hsiao et al. (1994). Use of the wavelet theory to de-noise the traffic data

increases the incident detection rate from 86.7% to 97.8%, reduces the false alarm rate from 5.3% to 1.8%, and reduces the incident detection time from 63.6 second to 48.9 seconds. Figure 4 shows the training convergence curve with and without DWT. It is observed that use of DWT improves the convergence of the training algorithm substantially.

Figures 5 and 6 show the effects of the size of the freeway (number of lanes) on the incident detection rate and time for detection, respectively. In general, the rate of detection reduces and the detection time increases with an increase in the number of lanes. Preprocessing of the traffic data by DWT, however, improves the performance of the algorithm substantially. Figure 7 shows the false alarm rate as a function of the number of lanes. The false alarm rate changes little with the number of lanes. It is 1.5-2.2% for the fuzzy wavelet neural network model and 4.9-6.1% for the fuzzy neural network model.

ACKNOWLEDGMENT

This article is based on the research sponsored by Ohio Department of Transportation, Federal Highway Administration, and The Ohio State University Center for Intelligent Transportation Research. The authors are also grateful to Mr. George Saylor of Ohio Department of Transportation (DOT), Mr. Leonard Palek of Minnesota DOT, and Phil Carter of Arizona DOT for their help in obtaining actual traffic data and Dr. Henry Lieu of Federal Highway Administration for providing access to the traffic

simulation package (TSIS/CORSIM) to test the computational models developed in this research.

REFERENCES

Adeli, H. and Hung, S.L. (1994), "An adaptive conjugate gradient algorithm for efficient training of neural networks", *Applied Mathematics and Computation*, Vol. 62, pp. 81-102.

Adeli, H. and Hung, S.L. (1995), *Machine Learning - Neural Networks, Genetic Algorithms, and Fuzzy System*, John Wiley, New York.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA.

Farge, M., Hunt, J.C.R., and Vassilicos, J.C. (1993), Ed., *Wavelets, Fractals, and Fourier Transform*, The institute of Mathematics and its Applications conference series, Clarendon Press, Oxford.

Hsiao, C.H., Lin, C.T., and Cassidy, M. (1994), "Application of fuzzy logic and neural networks to automatically detect freeway traffic incidents", *Journal of Transportation Engineering*, Vol. 120, No. 5, pp. 753-771.

Jameson, L.M., Hussaini, M.Y., Erlebacher, M. (1996), Ed., *Wavelets: Theory and Applications*, ICASE/LaRC Series in Computational Science and Engineering, Oxford University Press, NY.

Lee, C.C. (1990), "Fuzzy logic in control systems: Fuzzy logic controller part I & II", *IEEE trans. Systems, Man, and Cybern.*, SMC-20(2), 404-435.

Mallat, S.G. (1998), *Wavelet Tour of Signal Processing*, Academic Press, London.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986), "Learning internal representation by error propagation", in Rumelhart, D.E. et al., Eds. *Parallel Distributed Processing*, MIT Press, Cambridge, MA, pp. 318-362.

Samant, A. and Adeli, H. (2000), "Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis", *Computer-Aided Civil and Infrastructure Engineering*, submitted.

Wickerhauser, M.V. (1994), *Adapted Wavelet Analysis from Theory to Software*, A K Peters, Ltd. Wellesley, MA.

Zadeh, L.A., (1978) "Fuzzy set as a basis for a theory of possibility", *Fuzzy Sets and Systems*, Vol. 1, No. 1, NY, pp.3-28.

Table 1 Comparative performance of the incident detection models

Incident Detection Model	Incident detection Rate (%)	False Alarm Rate (%)	Mean time for detection
Fuzzy neural network (Hsiao et al., 1994)	86.7 (39/45)	5.3 (38/720)	63.6 sec.
Fuzzy wavelet neural network	97.8 (44/45)	1.8 (13/720)	48.9 sec.

CAPTIONS FOR FIGURES

Figure 1 Architecture of a fuzzy wavelet incident detection system

Figure 2 Enhanced fuzzy wavelet neural network

Figure 3 Membership functions for traffic volume, occupancy and average vehicle speed

Figure 4 Convergence curves for incident detection algorithms

Figure 5 Incident detection rate versus the number of lanes (in one direction)

Figure 6 Incident detection time versus the number of lanes (in one direction)

Figure 7 False alarm rate versus the number of lanes (in one direction)

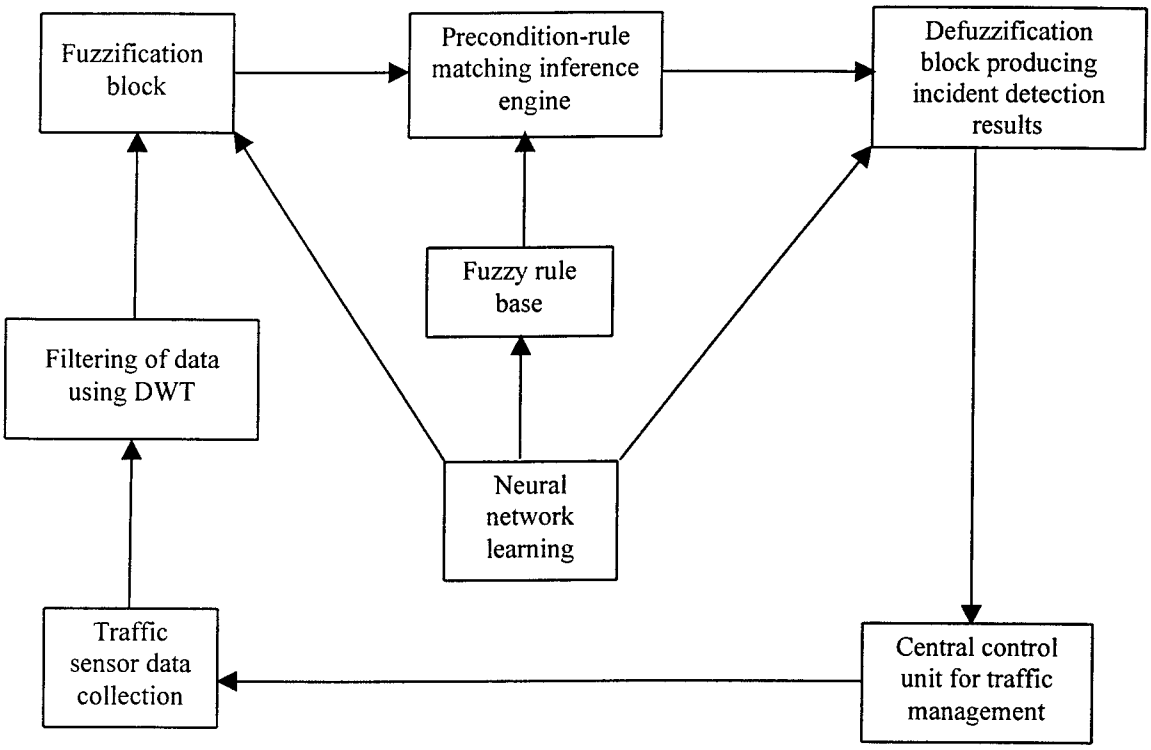


Figure 1 Architecture of a fuzzy wavelet incident detection system

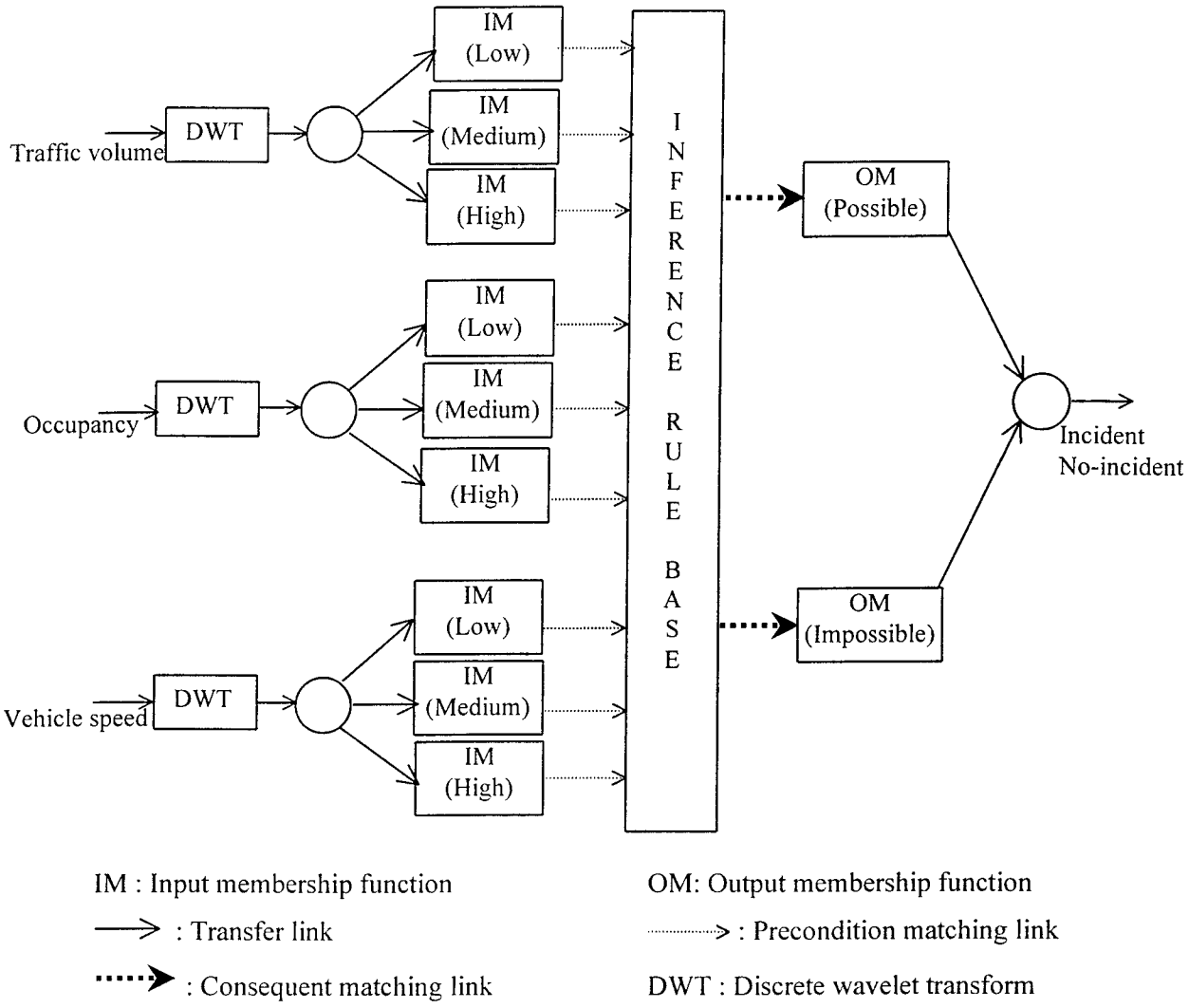


Figure 2 Enhanced fuzzy wavelet neural network

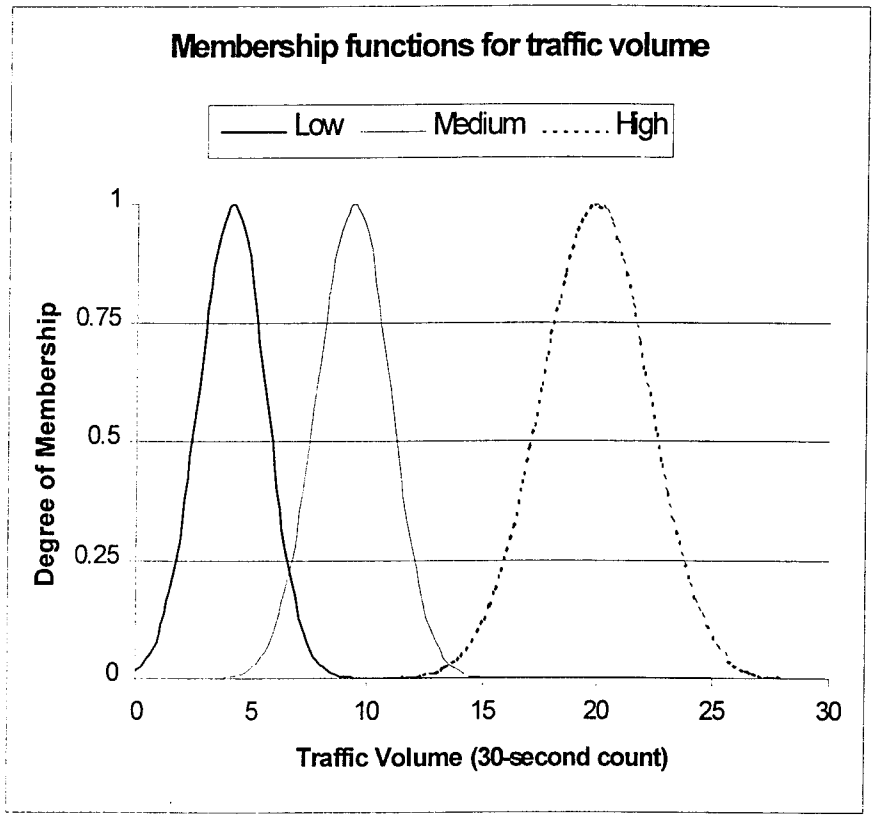

Membership functions for traffic volume
Low
Medium
High
Degree of Membership
1
0.75
0.5
0.25
0
0
5
10
15
20
25
30
Traffic Volume (30-second count)

Figure 3a

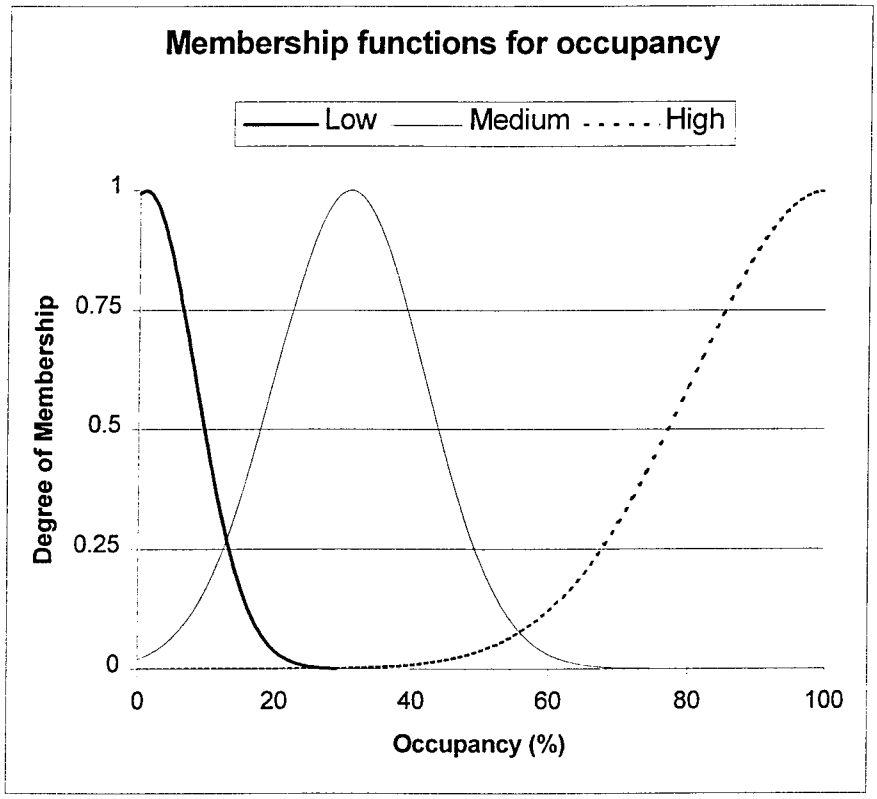

Membership functions for occupancy
Low
Medium
High
Degree of Membership
1
0.75
0.5
0.25
0
0
20
40
60
80
100
Occupancy (%)

Figure 3b

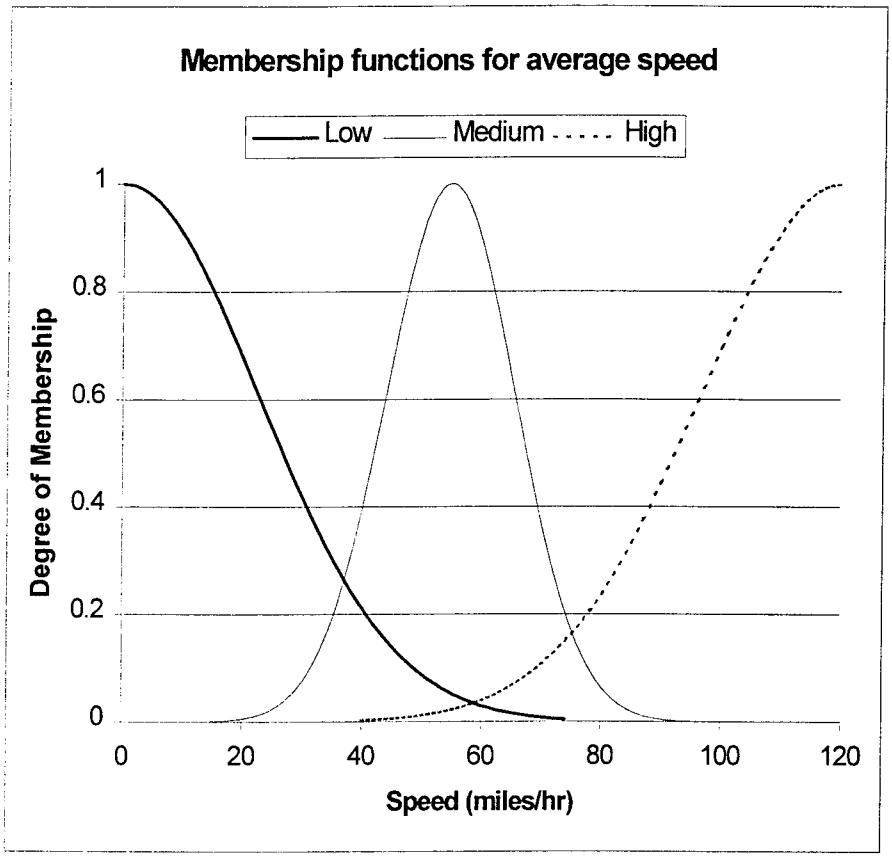

Membership functions for average speed
Low
Medium
High
Degree of Membership
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
Speed (miles/hr)

Figure 3c

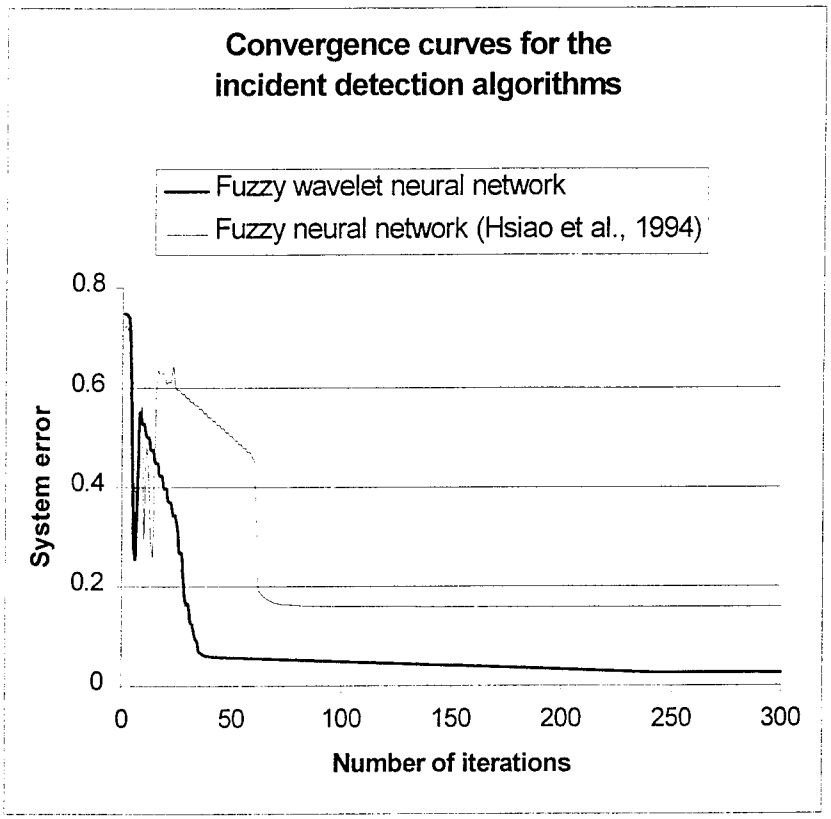

Convergence curves for the incident detection algorithms
Fuzzy wavelet neural network
Fuzzy neural network (Hsiao et al., 1994)
System error
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
300
Number of iterations

Figure 4

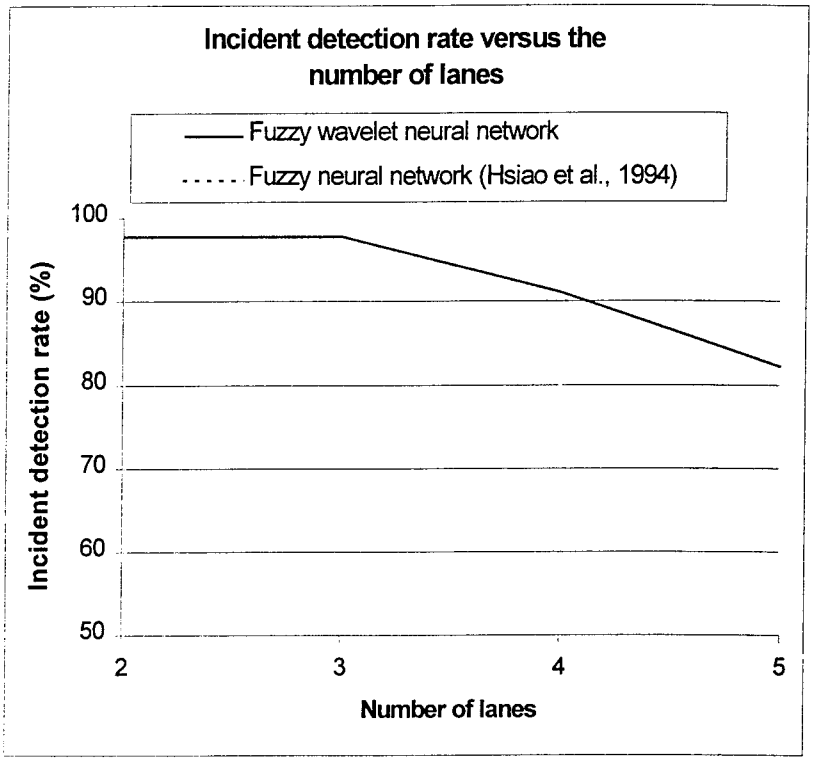

Incident detection rate versus the number of lanes
Fuzzy wavelet neural network
Fuzzy neural network (Hsiao et al., 1994)
Incident detection rate (%)
100
90
80
70
60
50
2
3
4
5
Number of lanes

Figure 5

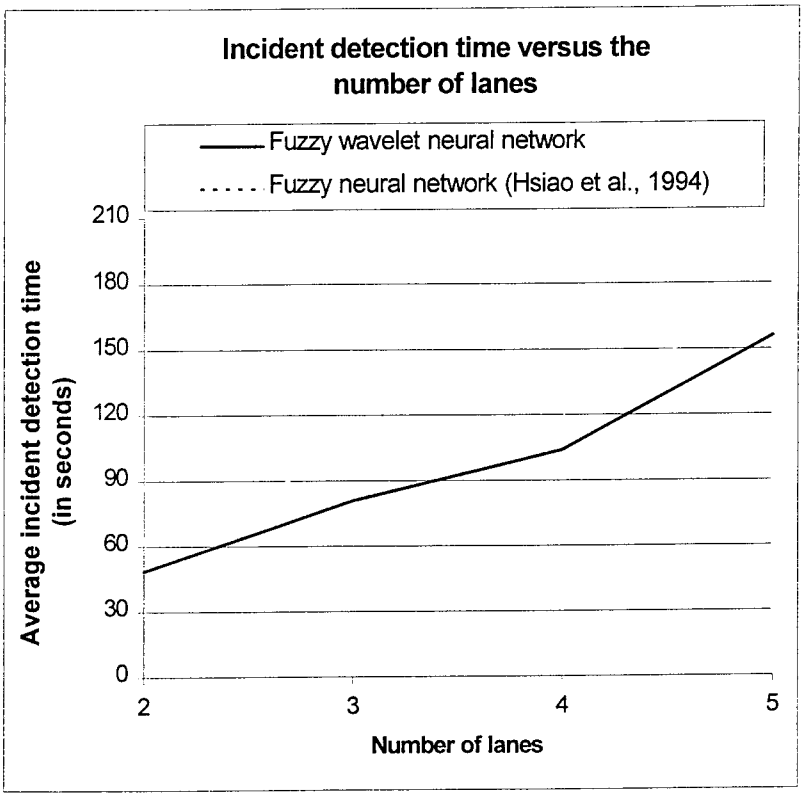

Incident detection time versus the number of lanes
Fuzzy wavelet neural network
Fuzzy neural network (Hsiao et al., 1994)
Average incident detection time (in seconds)
210
180
150
120
90
60
30
0
2
3
4
5
Number of lanes

Figure 6

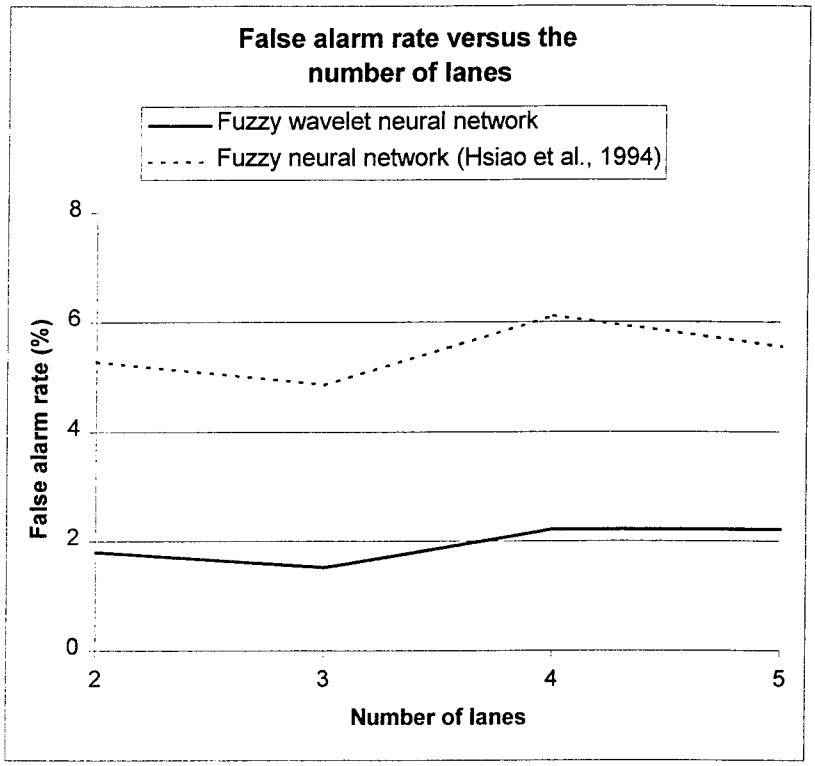

False alarm rate versus the number of lanes
Fuzzy wavelet neural network
Fuzzy neural network (Hsiao et al., 1994)
False alarm rate (%)
8
6
4
2
0
2
3
4
5
Number of lanes

Figure 7

Part 4

A FUZZY-WAVELET RBF NEURAL NETWORK MODEL FOR FREEWAY INCIDENT DETECTION

Hojjat Adeli[1] and Asim Karim[2]

ABSTRACT: Traffic incidents are non-recurrent and pseudo-random events that disrupt the normal flow of traffic and create a bottleneck in the road network. The probability of incidents is higher during peak flow rates when their system wide impact is most severe. Model-based solutions to the incident detection problem have not produced practically useful results primarily because the complexity of the problem does not lend itself to accurate mathematical and knowledge-based representations. A new multi-paradigm intelligent system approach is presented for the solution of the problem employing advanced signal processing, pattern recognition, and classification techniques. The methodology effectively integrates fuzzy, wavelet, and neural computing techniques to improve reliability and robustness. A wavelet-based de-noising technique is employed to eliminate undesirable fluctuations in observed data from traffic sensors. Fuzzy c-mean clustering is used to extract significant information from the observed data and to reduce its dimensionality. A radial basis function neural network is developed to classify the de-noised and clustered observed data. The new model produced excellent incident detection rates with no false alarms when tested using both real and simulated data.

[1] Professor. Dept. of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH 43210, USA.
[2] Graduate Research Associate. Dept. of Civil and Environmental Engineering and Geodetic Science, The Ohio State University.

INTRODUCTION

According to one estimate about 60 percent of the total vehicle-hours of delay on urban freeways is caused by traffic incidents (Lindley, 1987). In most urban areas the situation is worsening with increasing traffic and limited expansion of the existing highway infrastructure. In fact, most major urban freeways regularly operate at levels above their design capacities.

The Intermodal Surface Transportation Efficiency Act of 1991 and the National Highway System Designation Act of 1995 realize the significance of the situation and require all urban areas with populations greater than 200,000 to implement a congestion management system (Cottrell, 1998). A number of major U.S. cities already have a freeway management system in place with remote detection of traffic characteristics and a central operations center. However, few make use of an automatic incident detection algorithm for rapid identification and localization of incidents. In most cases, detection of incidents is done by human operators monitoring video camera outputs and/or from information obtained from the news media.

Considerable research has been done on the development of traffic incident detection algorithms in the past three decades. The lack of their widespread use is primarily due to their unreliability. In the simplest case, incident detection is a classification problem with two desired output classes: incident detected and no incident detected. The misclassification of an incident into no incident detected and no incident conditions into incident detected (false alarm) reduces the reliability of the algorithm and makes it less effective for general use.

In this article, we present a new systematic approach to the traffic incident detection problem employing advanced signal processing, pattern recognition, and classification techniques. The developed model judiciously integrates fuzzy logic, wavelet theory, and

neural network computation techniques into an efficient, reliable, and robust algorithm. One key feature of the new model is noise elimination and signal enhancement to improve detection and reduce false alarms. The collection and transmission of data introduces random noise that masks the observed signal and throws off any algorithm based on them. We present an advanced de-noising technique based on wavelet theory to overcome this problem and improve the efficiency and effectiveness of the algorithm.

INCIDENT DETECTION ALGORITHMS

Several algorithms have been suggested over the years for automatic freeway incident detection based on traffic data obtained from fixed detectors. The traffic characteristics obtained from these detectors and commonly used as input for the algorithms are the traffic occupancy (the fraction of time a location is occupied by a vehicle expressed as a percentage), flow rate (the number of vehicles passing a location in unit amount of time), and speed.

The approaches used for the incident detection algorithms range from simple magnitude comparisons to model-based predictions. The California algorithm (Payne and Tignor, 1978) is a popular algorithm that compares temporal and spatial occupancy data to predetermined thresholds in its algorithm logic. The thresholds are calibrated for each on-line implementation based on the trade-off desired between the detection rate and false alarm rate. The California algorithm is an example of a multi-detector, comparative algorithm. On the other hand, the McMaster algorithm (Persaud and Hall, 1989; Persaud et al., 1990) is a single detector algorithm that is based on a catastrophe theory/model of the traffic flow. The traffic model partitions the flow rate-occupancy behavior among different traffic states. This

information is then used in the algorithm logic together with the speed data to detect the onset of congestion due to a traffic incident.

Traffic data usually exhibit sudden and large changes in magnitude that reduce the reliability of algorithms. Statistical techniques for preprocessing the raw data have been proposed in the past (Dudek et al. 1974; Cook and Cleveland 1974; Ahmed and Cook, 1982; Stephanedes and Chassiakos 1993). Dudek et al. (1974) use the standard normal deviate of the data in their threshold-based algorithm, while Cook and Cleveland (1974) propose the use of double exponential smoothing of traffic data in a similar algorithm logic. Ahmed and Cook (1982) present a short-time time-series moving average model of occupancy data to determine large deviations and predict incidents. The Minnesota algorithm (Stephanedes and Chassiakos, 1993) uses a moving average smoothing approach to remove high frequency components in observed data. The smoothed data is then employed in the algorithm logic for incident detection.

More recently research has concentrated on model-free intelligent systems approaches to the solution of the incident detection problem. These algorithms are either based on fuzzy logic theory (Chang and Wang, 1994; Lin and Chang, 1998; Weil et al. 1998), neural network techniques (Cheu and Ritchie, 1995; Dia and Rose, 1997; Amin et al., 1998), or hybrid fuzzy logic and neural network approaches (Hsiao et al., 1994; Geng and Lee, 1998). Fuzzy logic theory provides a tool for reasoning about complex systems that effectively utilizes imprecise and linguistic input (Zadeh, 1978). Chang and Wang (1994) and Lin and Chang (1998) propose a fuzzy expert system approach for the incident detection problem. The idea is to build a fuzzy knowledge base from the raw data in the form of fuzzy rules that are then processed by a fuzzy inference system to identify and classify the relevant traffic states. The

authors of these articles describe the development of the fuzzy rules but present no tested implementation of the algorithm. Weil et al. (1998) propose a fuzzy logic model of traffic flow based on a fuzzy partitioning of the traffic data into daily and weekly flow patterns. Using an unsupervised learning technique the patterns in each partition are classified into two traffic states, normal or abnormal, where the abnormal state corresponds to congested flow. This research also does not present any implementation results.

Artificial neural networks (ANN) are powerful pattern recognizers and classifiers (Adeli and Hung, 1995). They operate as black box, model-free, and adaptive tools to capture and learn significant structures in data. The use of ANNs for the identification of incident patterns in traffic data is presented by Cheu and Ritchie (1995). Three ANN architectures—multi-layer perceptron, self-organizing feature map, and adaptive resonance theory model two (ART2)—are investigated and compared with three common conventional algorithms using simulated data. Dia and Rose (1997) use field data to test a multi-layer perceptron ANN as an incident detection classifier. Amin et al. (1998) propose a control model for advanced traffic management. The traffic flow prediction module is based on a radial basis function network that can potentially be used for congestion detection. Hsiao et al. (1994) present a hybrid fuzzy logic-neural network approach for the solution of the traffic incident detection problem. They use fuzzy logic rules to partition and classify observed occupancy, flow rate, and speed data into possible incident or no incident conditions. A neural network is used to learn the membership grades needed for fuzzy reasoning. Geng and Lee (1998) use the fuzzy cerebral model arithmetic computer (CMAC) ANN architecture to learn incident patterns in traffic data. The incorporation of fuzzy logic into ANN learning makes the process more amenable to

performance analysis and system output validation. The authors, however, do not present any numerical results.

A judicious combination of AI techniques and a multi-paradigm approach has the best potential to provide an effective solution to the incident detection problem (Adeli and Hung, 1995). Work during the past 30 years on developing a model-based solution, either mathematical or symbolic, has not produced reliable solutions that can be adopted widely in practice. Currently available algorithms can miss up to 30 percent of incidents and can produce a fraction of a percent of tests in false alarms. These performance indicators may look good but when the algorithm is implemented on an urban freeway management system with hundreds or even thousands of detector stations it can produce an unacceptable number of missed detections and false alarms. As a result, the total cost of operation of these algorithms in a practical environment is often too high to justify their deployment. The primary reason for the poor performance of incident detection algorithms is the complexity of the problem that does not lend itself to accurate conventional mathematical and knowledge-based representation. On the other hand, ANN techniques are self-organizing and learn from examples. However, it is imprudent to ignore known behavior of traffic flow completely. Our new approach to be described subsequently is based on a judicious integration of various problem-solving paradigms.

WAVELET, MULTIRESOLUTION, AND TIME-FREQUENCY ANALYSIS

Basic Concept

Wavelet analysis is a transformation method in which the original signal is transformed into and represented in a different domain that is more amenable to analysis and processing.

The concept of wavelet analysis is similar to that of Fourier analysis in that both techniques decompose the original signal into a linear combination of elementary functions. However, unlike the sine and cosine harmonics used in the Fourier analysis, wavelet analysis uses a more flexible wave function called a wavelet that is localized both in time and frequency. The result is a more informative and useful decomposition of the signal. For example, because of the compact support of wavelets (i.e. the function exists only over a subset of the input space and vanishes outside it) it is possible to localize signal features in both time and frequency by analyzing the magnitudes of the wavelet coefficients. Fourier analysis, on the other hand, uses periodic functions with infinite support (i.e. the functions exist over the entire input space) making it unsuitable for transient signal analysis. In the following paragraphs we introduce the mathematics of wavelet and multiresolution analysis briefly.

A signal $x(t) \in S$ can be written as a linear combination of elementary functions $\psi_{j,k}(t)$

$$x(t) = \sum_{j,k} w_{j,k} \psi_{j,k}(t) \qquad j,k \in Z \qquad (1)$$

where $\{w_{j,k}\}$ is the set of coefficients corresponding to the expansion set $\{\psi_{j,k}\}$ and Z is the space of integers. A two-dimensional decomposition is necessary to provide time and frequency resolution which is indicated by the subscripts j and k. The signal space S may be the space of discrete-time sequences or continuous-time functions. Equation (1) is an expansion series representation of the original signal. The choice of the set $\{\psi_{j,k}\}$ determines the usefulness of the transformation.

In general, the expansion set chosen must be able to represent the original signal in a compact manner. In other words, the choice should result in a representation in which most of the coefficients $\{w_{j,k}\}$ are insignificant in magnitude. Another consideration in the choice of

the expansion set is ease of computation of both the expansion set and the corresponding expansion coefficients. In wavelet analysis, elementary functions are obtained in a structured manner from a single function in the following form:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{j}}\psi\left(\frac{t-k}{j}\right) \qquad j>0, k \in Z \tag{2}$$

where ψ is called the mother or generating wavelet. The integers j and k represent the scaling and translation values, respectively. In most practical uses, the scaling in Eq. (2) is done in powers of two. For this dyadic formulation Eq. (2) can be rewritten as

$$\psi_{j,k}(t) = 2^{j/2}\psi\left(2^j t - k\right) \qquad j>0, k \in Z \tag{3}$$

When an orthonormal basis is used as the expansion set the coefficients of the expansion can be computed by an inner product of the signal with the corresponding wavelet:

$$w_{j,k} = \left\langle x, \psi_{j,k} \right\rangle = \int x(t)\psi_{j,k}(t)dt \tag{4}$$

Equation (1) with the coefficients given by Eq. (4) is called the discrete-time or continuous-time wavelet transform. It is called a discrete-time wavelet transform or discrete wavelet transform (DWT) when x is a discrete-time sequence and a continuous-time transform or continuous wavelet transform (CWT) when x is a continuous-time function. In the following discussion it is assumed that the signal is a discrete-time function and Eq. (1) represents the DWT of the function.

Multiresolution Analysis

Multiresolution analysis provides a powerful framework for analyzing functions at various levels of detail or resolution (Mallat, 1989). Multiresolution analysis entails a sequence of nested closed approximation subspaces V_m $(m \in Z)$, satisfying the following properties:

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \tag{5}$$

$$\overline{\bigcup_{m \in Z} V_m} = L^2(R) \tag{6}$$

$$\bigcap_{m \in Z} V_m = \{0\} \tag{7}$$

$$x(t) \in V_m \Leftrightarrow x(2t) \in V_{m+1} \tag{8}$$

$$x(t) \in V_0 \Rightarrow x(t-j) \in V_0 \qquad j \in Z \tag{9}$$

and there exist a scaling function $\varphi \in V_0$ such that $\varphi_{0,k}$ $(k \in Z)$ forms a basis of V_0. The scaling function $\varphi_{j,k}$ is defined as in Eq. (3). In Eqs. (5)-(9), $V_0 \subset V_1$ means that V_0 is a subspace of V_1, $\bigcup$ represents the union of spaces, $\bigcap$ represents the intersection of spaces, the over bar denotes the closure of the space, $L^2(R)$ is the space of all square integrable functions of real variables, and $\Rightarrow$ and $\Leftrightarrow$ stands for one way and two way implications, respectively.

If Eqs. (5)-(9) hold then there exists a set of functions $\psi_{j,k}$ (Eq. 3) such that $\psi_{j,k}$ $(k \in Z)$ spans W_j which is the orthogonal complement of the spaces V_j and V_{j+1}. More specifically, if $\{\varphi_{0,k}\}$ spans V_0 then $\{\psi_{0,k}\}$ spans W_0 such that

$$V_1 = V_0 \oplus W_0 \tag{10}$$

and, in general

$$L^2(R) = \cdots \oplus W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus W_2 \oplus \cdots \tag{11}$$

where $\oplus$ represents a direct sum. This means by starting from a representation of a function belonging to a coarse subspace higher detail or resolution can be obtained by adding spaces spanned by $\psi_{j,k}$ at a higher resolution (i.e. given by the next higher value of j).

The function $x(t)$ can then be represented as

$$x(t) = \sum_k c_{j_0,k}\varphi_{j_0,k}(t) + \sum_k \sum_{j=j_0} d_{j,k}\psi_{j,k}(t) \tag{12}$$

where the first term is a coarse resolution at scale j_0 and the second term adds details of increasing resolutions. Equation (12) can also be viewed as the time-frequency decomposition of *x(t)* where the second term provides the frequency and time breakdown of the signal. The nesting of spaces achieved by multiresolution and time-frequency analysis is shown conceptually in Figure 1. Note that spaces spanned by different scales of wavelets are orthogonal to each other because they do not overlap (non-overlapping functions are always orthogonal).

Computation of the DWT

In practical wavelet analysis of discrete signals we usually do not have to deal with the functions themselves but instead work with discrete coefficients. If $\{\varphi_{j,k}\}$ and $\{\psi_{j,k}\}$ form an orthonormal basis of $L^2(R)$, which is true for most wavelet systems used in practice, the expansion coefficients $c_{j,k}$ and $d_{j,k}$ can be found by taking the inner products of the basis functions and the original signal. Using the properties of the wavelet system, Eq. (4) can be written in terms of the coefficients as follows (Burrus et al., 1998):

$$c_{j,k} = c_j[k] = \sum_m h_0[m-2k]c_{j+1}[m] \tag{13}$$

$$d_{j,k} = d_j[k] = \sum_m h_1[m-2k]c_{j+1}[m] \tag{14}$$

The sequences h_0 and h_1 are called filter coefficients whose values are known for each type of wavelet system that may be used for analysis. The initial scaling coefficients c_j are taken equal to the original discrete signal. Equations (13)-(14) provide a recursive way to compute the DWT of a signal. Note that these computations have a finite time complexity as the

coefficients are of finite length. The inverse DWT is used to reconstruct the signal from the wavelet coefficients using Eq. (12). In this work we use Daubechies wavelet system of length 8 (Daubechies, 1992). For a more detailed coverage of DWT and its computation see Samant and Adeli (2000).

SELECTION OF TYPE AND NUMBER OF TRAFFIC DATA

It is important to carefully choose the number, type, and format of input data to be used for the incident detection algorithm. Most currently used sensors provide the speed, the occupancy, and the flow rate values at a given location every 20 or 30 seconds. Therefore, the choice for the type of traffic data has to be restricted to these three types. From these three data types only those that exhibit consistently identifiable patterns for incident and non-incident traffic flow conditions should be selected.

In this work, a pattern consists of a time-history of data rather than a single-time data value. This pattern preserves the temporal nature of traffic flow and makes distinguishing between patterns produced by incident and non-incident conditions easier. The distinguishing feature adopted in this work is the shape of the time-history and not any particular magnitude. To achieve this, each pattern is normalized to eliminate the effect of data magnitudes on the classification process. This approach also eliminates algorithm calibration and transferability issues caused by location specific conditions and temporal traffic flow variations. A single-station non-comparative approach is adopted in this research. This decision is based on the analysis of patterns on both the upstream and downstream side of an incident. The upstream and downstream patterns produced by an incident do not develop at the same time. Therefore, mixing them reduces the reliability of the algorithm. Furthermore, using patterns from

adjacent stations makes the algorithm dependent on several factors such as incident characteristics, distance between stations, and existence of on- and off-ramps in between the stations. The result is calibration problems and poor performance of the algorithm.

The speed and occupancy upstream of a capacity reducing obstruction are found to exhibit the most significant and consistent change relatively independent of the flow rate (Figure 2a, b). Consequently, the upstream speed and occupancy time-series data are used as input for the new model. Each pattern of traffic consists of N data points for the occupancy and the speed values obtained at the lane sensor immediately upstream of the incident location. From the algorithmic performance point of view the smallest number that can produce accurate results must be chosen. Computationally, however, DWT requires N to be a power of 2. Our numerical experiments indicate $N = 16$ provides accurate results and is therefore used in the model. The 16 data points constitute 5 minutes and 20 seconds of data, if data is obtained every 20 seconds. This represents a sufficient amount of data to characterize before and after incident traffic flow conditions and establish the defining shape of the traffic pattern. Eight data points did not produce good performance while the performance with 32 data points was identical to that for 16 data points. The normalized occupancy and speed data streams obtained from a given sensor location are denoted by the sequences $x_O[n]$ and $x_S[n]$, respectively, where $n = 1, 16$.

WAVELET-BASED DE-NOISING

When a signal is transformed into the wavelet domain it often becomes less complicated to effectively reduce noise and outliers in the signal. This ease is usually due to a degree of separation of noise and signal in the wavelet domain. For example, if the noise is made up of

localized high frequency components in a predominantly low frequency signal then the signal can be de-noised by the following procedure. Take the DWT of the signal, selectively discard the higher scale coefficients, and then reconstruct the signal by taking the inverse DWT. This technique is not optimal and automatic for use in a real-time intelligent system environment. In particular, no definite criteria are available to determine which wavelet coefficients to discard in order to produce the best results.

In recent years, formal wavelet-based de-noising techniques have been presented in the literature (Polchlopek and Noonan, 1997; Donoho, 1993, 1995). These techniques perform a nonlinear filtering on the transformed signal, modifying the wavelet coefficients in such a way that the inverse transformation yields a de-noised signal.

Donoho (1995) presented a technique in which the wavelet coefficients are passed through a nonlinear threshold filter. The resulting coefficients then represent an optimally de-noised DWT of the original signal. To de-noise each of the data sequences $x_O[n]$ and $x_S[n]$ the following procedure is employed:

- Calculate the DWT of $x[n]$ to obtain the noisy wavelet coefficients $\{d_{j,k}\}$. The 16 data points can be resolved into 4 different frequency bands or scales. The coarsest scale j_0 resolved in the DWT is 2 producing $2^2 = 4$ scaling coefficients. At this scale also the general shape of the original sequence is preserved. The number of wavelet coefficients obtained is $(2^4 - 2^2) = 12$ corresponding to the two highest scales. Applying the soft-thresholding on these coefficients will effectively remove the higher frequency components without distorting the signal.

- Filter the wavelet coefficients using the soft-thresholding nonlinearity $\eta(d) = \text{sgn}(d)(|d| - t)^+$ where $(.)^+$ is equal to (.) when (.) is positive and zero otherwise and the function sgn(.) returns the sign of its argument. The threshold t is given by $t = \sqrt{2\log(N)}$ where N (equal to 16 in our test example) is the total number of data points.
- Perform the inverse DWT using the scaling and the filtered wavelet coefficients.

The de-noised signals corresponding to $x_O[n]$ and $x_S[n]$ are denoted by $\bar{x}_O[n]$ and $\bar{x}_S[n]$. These signals will be cleaner versions of the original corrupted signal.

FUZZY DATA CLUSTERING

Data clustering techniques extract significant features from data based on given criteria. The goal is to reduce the dimensionality of the data without losing important information needed for a particular problem. Dimensionality reduction is needed to reduce data processing complexity and increase robustness and efficiency. The data clustering problem can be stated as follows: Given a set of vectors $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$ find the set $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \ldots, \mathbf{z}_c\}$ where $2 \le c < n$ and $\mathbf{x}, \mathbf{z} \in R^p$ such that $\mathbf{Z}$ properly characterizes $\mathbf{X}$. The vectors $\mathbf{z}_i$ represent classes or clusters in $\mathbf{X}$. In general, data clustering techniques are either based on statistical or fuzzy logic theory. It has been shown that most of these techniques have similar properties and produce comparable results (Dave and Krishnapuram, 1997). However, fuzzy logic approaches have the advantage of effective handling of imprecision.

The fuzzy c-means (FCM) clustering algorithm (Bezdek, 1981; Cannon et al., 1986) performs a fuzzy partitioning of the data set into classes. This is in contrast to crisp

assignment of data vectors to distinct classes employed in classical statistical clustering techniques. The prefix c in the fuzzy c-partitions refers to the number of classes in each partition. The clustering problem can be posed as a constrained optimization problem as follows:

Minimize

$$J_\beta(\mathbf{z}) = \sum_{i=1}^{n}\sum_{j=1}^{c} A_{ij}^{\beta} \left\| \mathbf{x}_i - \mathbf{z}_j \right\|^2 \tag{15}$$

subject to

$$\sum_{j=1}^{c} A_{ij} = 1 \quad 1 \le i \le n \tag{16}$$

$$A_{ij} \ge 0 \qquad 1 \le i \le n, 1 \le j \le c \tag{17}$$

where J_β is the objective function for a given value of β, A_{ij} is the membership grade of vector i in class j, and $\|.\|$ denotes the Euclidean norm. The parameter β represents the degree of fuzziness in the data. This value is often in the range $2 \ge \beta > 1$. Larger values are selected for fuzzier data situations. A value of $\beta = 1.5$ is chosen in the test example in this work. Note that c, the number of classes desired, is an input parameter. The classes are identified by the cluster centers $\mathbf{z}_i$ and the membership of a vector in a given class is determined by its Euclidean distance from the class center.

In a general FCM formulation the membership grades A_{ij} are also optimization variables. However, this formulation leads to a non-convex optimization problem that does not always produce a global optimal solution (Al-Sultan and Fediki, 1997). When using an iterative

procedure for solving the optimization problem we use the following membership grade function based on the Euclidean norm (Bezdek, 1981).

$$A_{ij}^{t+1} = \left[\sum_{k=1}^{c} \left(\frac{\left\| \mathbf{x}_i - \mathbf{z}_j^t \right\|^2}{\left\| \mathbf{x}_i - \mathbf{z}_k^t \right\|^2} \right)^{\frac{1}{\beta-1}} \right]^{-1} \qquad 1 \le i \le n, \;\; 1 \le j \le c \tag{18}$$

where the superscript t denotes the iteration number.

To cluster the de-noised data sequences $\bar{x}_O[n]$ and $\bar{x}_S[n]$ we define the feature or traffic pattern matrix $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N\}$ where the vector $\mathbf{x}_i$ is given by

$$\mathbf{x}_i = \{\bar{x}_O[i], \bar{x}_S[i]\} \; 1 \le i \le N . \tag{19}$$

and use the FCM algorithm in the following form.

1. Select an initial fuzzy c-partition by setting up the membership grades A_{ij} such that Eq. (16) is satisfied. Select a value for $\beta > 1$. Set the iteration counter $t = 0$.
2. Calculate the class centers for the traffic pattern $\mathbf{X}$.

$$\mathbf{z}_j^t = \frac{\sum_{i=1}^{n} A_{ij}^m \mathbf{x}_i}{\sum_{i=1}^{n} A_{ij}^m} \qquad 1 \le j \le c \tag{20}$$

3. Calculate the updated membership grade using Eq. (18).
4. If the maximum change in the membership grade is less than ε, or

$$\max \left| A_{ij}^{t+1} - A_{ij}^t \right| < \varepsilon \qquad 1 \le i \le n, 1 \le j \le c \tag{21}$$

stop. Otherwise, update $t = t + 1$ and go to step 2.

This algorithm is efficient and usually converges in a few iterations.

The FCM algorithm is used to reduce the dimensionality of the feature matrix to obtain c cluster centers $\mathbf{z}_i$ where $1 < c < N$. In the test example, the 16 pairs of occupancy and speed data are reduced to 4 (i.e. $c = 4$) representative samples. This reduced data set contains the most significant features of the original data and is then used for classification of traffic signals into incident and incident-free signals. It should be noted that these computations are efficient as the FCM algorithm converges in less than 10 iterations and the dimensionality of the data is small.

RADIAL BASIS FUNCTION NEURAL NETWORK CLASSIFIER

The radial basis function neural network (RBFNN) learns an input-output mapping by covering the input space with basis functions that transforms a vector from the input space to the output space (Moody and Darken, 1989; Poggio and Girosi, 1990). Conceptually, the RBFNN is an abstraction of the observation that biological neurons exhibit a receptive field of activation such that the output is large when the input is closer to the center of the field and small when the input moves away from the center. Structurally, the RBFNN has a simple topology with a hidden layer of nodes having nonlinear basis transfer functions and an output layer of nodes with linear transfer functions.

Figure 3 shows the topology of the RBFNN for the classification of traffic data into two states: incident and no incident. Therefore, only a single node in the output layer is required. The input vector is denoted by $\mathbf{x}$ and the output is denoted by y. The number of input nodes is equal to N_i which is equal to the product of the number of clusters, c (equal to 4 in our test example), and the dimension of each cluster (equal to 2, when occupancy and speed is used as in our example). The number of nodes in the hidden layer is equal to the number of cluster

centers, $1 < N_c < N_p$, for the entire training instances where N_p is the total number of training instances. The cluster centers $\boldsymbol{\mu}_i$ $(1 \le i \le N_c)$ is obtained using the FCM algorithm.

The connection from the input node i to the hidden node j is assigned the weight μ_{ji} corresponding to the ith component of the vector $\boldsymbol{\mu}_j$. Each hidden node produces an output that is a function of the Euclidean distance of the input vector $\mathbf{x}$ from the cluster center $\boldsymbol{\mu}_j$. In this work, we use the Gaussian (bell-shaped) function as the transfer function for the hidden nodes. The output of the hidden node j is then given by

$$\phi_j = \exp\left(-\frac{\|\mathbf{x}-\boldsymbol{\mu}_j\|^2}{2\sigma_j^2}\right) \tag{22}$$

where the factor σ_j controls the spread or range of influence of the Gaussian function centered at $\boldsymbol{\mu}_j$. The output y of the network is given by

$$y = \sum_{j=1}^{N_c} \phi_j \lambda_j \tag{23}$$

where λ_j is the weight of the link from the hidden node j to the output node. The output value of 1 corresponds to an incident classification while a value of -1 corresponds to a no incident classification.

The variables λ_j's and μ_{ji}'s are found by training the neural network off-line. The FCM algorithm is used to obtain N_c cluster centers $\boldsymbol{\mu}_i$ from the N_p training instances $\mathbf{x}$. The RBFNN is trained to find the weights λ_j by minimizing the error between the network computed output y and the desired output y_d. In other words, to train the network for λ_j's we solve the following unconstrained optimization problem:

$$\text{Minimize } E(\boldsymbol{\lambda}) = \sum_{i=1}^{N_c} \left| y^i - y_d^i \right| \tag{24}$$

The gradient descent optimization algorithm is used to solve this optimization problem.

The spread parameters σ_j's can also be treated as variables. However, we found that there was no improvement in the performance of the classification when the spread parameter is allowed to adapt. At the same time, including the parameter in the learning process slows down the training. In this work, the following expression is used to pre-assign the value of σ_j:

$$\sigma_j = \frac{1}{3N_c} \sum_{i=1}^{N_c} \left\| \boldsymbol{\mu}_j - \boldsymbol{\mu}_i \right\| \quad 1 \le j \le N_c \tag{25}$$

This equation approximates the spread parameter σ_j as one third of the mean distance between the cluster center at j and all other cluster centers. In this way an adequate amount of overlap of the basis functions is achieved for classification purposes.

EXAMPLE

The new incident detection algorithm is tested using both simulated and real traffic data. The simulated data is generated from the simulation software TSIS (Traffic Software Integrated System) (http://www.fhwa-tsis.com). TSIS uses a microscopic stochastic model to simulate traffic flow on freeways. A variety of parameters can be specified to simulate different traffic flow scenarios. By changing the random number seeds for each simulation run a representative sample is obtained for training and testing. The real traffic data is obtained from the Freeway Service Patrol Project's I-880 database in California

(http://www.path.berkeley.edu/FSP/). The model is trained using simulated data only. The trained model is then tested using both simulated and real traffic data.

The simulated training and testing data is generated from simulating traffic on a straight stretch of a two-lane (in one direction) freeway. Traffic enters the freeway section from one end and exits from the other. Pairs of loop detectors are spaced 450-750 m (1500-2500 feet) apart. A total of one hundred and fifty 800-second simulations were performed with data obtained in 20-second intervals. Ninety of these simulations involve a traffic incident while the remaining sixty do not have any incident. Each incident is modeled by the blockage of one lane and the reduction in capacity of the adjacent lane. The blockages are evenly distributed between the two lanes and are located at varying distances from an upstream detector station. The entry flow rate is varied in the range 2000-2500 vehicles per hour. Low demand conditions are adopted for evaluation because these are the conditions under which currently available incident detection algorithms perform poorly.

Thirty incident and thirty non-incident patterns were used for training. It was found that the basic shapes of the occupancy and speed plots are similar in different incident simulation runs; the primary difference is that they are time shifted depending on the location of the incident downstream of a detector station and the flow rate at the time of the incident. Therefore, to ensure that the incident patterns are consistent they are extracted from the 800-second simulations such that the effects of the blockage is pronounced during the last few values of the sample. Figure 4 shows the normalized occupancy plots for two simulation runs. Figure 4a is for an incident 244 m downstream of the detector station while Figure 4b is for an incident 122 m downstream of the detector station. Figure 5 shows the corresponding occupancy incident patterns extracted from these simulations and used for training. Notice the

similarity of the form of the two patterns. This pattern extraction is essential for robust classification. For the test example, the RBFNN learned the patterns with a cumulative mean square error of less than 0.003 in a few seconds on a Pentium II 400 MHz machine.

Testing of Algorithm Using Simulated Data

To test the algorithm the output from the RBFNN is passed through a threshold, t, of 0.3. An output greater than or equal to 0.3 is classified as an incident. Otherwise, it is classified as a non-incident. The model is tested using the simulated data by presenting each of the ninety 800-second simulation as a continuous stream of data. An output is produced every 20-second after the first 320-second (16 data points). An incident is detected when the output becomes greater than the threshold for the first time. All the 60 incidents were detected correctly during the testing of the model. Therefore, the detection rate is 100 percent. Also, none of the non-incident simulations or the incident simulations before the occurrence of the incident (a total of 360 patterns) were misclassified as an incident. Therefore, the false alarm rate is zero.

The time to detection tends to be somewhat large for flow rates less than the freeway capacity. Figure 6 shows the variation of the mean detection time of the algorithm with pre incident flow rate and distance from the upstream detector station.

Testing of Algorithm Using Real Data

The I-880 database contains loop detector and incident data for a 14.8 km (9.2-mile) long segment of the freeway from Oakland to San Jose, California. The number of lanes in each direction varies from three to five. The incident data is recorded by human observers traversing this segment of the freeway in patrol vehicles. Several incident characteristics are recorded including the type of the incident, the location of the incident, and the time of occurrence of the incident. For the testing of the new incident detection algorithm, the

southbound data is processed to extract 21 incidents that block one or more lanes. The loop detector data are averaged over a 30-second time interval. Our incident detection model detected 20 of the 21 incidents, resulting in a detection rate of 95.2 percent. The traffic pattern corresponding to the missed incident did not exhibit the characteristics of an incident condition. This appears to be an error in the incident data. The incident data, in general, is not accurate as the location of incidents are reported approximately (like 1 mile from exit) and the time of the incident is actually the time at which a patrol vehicle observed the incident and not the time at which the incident occurred. As a result, it is not possible to determine the time to detection which in our tests varied from negative to positive values.

Four hours of incident free traffic data are used for testing the false alarm performance. In all, 30 patterns were presented to the model. Our new incident detection model correctly identified all 30 patterns as non -incident patterns. Thus, the false alarm rate is zero.

Note that the model trained using simulated is tested on both simulated and real data without modification. Also, the simulated data is available at 20-second interval while the real data is available at 30-second intervals. The model does not require any calibration and can be used at all locations once it has been trained.

CONCLUSION

A new multi-paradigm intelligent system methodology is presented for the solution of the traffic incident detection problem. The methodology effectively integrates fuzzy, wavelet, and neural computing techniques to improve reliability and robustness of the algorithm. A wavelet-based de-noising technique is employed to eliminate undesirable fluctuations in observed data from traffic sensors. Fuzzy clustering is used to extract significant information

from the observed data and to reduce its dimensionality. A radial basis function neural network is developed to classify the de-noised and clustered observed data. The new methodology has been implemented in the combination of C++ and MATLAB programming environments.

The algorithm was tested using both simulation and real data. One hundred and fifty simulation runs were performed by changing the blocked lane, the distance of the blockage from the upstream sensor, and the flow rate. Under these conditions the algorithm produces the detection rate of 100 percent and the false alarm rate of zero. Real traffic data was obtained from the I-880 database. The algorithm correctly identified 20 out of 21 lane-blocking incidents and did not signal a false alarm in four hours of incident free data.

The methodology presented provides a solid foundation for further research and development. We are currently investigating approaches to improve the mean detection time without sacrificing the excellent reliability of the algorithm.

ACKNOWLEDGMENT

This manuscript is based on a research project sponsored by the Ohio Department of Transportation and Federal Highway Administration.

REFERENCES

Adeli, H. and Hung, S. L. (1995), *Machine Learning--Neural Networks, Genetic Algorithms, and Fuzzy Systems*, John Wiley & Sons, Inc., New York, NY.

Adeli, H. and Park, H. S. (1998), *Neurocomputing for Design Automation*, CRC Press, Boca Raton, FL.

Ahmed, S. A. and Cook, A. R. (1982), "Application of Time-Series Analysis Techniques to Freeway Incident Detection," *Transportation Research Record*, No. 841, pp. 19-28.

Al-Sultan, K. S. and Fediki, C. A. (1997), "A Tabu Search-Based Algorithm for the Fuzzy Clustering Problem," *Pattern Recognition*, Vol. 30, No. 12, pp. 2023-2030.

Amin, S. M., Rodin, E. Y., Liu, A. -P., Rink, K. and Garcia-Ortiz, A. (1998), "Traffic Prediction and Management Via RBF Neural Nets and Semantic Control," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 13, pp. 315-325.

Bezdek, J. C. (1981), *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY.

Burrus, C. S., Gopinath, R. A. and Guo, H. (1998), *Introduction to Wavelets and Wavelet Transforms--A Primer*, Prentice-Hall, Inc., Upper Saddle River, NJ.

Cannon, R. L., Dave, J. V. and Bezdek, J. C. (1986), "Efficient Implementation of the Fuzzy c-Means Clustering Algorithms," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 2, pp. 248-255.

Chang, E. C. and Wang, S. -H. (1994), "Improved Freeway Incident Detection Using Fuzzy Set Theory," *Transportation Research Record*, No. 1453, pp. 75-82.

Cheu, R. L. and Ritchie, S. G. (1995), "Automated Detection of Lane-Blocking Freeway Incidents Using Artificial Neural Networks," *Transportation Research-C*, Vol.3, No. 6, pp. 371-388.

Cook, A. R. and Cleveland, D. E. (1974), "Detection of Freeway Capacity-Reducing Incidents by Traffic Stream Measurements," *Transportation Research Record*, No. 495, pp. 1-11.

Cottrell, W. D. (1998), "Estimating the Probability of Freeway Congestion," *Transportation Research Record*, No. 1634, pp. 19-27.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA.

Dave, R. N. and Krishnapuram, R. (1997), "Robust Clustering Methods: A Unified View," *IEEE Transactions on Fuzzy Systems*, Vol. 5, No. 2, pp. 270-293.

Dia, H. and Rose, G. (1997), "Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data," *Transportation Research--C*, Vol. 5, No. 5, pp. 313-331.

Donoho, D. L. (1993), "Nonlinear Wavelet Methods for Recovery of Signals, Images, and Densities from Noisy and Incomplete Data," Daubechies, I., Ed., *Different Perspectives on Wavelets*, AMS, Providence, RI.

Donoho, D. L. (1995), "De-Noising by Soft Thresholding," *IEEE Transactions on Information Theory*, Vol. 41, pp. 613-627.

Dudek, C. L., Messer, C. J. and Nuckles, N. B. (1974), "Incident Detection on Urban Freeways," *Transportation Research Record*, No. 495, pp. 12-24.

Geng, J. and Lee, T. N. (1998), "Freeway Traffic Incident Detection Using Fuzzy CMAC Neural Networks," *IEEE Internation Conference on Fuzzy Systems*, Anchorage, AK, 5/4-5/9, Vol. 2, pp. 1164-1169.

Hsiao, C. -H., Lin, C. -T. and Cassidy, M. (1994), "Applications of Fuzzy Logic and Neural Networks to Automatically Detect Freeway Traffic Incidents," *Journal of Transportation Engineering*, ASCE, Vol. 120, No. 5, pp. 753-772

Lin, C. -K. and Chang, G. -L. (1998), "Development of a Fuzzy-Expert System for Incident Detection and Classification," *Mathematical and Computer Modelling*, Vol. 27, No. 9-11. pp. 9-25.

Lindley, J. A. (1987), "Urban Freeway Congestion: Quantification of the Problem and Effectiveness of Potential Solutions," *ITE Journal*, Vol. 57, No. 1, pp. 27-32.

Mallat, S. G. (1989), "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 674-693.

Moody, J. and Darken, C. J. (1989), "Fast Learning in Networks of Locally-Tuned Processing Units," *Neural Computation*, Vol. 1, pp. 281-294.

Payne, H. J. and Tignor, S. C. (1978), "Freeway Incident-Detection Algorithms Based on Decision Trees With States," *Transportation Research Record*, No. 682, pp. 30-37.

Persaud, B. N. and Hall, F. L. (1989), "Catastrophe Theory and Patterns in 30-Second Freeway Traffic Data--Implications for Incident Detection," *Transportation Research--Part A*, Vol. 23A, No. 2, pp. 103-113.

Persaud, B. N., Hall, F. L. and Hall, L. M. (1990), "Congestion Identification Aspects of the McMaster Incident Detection Algorithm," *Transportation Research Record*, No. 1287, pp. 167-175.

Poggio, T. and Girosi, F. (1990), "Networks for Approximation and Learning," *Proceedings of the IEEE*, Vol. 78, pp. 1481-1497.

Polchlopek, H. M. and Noonan, J. P. (1997), "Wavelets, Detection, Estimation, and Sparsity," *Digital Signal Processing*, Vol. 7, pp. 28-36.

Samant, A. and Adeli, H. (2000), "Feature Extraction for Traffic Incident Detection Using Wavelet Transform and Linear Discriminant Analysis," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 15, No. 4, pp. 241-250.

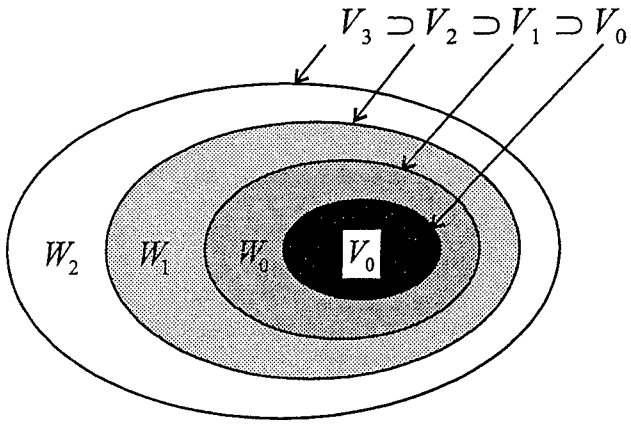
Stephanedes, Y. J. and Chassiakos, A. P. (1993), "Application of Filtering Techniques for Incident Detection," *Journal of Transportation Engineering*, ASCE, Vol. 119, No. 1, pp. 13-26.

Weil, R., Wootton, J. and Garcia-Ortiz, A. (1998), "Traffic Incident Detection: Sensors and Algorithms," *Mathematical and Computer Modelling*, Vol. 27, No. 9-11, pp. 257-291.

Zadeh, L. A. (1978), "Fuzzy Set as a Basis for a Theory of Possibility," *Fuzzy Sets and Systems*, Vol. 1, No. 1, pp. 3-28.

LIST OF CAPTIONS FOR FIGURES

1. Multiresolution function space decomposition using wavelet analysis
2. Typical time-histories upstream of an incident (a) occupancy plot, and (b) speed plot
3. Radial basis function neural network for discriminating incident and non-incident patterns
4. Normalized occupancy plots obtained from simulating traffic on a two-lane freeway (incident occurs at time 400 second)
5. The occupancy incident patterns extracted from the simulations presented in Figure 7
6. Mean detection time of an incident as a function of flow rate and distance from upstream sensor



$$W_2 \perp W_1 \perp W_0 \perp V_0$$

$X \supset Y$ Y is a subspace of X
$X \perp Y$ Y is orthogonal to X

Figure 1

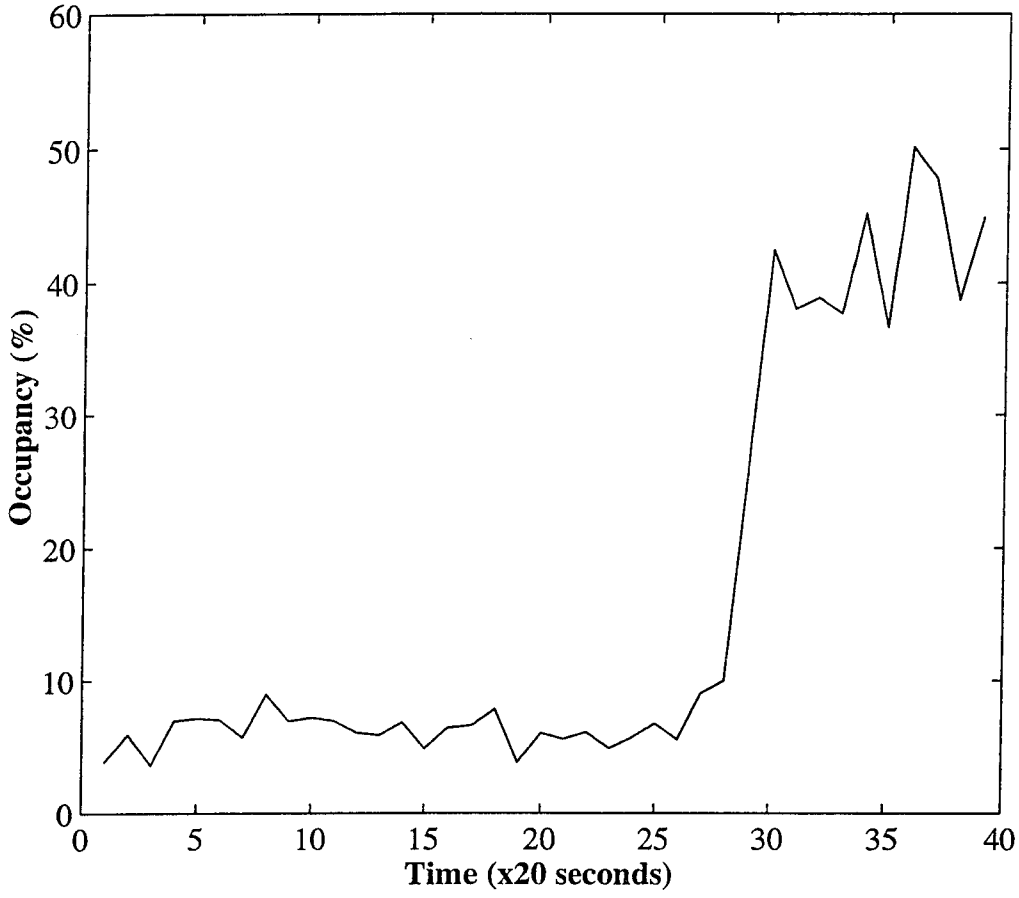

60
50
40
30
20
10
0
Occupancy (%)
0
5
10
15
20
25
30
35
40
Time (x20 seconds)

Figure 2a

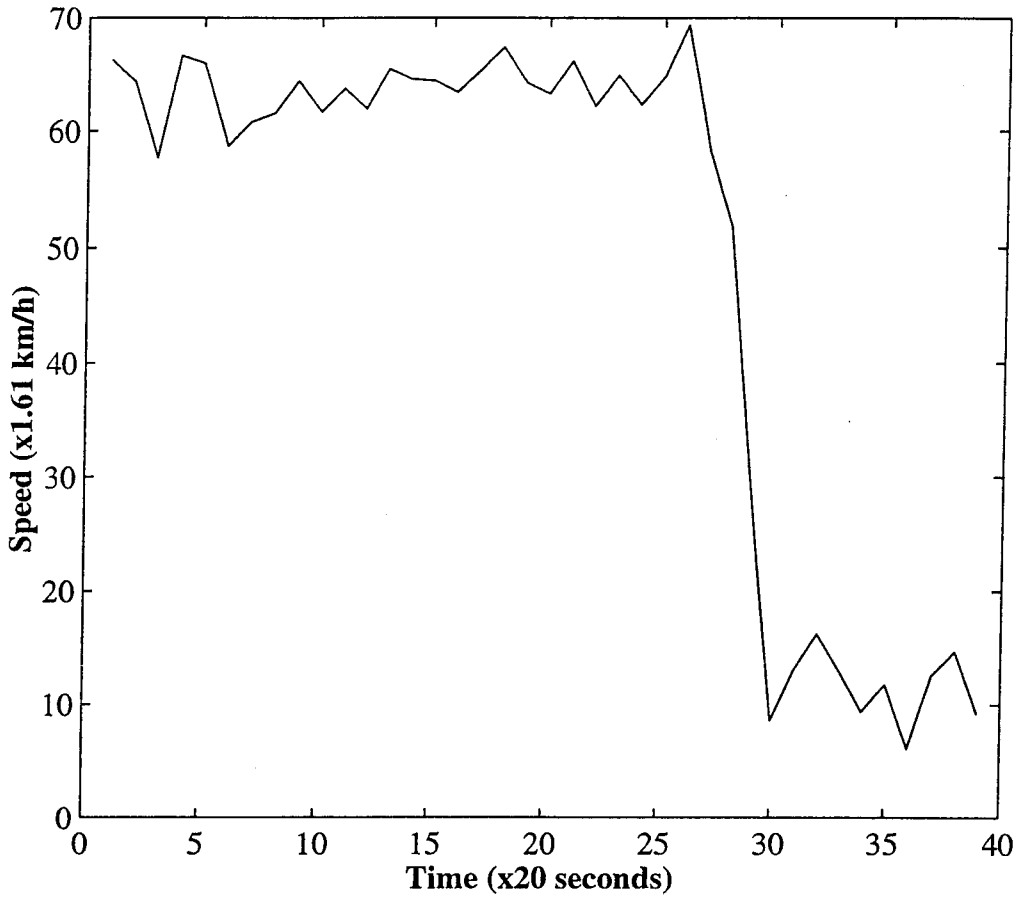
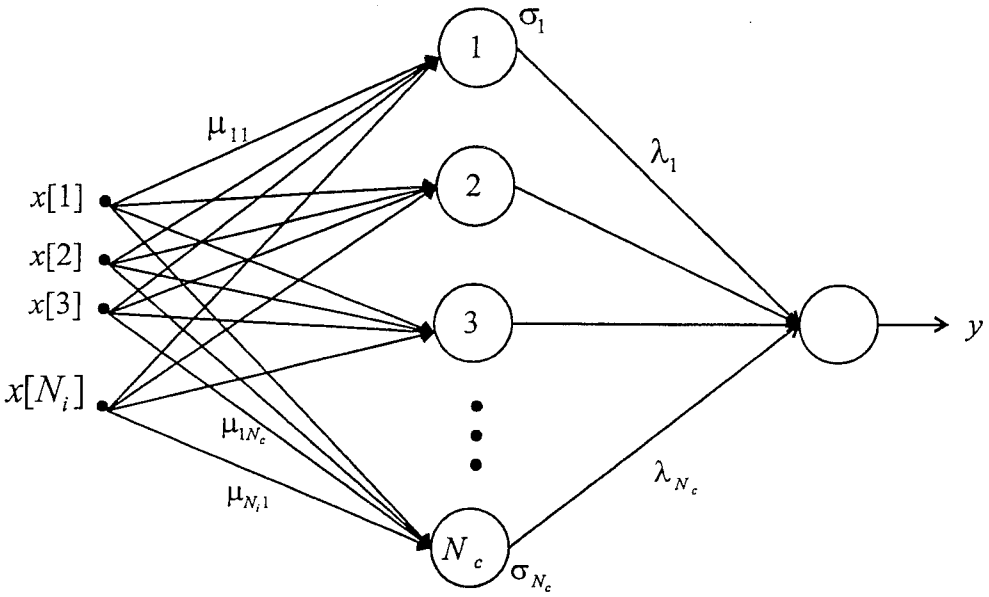

70
60
50
40
30
20
10
0
Speed (x1.61 km/h)
0
5
10
15
20
25
30
35
40
Time (x20 seconds)

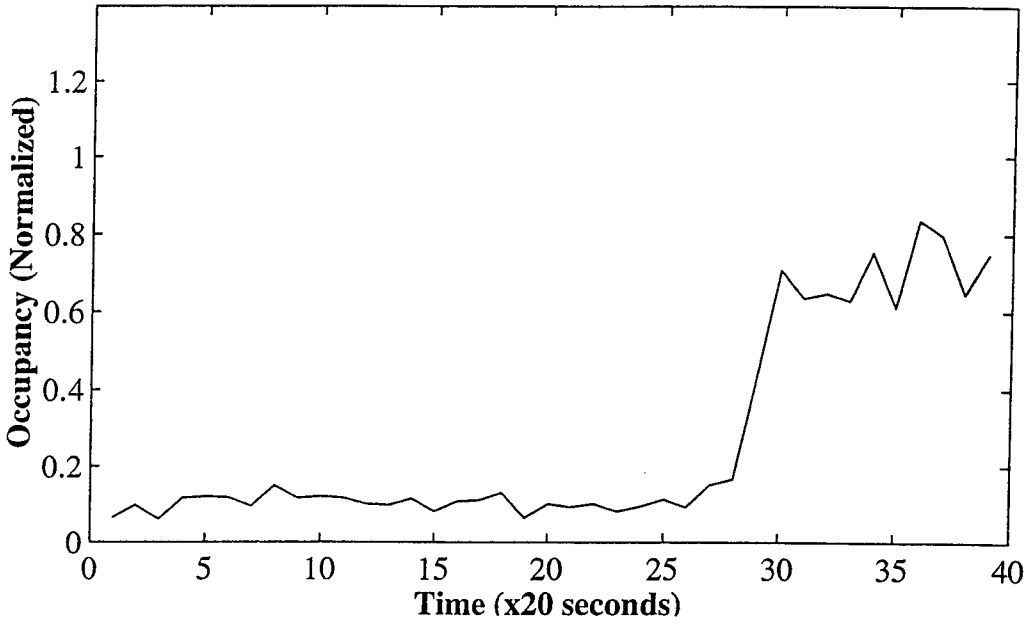
Figure 2b



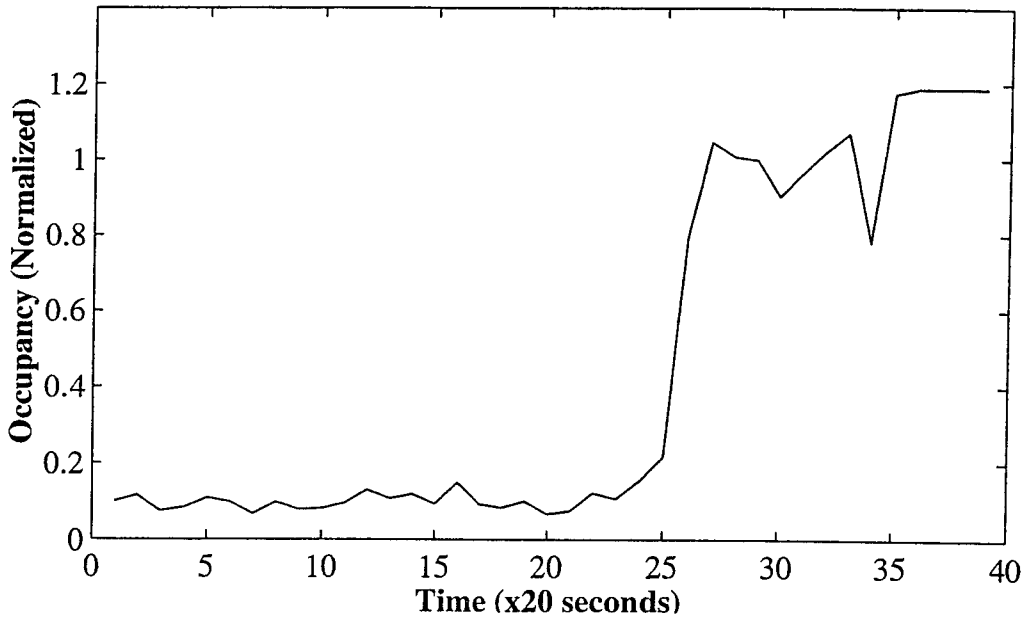
N_i = Number of inputs

N_c = Number of cluster centers

Figure 3

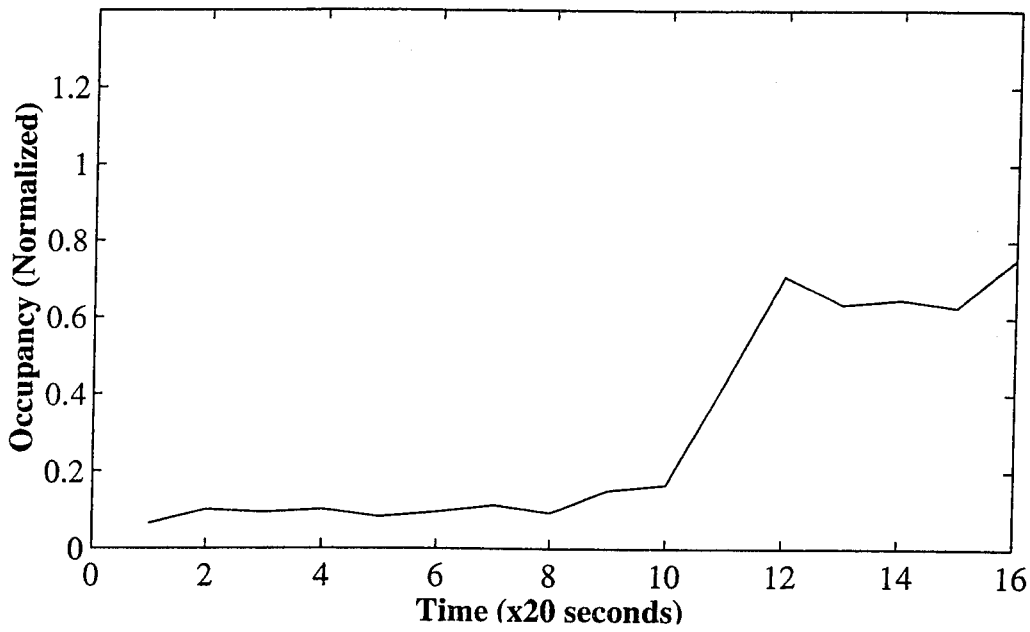


(a) Incident located 122 m downstream of sensor

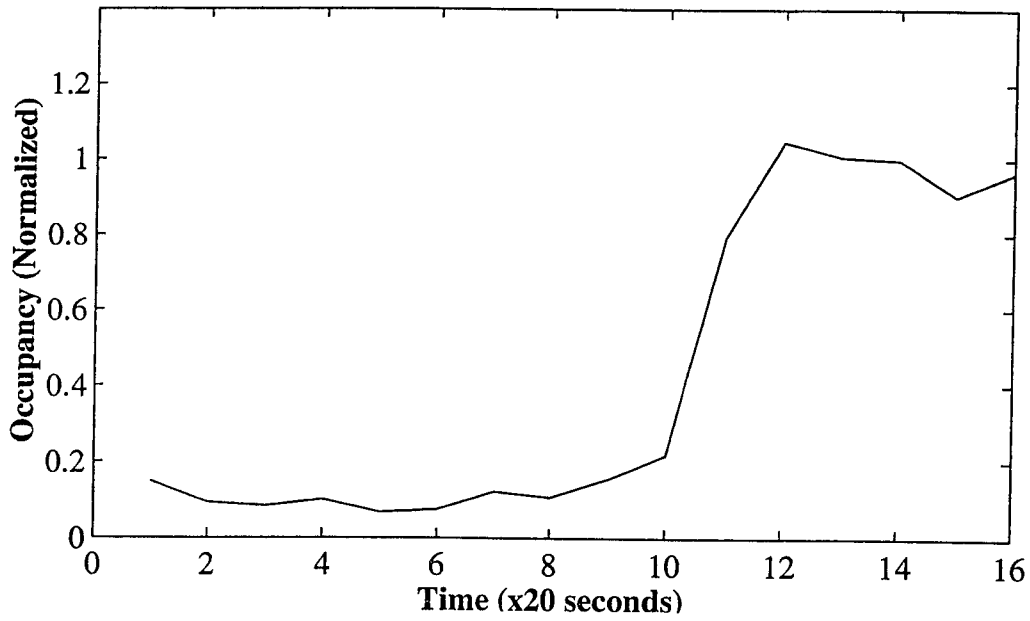


(b) Incident located 244 m downstream of sensor

Figure 4



(a) Incident located 122 m downstream of sensor



(b) Incident located 244 m downstream of sensor

Figure 5

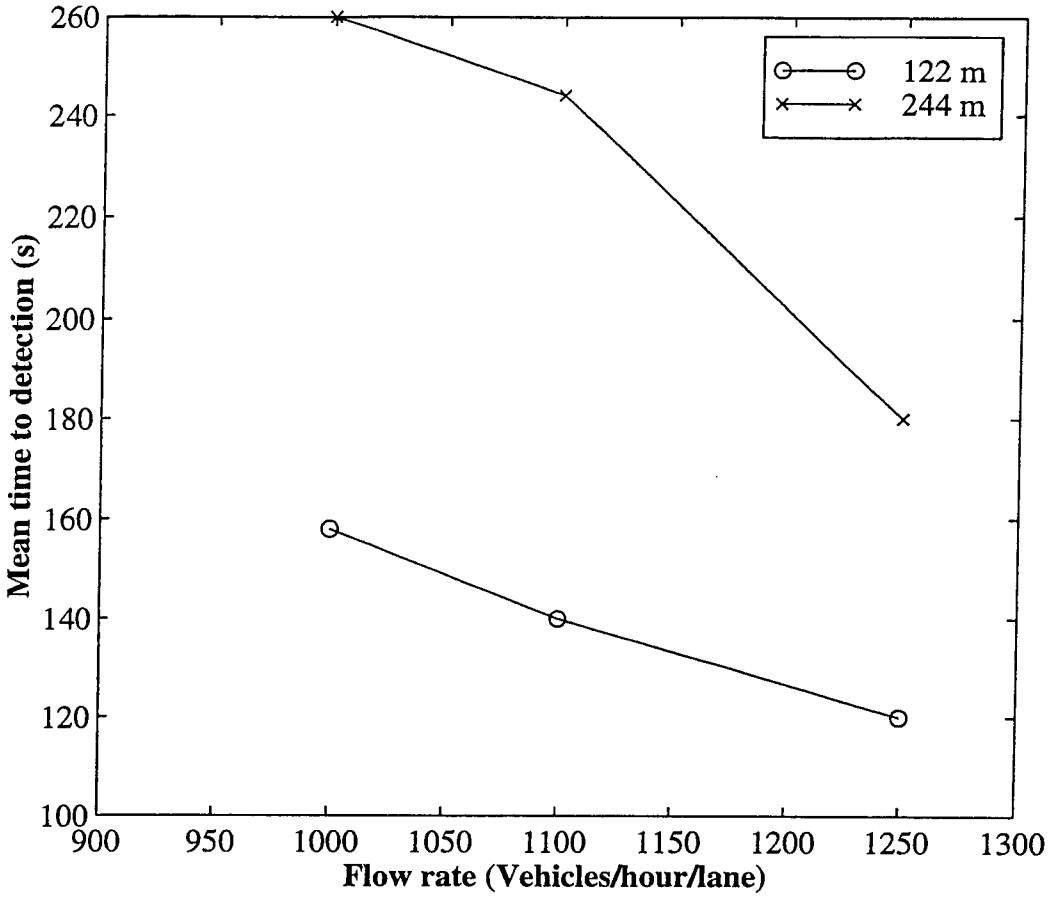

122 m
244 m
Mean time to detection (s)
260
240
220
200
180
160
140
120
100
900
950
1000
1050
1100
1150
1200
1250
1300
Flow rate (Vehicles/hour/lane)

Figure 6

Part 5

INCIDENT DETECTION ALGORITHM USING WAVELET ENERGY REPRESENTATION OF TRAFFIC PATTERNS

Asim Karim[1] and Hojjat Adeli[2]

Abstract: Automatic freeway incident detection is an important component of advanced transportation management systems that provides information for emergency relief and traffic control and management purposes. Earlier algorithms for the freeway incident problems have produced less reliable results especially in recurrent congestion and compression wave traffic conditions. This article presents a new two-stage single-station freeway incident detection model based on advanced wavelet analysis and pattern recognition techniques. Wavelet analysis is used to de-noise, cluster, and enhance the raw traffic data, which is then classified by a radial basis function (RBF) neural network. An energy representation of the traffic pattern in the wavelet domain is found to best characterize incident and non-incident traffic conditions. False alarm during recurrent congestion and compression waves is eliminated by normalization of a sufficiently long time-series pattern. The model is tested under several traffic flow scenarios including compression wave conditions. It produced excellent detection and false alarms characteristics. The model is computationally efficient and can readily be implemented on-line in any ATMS without any need for re-calibration.

[1] Graduate Research Associate. Dept. of Civil and Environmental Engineering and Geodetic Science, The Ohio State University.
[2] Professor. Dept. of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Ave., Columubus, OH, 43210, USA.

INTRODUCTION

An important component of any advanced transportation management system (ATMS) is the reliable and efficient detection of traffic incidents. Traffic incidents on heavy demand freeways can seriously disrupt the performance of the entire highway network. From an engineering point of view the challenge is to localize the disruptive effects of an incident. The key to this problem is the development of an automatic algorithm that immediately recognizes the presence of a congestion-inducing incident so that effective control measures can be taken to prevent the spread of the congestion. A typical urban highway network often has excess capacity at any given time. The goal is to effectively utilize this extra capacity when a bottleneck occurs.

Traffic incident detection algorithms must rely on data obtained at periodic time intervals from traffic sensors or detectors. The common traffic data available for use in incident detection algorithms are the lane occupancy, speed, and flow rate obtained from road sensors located every 500 m to 2 km at usually 20- or 30-second time intervals. Incident detection algorithms must be able to process this information to determine changes in patterns that may indicate an incident condition. However, incident-like patterns may also be produced by non-incident conditions such as recurrent congestion during rush hours and banding of vehicles or compression waves. Traffic incident detection algorithms also have to be able to deal with erroneous data from mal-functioning traffic sensors effectively.

Over the years researchers have developed numerous algorithms for the traffic incident detection (ID) problem (Cook and Cleveland, 1974; Payne and Tignor, 1978; Ahmed and Cook, 1982; Persaud and Hall, 1989; Chassiakos and Stephanedes, 1993; Hsiao et al., 1994; Cheu and Ritchie, 1995; Dia and Rose, 1997; Lin and Daganzo, 1997; Ishak and Al-Deek, 1998; Lin and Chang, 1998; Xu et al., 1998). These algorithms range from earlier simple comparative approaches to more recent pattern recognition and decision-making techniques. The results, in general, have not

been very satisfactory and few freeway management systems today employ an automatic ID algorithm. The complexity arises from both the dynamic and unpredictable nature of traffic flow and the unreliability of the installed traffic sensors, which in turn make simple approaches unreliable.

When a traffic incident reduces the capacity below the prevailing flow rate a queue will form on the upstream direction producing significant reduction in lane speed and significant increase in lane occupancy. This change in pattern is well pronounced. The queue, however, may develop slowly depending on the prevailing flow conditions and the number of lanes closed. Hence the detection time can be large. On the other hand, the change in the flow pattern downstream of a capacity-reducing incident can take place within seconds, independent of the prevailing flow rate before the occurrence of the incident. This change (decrease in lane flow rate and occupancy), however, is not as significant compared with that occurring on the upstream of the incident. It has been argued that an algorithm that uses only the downstream readings produces a high false alarm rate and has difficulty in distinguishing compression waves from incident producing patterns (Weil, et al., 1998). This argument, however, is often based on using algorithms incapable of reliably distinguishing the patterns.

Recently, Adeli and Karim (2000) presented a computational model for automatic traffic incident detection using discrete wavelet transform, fuzzy logic, and neural networks. In their model, the upstream lane occupancy and speed time series data is adopted as the characterizing pattern for traffic state classification. The raw data is first de-noised by soft thresholding in the wavelet domain. Subsequently, the de-noised data is clustered by the fuzzy c-means technique to reduce data dimensionality and enhance feature separation. Finally, a radial basis function neural network is developed to reliably classify the de-noised and clustered pattern. The model is tested with both simulated and real traffic data producing excellent incident detection and false alarm

characteristics. However, the time to detection for the model is long, and depending on the traffic and incident characteristics can be as large as 5 minutes.

In this article, a new traffic incident detection algorithm is presented that distinguishes effectively patterns produced by capacity reducing incidents from those produced by compression waves and recurrent congestion. Furthermore, in most traffic and incident conditions, it signals the presence of an incident within a minute of its occurrence. Only data available locally at each detector station are used for processing Computationally, the algorithm is based on an advanced energy representation of the time-series pattern developed using wavelet theory. This approach effectively enhances the desirable features and de-noises the traffic patterns, which are then classified using a radial basis function (RBF) neural network. The new algorithm is developed, described, and evaluated in the subsequent sections.

FREEWAY INCIDENT DETECTION AND PATTERNS IN TRAFFIC FLOW

A freeway incident detection algorithm determines the presence or absence of an incident condition based on patterns in traffic flow. Therefore, the selection of the number, type, and format of the traffic data to be used is essential to the reliability of the algorithm. Currently, most advanced transportation management systems can provide lane occupancy, speed, and flow rate data from irregularly spaced sensors at regular time intervals. Hence, a reliable incident detection algorithm must be based on the use of such data only. In selecting appropriate patterns for an effective incident detection algorithm we set three goals.

- First, the selected patterns must consistently characterize traffic incident conditions and, at the same time, be distinguishable from other flow conditions such as compression waves.
- Second, the selected patterns by and large should be independent of prevailing roadway and traffic conditions to avoid calibration problems.

- Third, the patterns should indicate an incident condition in less than one minute after the occurrence of incidence.

In this section patterns in traffic data before, during, and after an incident are investigated to determine the most appropriate input for the incident detection algorithm. Note that raw traffic data are analyzed. The pattern identified from this analysis will be processed further to enhance desirable features. The data presented in this section are obtained from TSIS (http://www.fhwa-tsis.com), a traffic simulation software.

Single-Station Versus Two-Station Incident Detection Approaches

A capacity-reducing traffic incident will produce observable changes in flow conditions at the detector stations immediately upstream and downstream of the incident. In general, these changes consist of an increase in traffic congestion upstream and a decrease in traffic congestion downstream of the incident. Based on these observations, two different approaches—called two-station comparative and single-station approaches—have been used to develop traffic incident detection algorithms. The single-station approach relies on data obtained from only one station while the two-station approach makes use of data from two adjacent stations.

The two-station comparative approach, exemplified by the California algorithm (Payne and Tignor, 1978), employs both spatial and temporal data in its algorithm logic. The premise is that using spatial data will reduce false alarms that are produced as a result of changing roadway and traffic conditions because of the natural canceling effect of comparative analysis (Weil et al., 1998; Persaud and Hall, 1989; Payne and Tignor, 1978). The California algorithm is a simple threshold-based algorithm that uses only one flow parameter (occupancy). Also, because of its comparative approach it has to be calibrated at each station to optimize it for the particular roadway geometry.

The two-station comparative approach, in general, has several disadvantages even when advanced pattern recognition techniques are employed. Traffic incidents are temporal events whose

effects develop over time both in the upstream and downstream directions. However, the characteristics of the traffic patterns developed in the upstream and downstream directions are different. Therefore, combining data from both stations is likely to produce less reliable detection of incidents because of the mixing of two different temporal patterns. Two-station comparative algorithms are also more difficult to calibrate because they are affected by the geometry of the roadway, the distance between the stations, the presence of on- and off-ramps, and the prevailing flow conditions.

Figures 1 and 2 show typical time-series plots of lane occupancy, lane speed, and lane flow rate at a station upstream and downstream, respectively, of a lane-blocking incident on a two-lane freeway. Three time-series plots are displayed for three different traffic flow rates of 1000, 1250, 1500 vehicles per hour (vph) per lane. The incident occurs at time 400 second. Note that the time at which the upstream traffic occupancy and speed change (Figures 1a and b) depends on the pre-incident flow rate. The formation of a queue, which produces the significant changes in the traffic occupancy and speed patterns, also depends on the reduction in the capacity and roadway conditions (not presented in the figures). Figure 1c indicates that there is no significant change in the traffic flow on the upstream side. On the other hand, on the downstream side, there are significant changes in the traffic occupancy and flow rate (Figures 2a and c) but no significant change in the traffic speed (Figure 2b). As a result, the two-station comparative algorithms that employ upstream and downstream data together are difficult to calibrate and are likely to produce unreliable detection.

Single-station approaches (Persaud and Hall, 1989; Cook and Cleveland, 1974) do not require data from more than one station to make a decision on the presence or absence of an incident condition. As such, their on-line implementation does not require expensive continuous communication between different detector stations. Furthermore, single-station patterns are not affected by the freeway layout and geometry. Recurring changes in traffic flow such as those

produced by daily rush time traffic and bad weather can be handled effectively by using a normalization technique, as explained later.

In this research our computational model relies on single-station patterns. Our model can handle patterns from both upstream and downstream stations. But, there is no comparison of patterns from the upstream and downstream stations. Rather, each set of patterns are processed independently.

Upstream and Downstream Flow Patterns

From Figures 1 and 2 the pattern formed on the upstream or the downstream side of a capacity-reducing incident each can be used as the basis for an incident detection algorithm. On the upstream side, the dominant flow pattern is the increase in occupancy and the decrease in speed. The flow rate, however, does not show a consistent and significant change as compared to the occupancy and the speed. A pattern based on the upstream time histories of the lane occupancy and speed is therefore most appropriate for reliable incident detection purposes. This conclusion is confirmed by Figure 3, which shows a scatter plot of occupancies and speeds before and after an incident. In this figure, regions of congested and normal flow are generally distinguishable (they can be clearly separated after data de-noising and feature enhancement). On the other hand, the scatter plot of occupancy and flow rate (Figures 4) does not indicate a clear demarcation between normal and congested flow conditions. One limitation of using only the upstream data for an incident detection algorithm is that the detection time may be unacceptably large under low flow rate conditions. The detection time is also dependent on other factors such as distance between detector stations and weather conditions.

Three observations can be made from the time series plots of traffic data on the downstream side of an incident (Figures 2a through c). First, the occupancy and the flow rate decrease rapidly after the occurrence of the incident (in about 20 s or one time interval reported by sensors in the examples of Figures 2a and c). This change, however, is less marked as compared to the increase in

lane occupancy and decrease in lane speed seen on the upstream side. Second, the speed downstream of an incident is not a good indicator of an incident condition, as observed in Figure 2b. After passing through an incident region, vehicles will accelerate and reach free flow speeds rather quickly. Third, the times at which the occupancy and the flow rate decrease appreciably are about the same and relatively independent of the flow rate.

The scatter plots of occupancy and speed (Figures 5) and occupancy and flow rate (Figure 6) for data from a location downstream of an incident show that there are no discernable and separable regions for before and after incident flow conditions. Because of this the development of a reliable algorithm for incident detection based on data from the downstream side has proven to be more difficult. Using the downstream data poses two additional challenges. First, there is the risk of false alarms as a result of compression waves because a compression wave's occupancy and flow rate downstream patterns resemble those of an incident. Second, the magnitudes of the flow rate data on the downstream side may vary because of weather conditions, the severity of the capacity reduction as a result of the incident, and other daily changes in the flow rate. On the other hand the major advantage of using the downstream data is that the change in pattern after an incident is almost immediate and independent of the prevailing flow rate.

Based on these observations a new incident detection logic and computational model is developed that utilizes both upstream and downstream traffic patterns independently. A two-stage logic is employed. In the first stage, the presence or absence of an incident condition is determined from the downstream occupancy and flow rate time-series data. The second stage confirms the presence or the absence of an incident condition by using the upstream occupancy and speed time-series data. To minimize the possibility of a missed detection and eliminate false alarms an advanced wavelet-based feature enhancement and de-noising approach is adopted to process the data. False alarms from compression waves are avoided by using a sufficiently long time series as

input. Recurrent congestion is handled by a normalization technique. These models are developed in detail in subsequent sections.

DISCRETE WAVELET TRANSFORM AND SIGNAL ENERGY

The discrete wavelet transform (DWT) provides a powerful and efficient technique for analyzing, decomposing, de-noising, and compressing signals. In particular, the DWT of a signal breaks it down into several time-frequency components that enable the extraction of features desirable for signal identification and recognition. The DWT and wavelet theory in general have been developed rapidly in the last 10 years (Daubechies, 1992, Burrus et al., 1998). In this section the basic concepts of DWT and its energy representation employed in this research are presented briefly. Additional details of DWT and its application in ITS problems can be found in Samant and Adeli (2000).

A one-dimensional signal $f(t) \in L^2(R)$ can be decomposed into multiresolution components that are indexed by the scale j (indicator of frequency) and the translation k (indicator of time):

$$f(t) = \sum_k c_{j_0,k}\varphi_{j_0,k}(t) + \sum_k \sum_{j=j_0} d_{j,k}\psi_{j,k}(t) \qquad (1)$$

where $L^2(R)$ is the space of all square integrable functions defined in the one-dimensional real space R, $c_{j,k}$ is the scaling coefficient corresponding to the scaling function $\varphi_{j,k}(t)$, and $d_{j,k}$ is the wavelet coefficient corresponding to wavelet $\psi_{j,k}(t)$. The index j_0 represents the lowest resolution that is decomposed by the DWT. The functions $\varphi_{j,k}(t)$ $(j,k \in Z)$ and $\psi_{j,k}(t)$ $(j,k \in Z)$ (Z is the space of integers), each forming a basis of $L^2(R)$, are defined by the following equations:

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k) \qquad (2)$$

$$\varphi(t) = \sum_k h_0[k]\sqrt{2}\varphi(2t-k) \qquad k \in Z \qquad (3)$$

$$\psi(t) = \sum_k h_1[k]\sqrt{2}\varphi(2t-k) \qquad k \in Z \tag{4}$$

where h_0 and h_1 are filter coefficients and the constant $\sqrt{2}$ maintains the unity norm of the functions. In this work, the Daubechies wavelet system of order eight (Daubechies, 1992), defined by eight h_1 and h_0 coefficients, is used. This wavelet basis system is selected because of its orthonormality property and compact support providing a DWT with a finite length and number of wavelet coefficients.

When an orthonormal basis is used the coefficients $c_{j,k}$ and $d_{j,k}$ are given by the inner product of the signal with the appropriate function:

$$c_{j,k} = c_j[k] = \int f(t)\varphi_{j,k}(t)dt \quad \forall j,k \tag{5}$$

$$d_{j,k} = d_j[k] = \int f(t)\psi_{j,k}(t)dt \quad \forall j,k \tag{6}$$

which can be reduced to the following recursive equations (Burrus et al., 1998):

$$c_j[k] = \sum_m h_0[m-2k]c_{j+1}[m] \tag{7}$$

$$d_j[k] = \sum_m h_1[m-2k]c_{j+1}[m] \tag{8}$$

In these equations it is assumed that the scaling coefficients of the signal at the highest resolution are known.

The traffic data are available as a discrete sequence $f[k]$ of finite length $L = 2^J$ where J is an integer. The highest resolution part of the scaling function $\varphi_{j,k}(t)$, $\varphi_{J,k}(t)$, will approach a Dirac delta function and Eq. (5) will represent a sampling of $f[k]$. Therefore, $c_J[k]$ can be approximated by $f[k]$. Use of the recursive Eqs. (7) and (8) for calculating the DWT coefficients requires that $f[k]$ be extended periodically. In other words, the following equation should hold:

$$f[k] = f[k+Ln] \quad n = 1,2,3,... \tag{9}$$

However, traffic time-series data, such as those shown in Figures 1 and 2, are not periodic. In other words, generally the end values $f[1]$ and $f[L]$ are not equal. As a result of the incompatibility of the traffic data with the periodic boundary condition, the wavelet representation can distort the shape of the original traffic pattern. To overcome this problem the traffic pattern is extended on either ends before its DWT is found. This procedure is explained in detail in the next section.

An advantage of using an orthonormal basis to find the DWT of a signal is that the energy of the signal can be partitioned into its various time-frequency components. The energy contribution from each component is expressed as a function of the wavelet and scaling coefficients. This is known as Parseval's theorem and is expressed mathematically in the form of the following energy functional (Burrus et al., 1998):

$$\int |f(t)|^2 dt = \sum_k |c_{j_0,k}|^2 + \sum_k \sum_{j=j_0} |d_{j,k}|^2 \qquad (10)$$

We use this functional to enhance the traffic data streams for the purpose of pronouncing the traffic incident patterns, as explained in the next section.

TRAFFIC PATTERN FEATURE ENHANCEMENT AND DE-NOISING

In our traffic incident detection model, we process the three time-series traffic data (lane occupancy, speed, and flow rate) obtained at each detector station with the objectives of reducing the noise and enhancing the desirable features. This processing is essential to ensure that no incidents go undetected and no false alarms are triggered. The upstream lane occupancy ($f_O[i]$) and speed ($f_S[i]$) form one pattern for identifying incident conditions. The downstream lane occupancy ($f_O[i]$) and flow rate ($f_F[i]$) form another pattern for identifying incident conditions.

Sixteen data points are selected for each one of the three traffic parameters. That is, the sequences $f_O[i]$, $f_F[i]$, and $f_S[i]$ consist of sixteen values indexed from 1 to 16. There are two

reasons for selecting this length for each time-series. The DWT used in this work (and in fact in most cases) requires that the number of data points to be equal to some power of 2 (4, 8, 16, etc.). For algorithmic efficiency, the smallest number is preferred. We found 16 to be the minimum number needed to avoid false alarms that may be caused by compression waves. We found this necessary for the downstream pattern ($f_O[i]$ and $f_F[i]$) which may exhibit similar patterns for both compression waves and incident conditions.

When the time interval between successive readings is 20 seconds (which is the minimum available from current detector stations) sixteen data points constitute 5 minute and 20 second of data. Compression waves are usually temporary conditions and not very likely to exist for as long as 5 minutes. In other words, it is unlikely that a pattern in which the values of $f_O[i]$ and $f_F[i]$ (i = 15, 16) are much smaller than the values of $f_O[i]$ and $f_F[i]$ (i = 1, 2,...,14) is caused by a compression wave. This data sampling strategy prevents the downstream pattern from signaling an incident condition erroneously whenever a compression wave passes by.

The traffic time-series data are normalized by dividing them by the average of the highest two values in each series. Normalization reduces the significance of magnitude in the pattern recognition process and the undesirable domination of a single large value. Patterns are distinguished primarily on the basis of their shape and form and not on the basis of magnitude. As a result, the normalization technique also eliminates the need for re-calibration whenever the flow condition changes. Flow variations caused by daily rush time traffic, weather conditions, geometry, and other situations can therefore be handled automatically and transparently. The normalized occupancy, speed, and flow rate sequences are represented as $\bar{f}_O[i]$, $\bar{f}_S[i]$, and $\bar{f}_F[i]$, respectively.

The normalized data series are extended by 8 points at each end before their DWT's are calculated as follows:

$$\hat{f}[i] = \begin{cases} 0.5\left(\bar{f}[1] + \bar{f}[2]\right) & 1 \le i \le 8 \\ \bar{f}[i-8] & 9 \le i \le 24 \\ 0.5\left(\bar{f}[15] + \bar{f}[16]\right) & 25 \le i \le 32 \end{cases} \tag{11}$$

The length L of each data series now becomes 32 (i.e. $L = 2^5$ and $J = 5$). The need for extending the data series is shown in Figures 7a and b. Figure 7a shows a typical flow rate data series, $\bar{f}_F[i]$ (solid line), on the downstream side of an incident and its scale 3 (i.e. $j = 3$) wavelet approximation (dashed line). Notice how the shape of the wavelet approximation is distorted at the left edge because of the periodic boundary condition assumption. Figure 7b shows the same data series extended using Eq. (11) (solid line) and its scale 3 wavelet approximation (dashed line). In this figure the wavelet distortion has been pushed aside to the outer edges, outside the usable region of data, the segment from data points 9 to 24. In this segment the basic shape of the original data series is preserved without distortions.

In the new traffic incident detection model, the DWT is employed to reduce the dimensionality of input data for the neural network pattern classifier, eliminate the traffic noise, and enhance the desirable features in each data series. The extended data series has a length of 2^5 and is represented by scale $J = 5$ in Eq. (5). Equation (7) is applied two times recursively to calculate the scaling coefficients at scale $j = 3$. This operation corresponds to a two-stage low-pass filtering of $c_J[k]$ with h_0 (Samant and Adeli, 2000). At this reduced resolution the higher frequency noise-like components are eliminated leaving a smoother de-noised shape or form. Also, through the two-stage low-pass filtering the 32-point time-series is now reduced to an 8-coefficient representation. However, this DWT is for the extended 32-point data series. The DWT of the original 16-point data series is given by the middle 4 values of the 8 coefficients ($c_3[k]$, $k = 3, 4, 5, 6$). Let these reduced sets of coefficients be defined as $c_O[i]$, $c_S[i]$, and $c_F[i]$ for occupancy, speed, and flow rate, respectively, where $i = 1, 2, 3, 4$.

Notice from Figures 1 and 2 that an incident condition pattern exhibits either a sudden decrease or a sudden increase in magnitude of data values which occur in the last few data points. This feature, which distinguishes an incident condition from a non-incident condition, can be enhanced by using the energy representation capability of wavelet transforms (Eq. 10). The squares of the absolute values of the coefficients $c[i]$ represent the energy of the de-noised time-series data at each time location defined by index i. The energy (or the area under a squared time-series plot) enhances incident condition patterns and distinguishes them from non-incident condition patterns. Thus, the scaling coefficients are modified as follows:

$$\hat{c}[i] = |c[i]|^2 \quad \forall i \tag{12}$$

The benefit of DWT-based de-noising and feature enhancement is demonstrated in Figures 8 and 9. Figure 8 is a scatter plot of $\hat{c}_O[i]$ and $\hat{c}_S[i]$ based on the same data used in Figure 3. Figure 9 is a scatter plot of $\hat{c}_O[i]$ and $\hat{c}_F[i]$ based on the same data used in Figure 6. Comparisons of Figure 3 with Figure 8 and Figure 6 with Figure 9 indicate the improvement in pattern separation achieved by wavelet-based de-noising and feature enhancement. The points between cluster regions seen in these figures are intermediate conditions that will move to one of the clusters as the time-series pattern becomes more defined with time.

The enhanced traffic pattern at the upstream side, $x_U[i]$, is then formed by concatenating the 4 coefficients from the occupancy and the speed data series. Similarly, the enhanced traffic pattern on the downstream side, $x_D[i]$, is formed by concatenating the occupancy and flow rate data series coefficients. Mathematically, the patterns are given by

$$x_U = \{\hat{c}_O[i], \hat{c}_S[i]\} \quad i = 1,2,3,4 \tag{13}$$

$$x_D = \{\hat{c}_O[i], \hat{c}_F[i]\} \quad i = 1,2,3,4 \tag{14}$$

PATTERN CLASSIFICATION USING RADIAL-BASIS FUNCTION NEURAL NETWORK

Neural networks are powerful model-free pattern classifiers (Adeli and Hung, 1995). However, they can be computationally very expensive when the size or dimensionality of the input data is large requiring a very large number of training instances. Training instances of the traffic patterns defined by Eqs. (13) and (14) are used to develop a mapping from an 8-dimensional space to a one-dimensional space. For this purpose, the radial basis function (RBF) neural network is adopted. The RBF neural network is an efficient universal classifier (Moody and Darken, 1989) that has a simple topology consisting of a hidden layer of nodes with nonlinear transfer functions and an output layer of nodes with linear transfer functions.

The topology of the RBF neural network developed for the traffic pattern classification is shown in Figure 10. The input layer has 8 nodes corresponding to the eight data points in each pattern ($x_U[i]$ or $x_D[i]$, henceforth called vector $\mathbf{x}$). The number of nodes in the hidden layer, N_h is equal to the number of cluster centers used to characterize the input training space. The output layer has one node (y). The number of nodes in the hidden layer is chosen as a fraction of the total number of training instances. This choice is based on numerical experimentation to determine which number adequately covers the input space and produces the best mapping. We found a number within the range of 10 to 30% of the number of training instances to provide satisfactory results. The cluster centers $\boldsymbol{\mu}_i$ $(1 \le i \le N_h)$ is obtained using the fuzzy c-means algorithm (Bezdek, 1981; Cannon et al., 1986).

The connection from the input node i to the hidden node j is assigned the weight μ_{ji} corresponding to the ith component of the vector $\boldsymbol{\mu}_j$. The output of a hidden node j is given by the following Gaussian transfer function:

$$\phi_j = \exp\left(-\frac{\|\mathbf{x}-\boldsymbol{\mu}_j\|^2}{2\sigma_j^2}\right) \tag{15}$$

where the factor σ_j controls the spread or range of influence of the Gaussian function centered at $\boldsymbol{\mu}_j$. In this work σ_j is calculated as

$$\sigma_j = \frac{1}{48}\sum_{i=1}^{N}\|\boldsymbol{\mu}_j - \boldsymbol{\mu}_i\| \quad 1 \le j \le 12 \tag{16}$$

where N is the total number of training instances. Equation (16) approximates the spread parameter σ_j as one third of the mean distance between cluster centers. The connection from the hidden node j to the output node is assigned the weight λ_j. The output y of the network is then given by

$$y = \sum_{j=1}^{N}\phi_j\lambda_j \tag{17}$$

Theoretically an output value of 1 corresponds to an incident classification while an output value of –1 corresponds to a no incident classification. Practically, however, one has to choose a threshold value for distinguishing between the two classes as the output from Eq. (17) can take any value in the range -1 and 1.

The weights λ_j are calculated by minimizing the error between the network computed output y and the desired output y_d based on training examples. In other words, to train the network for λ_j's we solve the following unconstrained optimization problem:

$$\text{Minimize } E(\boldsymbol{\lambda}) = \sum_{i=1}^{N}\left|y^i - y_d^i\right| \tag{18}$$

The gradient descent optimization algorithm is used to solve this optimization problem.

MODEL TESTING

Introduction

The new computational model for freeway incident detection is tested using both real and simulated traffic data. More than 40 hours of simulated traffic data is generated from the traffic simulation software TSIS/CORSIM while real traffic data is obtained from the freeway service patrol (FSP) project's I-880 database. A large portion of the simulated data is made up of incident or incident-like conditions on two- and three-lane freeways. This is an advantage of employing a simulation software for testing purposes as sufficient quantities of reliable real data with traffic incidents are not readily available. Furthermore, with a data generating software it is possible to study the performance of the model under various traffic flow scenarios. The real data is used for further validation of the model.

Training

The model is trained using a sample of 30 incident and 30 non-incident patterns extracted from the simulated data. Two RBF neural networks are trained: one for the upstream detector station and the other for the downstream detector station. Training is done only once and no re-calibration or re-training is needed. The RBF classifier can therefore be implemented on-line on all stations after the training is done off-line.

First Test Using Simulated Data: Two-lane Freeway

The performance of the incident detection model on a two-lane freeway (in each direction) is shown in Table 1. The prevailing flow rate per lane is varied from 1000 to 2000 vehicles per hour (vph). The traffic incident consists of the blockage of one lane (the blockages are distributed evenly between the lanes) and a 50 percent reduction in capacity of the adjacent lane. In 600 different simulations the algorithm detects all incidents both at the downstream and the upstream detector stations. One false alarm is produced at the downstream station when the demand is a low 1000 vph

per lane. The data that caused this false alarm exhibited a pattern similar to that of an incident condition pattern. This situation will occur rarely in practice and only in low flow conditions. A sensor malfunction may also cause a false alarm. But this can be handled easily in the preprocessing logic as most sensors report their operation status regularly. False alarms can be eliminated completely by using a slightly higher transition threshold from non-incident to incident condition on the RBF classifier output. In this first test scenario the threshold was kept at zero to validate the pattern recognition properties of the model.

The average incident detection time for the downstream detector station is 46.5 seconds with a range varying from 40 to 54 seconds. This is an acceptable delay for practically all emergency and control purposes. Also, there is practically no variation of this time with any change in flow rate and location of the incident. This result is significantly better than that reported by Adeli and Karim (2000) where the detection time is as large as 5 minutes. The time to detection for the upstream detector station, on the other hand, does vary significantly with the flow rate and the distance of the incident from the detector station. It varies from 70 to 228 seconds. The upstream pattern is based on the formation of a queue that takes a rather long time to develop (in the order of one to four minutes).

In the subsequent test scenarios the threshold value was taken as 0.2 where an output greater or equal to 0.2 was signaled as an incident while a value less than 0.2 was labeled as a non-incident. This was intended to eliminate the false alarms but at the expense of slightly more detection times.

Second Test Using Simulated Data: Three-lane Freeway

Table 2 shows the performance of the model on a 3-lane freeway for flow rates ranging from 1250 vph to 2000 vph per lane. Only one lane (either the lane adjacent to the shoulder or the median) is blocked in this scenario with no reduction in capacity of the other lanes. This scenario simulates a shoulder or median obstruction that also requires the closure of the adjacent traffic lane.

Under this scenario in 600 different traffic simulations the downstream detector station produced perfect results while the upstream detector station missed 4 incidents during low demand conditions. The missed detections by the upstream detection station are understandable because the remaining capacity (about 4000 vph) is still able to handle the demand (3750 vph) without the development of significant congestion on the upstream side. On the other hand, the downstream detector station is able to detect all incidents within about a minute of its occurrence. This test scenario illustrates the capability of the model under low demand conditions and minor obstructions, situations in which many algorithms produce poor detection and numerous false alarms.

Third Test Using Simulated Data: Compression Waves

To test the model's performance under compression wave-like conditions one hundred minutes of data are generated for a two-lane freeway with moderate flow rate and with several periods of increased flow rate by up to 500 vph. The periods of increased flow rate are limited to 5 minutes or less based on the assumption that compression waves are temporary conditions. A typical 25-minute plot of lane occupancy is shown in Figure 11. The higher flow rate period lasts from 600 to 900 seconds. In all, there are 374 patterns in this 100-minute data. The model correctly identified all of them as non-incident conditions.

Fourth Test Using Real Data: FSP Project's I-880 Database

The freeway service patrol (FSP) project's database contains traffic data for a 14.8 km (9.2 mile) long segment of the I-880 freeway between Oakland and San Jose, California. This segment has a varied geometry of 3 to 5 lanes (in each direction), single and multiple lane on- and off-ramps, and mild horizontal and vertical curvatures. Over the duration of the project observers in patrol vehicles traversed this freeway segment and recorded the occurrence of incidents by noting down key incident characteristics such as location, time, and type of incident. By correlating this

information with data obtained from sensors, samples for 21 lane blocking incidents are extracted from the database. To test the false alarm rate performance, 4 hours of incident free data are also extracted. Table 3 shows the performance of the new incident detection model using real data. Both downstream and upstream stations produced a detection rate of 95.2 percent and a false alarm rate of zero. This result is identical to that reported by Adeli and Karim (2000). Accurate information for the time of occurrence of incidents is not available from the database. Thus, the detection times for the model cannot be computed.

Result Summary and Comparison

The results of the new incident detection model indicate that the downstream detector station data and logic by themselves provide satisfactory results. In an ATMS that does not provide speed data the upstream station logic can be eliminated. However, in situations where the speed data is available the upstream detector station logic provides an additional level of reliability without any significant increase in computation. The results also show the calibration free transferability of the model where the model trained using simulated data performs reliably when tested using both real and simulated data. As compared to the fuzzy-wavelet RBFNN model presented by Adeli and Karim (2000), the new model produces significantly shorter detection times without any loss in detection and false alarm rate performance. Furthermore, the new model is computationally more efficient as it does not require the compuation of the inverse wavelet transform and the fuzzy c-mean at each time interval.

CONCLUSION

A new traffic incident detection logic and computational model is presented that overcomes several shortcomings of earlier algorithms. The model uses a two-stage single-station detection logic. In the first stage a decision is made based on data obtained from the downstream detector

station only while in the second stage the decision is confirmed based on data obtained from the upstream detector station only. Wavelet domain processing is used to de-noise, compress, and enhance the raw traffic data for classification. It is found that an energy representation of the data best characterizes incident and non-incident conditions. The model determines the state of the traffic flow from the shape of the time-series data rather than the magnitude. A radial basis function neural network is developed to classify the processed traffic data into incident and non-incident states.

The new model has the following five advantages and desirable characteristics. No other existing incident detection algorithm can provide all of them simultaneously.

- The new model is capable of detecting all incidents even when the reduced freeway capacity after the incident is greater than the prevailing flow rate (normally occurring under low flow rate conditions).
- The model can reliably identify recurrent congestion and compression waves a non-incident conditions without triggering a false alarm.
- The model signals the presence of an incident within one minute of its occurrence, to a great extent independent of the prevailing traffic and roadway conditions.
- The model does not require re-calibration for its on-line implementation and thus is readily transferable.
- The model is computationally highly efficient because a) DWT operations require a small number of multiplications and additions in every sensor reporting interval (say 20 seconds) and b) we have reduced the dimensionality of the RBF neural network through wavelet-based energy representation of input.

These characteristics make our new traffic incident detection model ideal for widespread practical adoption in urban ATMS. The model was tested under several traffic flow scenarios. In

general, it produced excellent results across a wide range of prevailing flow conditions. The model also correctly identified compression wave conditions and none of them were signaled as false alarms.

ACKNOWLEDGMENT

This manuscript is based on a research project sponsored by the Ohio Department of Transportation and Federal Highway Administration.

APPENDIX I. REFERENCES

Adeli, H. and Hung, S. L. (1995), *Machine Learning--Neural Networks, Genetic Algorithms, and Fuzzy Systems*, John Wiley & Sons, Inc., New York, NY.

Adeli, H. and Karim, A. (2000), "A Fuzzy-Wavelet RBFNN Model for Freeway Incident Detection," *Journal of Transportation Engineering*, ASCE, Vol. 126, No. 6, pp. 464-471.

Adeli H. and Samant, A (2000), "An Adaptive Conjugate Gradient Neural Network-Wavelet Model for Traffic Incident Detection", *Computer-Aided Civil and Infrastructure Engineering*, Vol. 15, No. 4, pp. 251–260.

Ahmed, S. A. and Cook, A. R. (1982), "Application of Time-Series Analysis Techniques to Freeway Incident Detection," *Transportation Research Record*, No. 841, pp. 19-28.

Bezdek, J. C. (1981), *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY.

Burrus, C. S., Gopinath, R. A. and Guo, H. (1998), *Introduction to Wavelets and Wavelet Transforms--A Primer*, Prentice-Hall, Inc., Upper Saddle River, NJ.

Cannon, R. L., Dave, J. V. and Bezdek, J. C. (1986), "Efficient Implementation of the Fuzzy c-Means Clustering Algorithms," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 2, pp. 248-255.

Chang, E. C. and Wang, S. -H. (1994), "Improved Freeway Incident Detection Using Fuzzy Set Theory," *Transportation Research Record*, No. 1453, pp. 75-82.

Chassiakos, A. P. and Stephanedes, Y. J. (1993), "Smoothing Algorithms for Incident Detection," *Transportation Research Record*, No. 1394, pp. 8-16.

Cheu, R. L. and Ritchie, S. G. (1995), "Automated Detection of Lane-Blocking Freeway Incidents Using Artificial Neural Networks," *Transportation Research-C*, Vol.3, No. 6, pp. 371-388.

Cook, A. R. and Cleveland, D. E. (1974), "Detection of Freeway Capacity-Reducing Incidents by Traffic Stream Measurements," *Transportation Research Record*, No. 495, pp. 1-11.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA.

Dia, H. and Rose, G. (1997), "Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data," *Transportation Research--C*, Vol. 5, No. 5, pp. 313-331.

Hsiao, C. -H., Lin, C. -T. and Cassidy, M. (1994), "Applications of Fuzzy Logic and Neural Networks to Automatically Detect Freeway Traffic Incidents," *Journal of Transportation Engineering*, ASCE, Vol. 120, No. 5, pp. 753-772

Ishak, S. S. and Al-Deek, H. M. (1998), "Fuzzy ART Neural Network Model for Automated Detection of Freeway Incidents," *Transportation Research Record*, No. 1634, pp. 56-63.

Lin, C. -K. and Chang, G. -L. (1998), "Development of a Fuzzy-Expert System for Incident Detection and Classification," *Mathematical and Computer Modelling*, Vol. 27, No. 9-11. pp. 9-25.

Lin, W. -H. and Daganzo, C. F. (1997), "A Simple Detection Scheme for Delay-Inducing Freeway Incidents," *Transportation Research--Part A*, Vol. 31, No. 2, pp. 141-155.

Moody, J. and Darken, C. J. (1989), "Fast Learning in Networks of Locally-Tuned Processing Units," *Neural Computation*, Vol. 1, pp. 281-294.

Payne, H. J. and Tignor, S. C. (1978), "Freeway Incident-Detection Algorithms Based on Decision Trees With States," *Transportation Research Record*, No. 682, pp. 30-37.

Persaud, B. N. and Hall, F. L. (1989), "Catastrophe Theory and Patterns in 30-Second Freeway Traffic Data--Implications for Incident Detection," *Transportation Research--Part A*, Vol. 23A, No. 2, pp. 103-113.

Samant, A. and Adeli, H. (2000), "Feature Extraction for Traffic Incident Detection using Wavelet Transform and Linear Discriminant Analysis", *Computer-Aided Civil and Infrastructure Engineering*, Vol. 15, No. 4.

Weil, R., Wootton, J. and Garcia-Ortiz, A. (1998), "Traffic Incident Detection: Sensors and Algorithms," *Mathematical and Computer Modelling*, Vol. 27, No. 9-11, pp. 257-291.

Xu, H., Kwan, C. M., Haynes, L. and Pryor, J. D. (1998), "Real-Time Adaptive On-Line Traffic Incident Detection," *Fuzzy Sets and Systems*, Vol. 93, pp. 173-183.

Table 1. Performance of the new incident detection model on a two-lane freeway

Flow rate (vph per lane)	Location (m)*	Downstream station **			Upstream station **		
		Detections	False alarms	Detection time (s)	Detections	False alarms	Detection time (s)
1000	244	10/10	1/40	50	10/10	0/40	192
1000	122	10/10	0/40	40	10/10	0/40	142
1100	244	10/10	0/40	40	10/10	0/40	228
1100	122	10/10	0/40	40	10/10	0/40	126
1250	244	10/10	0/40	48	10/10	0/40	172
1250	122	10/10	0/40	46	10/10	0/40	110
1500	244	10/10	0/40	48	10/10	0/40	130
1500	122	10/10	0/40	48	10/10	0/40	82
1750	244	10/10	0/40	44	10/10	0/40	114
1750	122	10/10	0/40	48	10/10	0/40	70
2000	244	10/10	0/40	54	10/10	0/40	88
2000	122	10/10	0/40	52	10/10	0/40	70
Totals		120/120 100 %	1/480 0.2 %		120/120 100 %	0/480 0 %	

* Distance of the traffic incident from the upstream station. Distance between stations is 460 m.
** Numbers after / indicate the total number of simulations.

Table 2. Performance of the new incident detection model on a 3-lane freeway

Flow rate (vph per lane)	Location (m)*	Downstream station **			Upstream station **		
		Detections	False alarms	Detection time (s)	Detections	False alarms	Detection time (s)
1250	244	10/10	0/140	40	6/10	0/140	435
1500	244	10/10	1/140	42	10/10	0/140	320
1833	244	10/10	0/140	48	10/10	0/140	292
2000	244	10/10	0/140	66	10/10	0/140	248
Totals		40/40 100 %	1/560 0.18 %		36/40 90 %	0/560 0 %	

* Distance of the traffic incident from the upstream station. Distance between stations is 460 m.
** Numbers after / indicate the total number of simulations.

Table 3. Performance of the new incident detection model using real data from the FSP project's database

Dowstream station *		Upstream station *	
Detections	False alarms	Detections	False alarms
20/21	0/480	20/21	0/480
95.2 %	0 %	95.2 %	0 %

* Numbers after / indicate the total number of tests

LIST OF CAPTIONS FOR FIGURES

1. Time-series plots of upstream traffic data on a two-lane freeway with three prevailing flow rates of 1000, 1250, and 1500 vph per lane before and after an incident. (a) lane occupancy plot, (b) lane speed plot, (c) lane flow rate plot
2. Time-series plots of downstream traffic data on a two-lane freeway with three prevailing flow rates of 1000, 1250, and 1500 vph per lane before and after an incident. (a) lane occupancy plot, (b) lane speed plot, (c) lane flow rate plot
3. Scatter plot of upstream lane occupancy and speed before and after incidents
4. Scatter plot of upstream lane occupancy and flow rate before and after incidents
5. Scatter plot of downstream lane occupancy and speed before and after incidents
6. Scatter plot of downstream lane occupancy and flow rate before and after incidents
7. (a) DWT of a 16-point flow rate traffic pattern
   (b) DWT of an extended 32-point flow rate traffic pattern (based on the data of Figure 7a)
8. Scatter plot of upstream lane occupancy and speed wavelet energy coefficients before and after incidents
9. Scatter plot of downstream lane occupancy and flow rate wavelet energy coefficients before and after incidents
10. Topology of radial basis function neural network for traffic pattern classification
11. Typical lane occupancy time-series plot for compression wave traffic condition

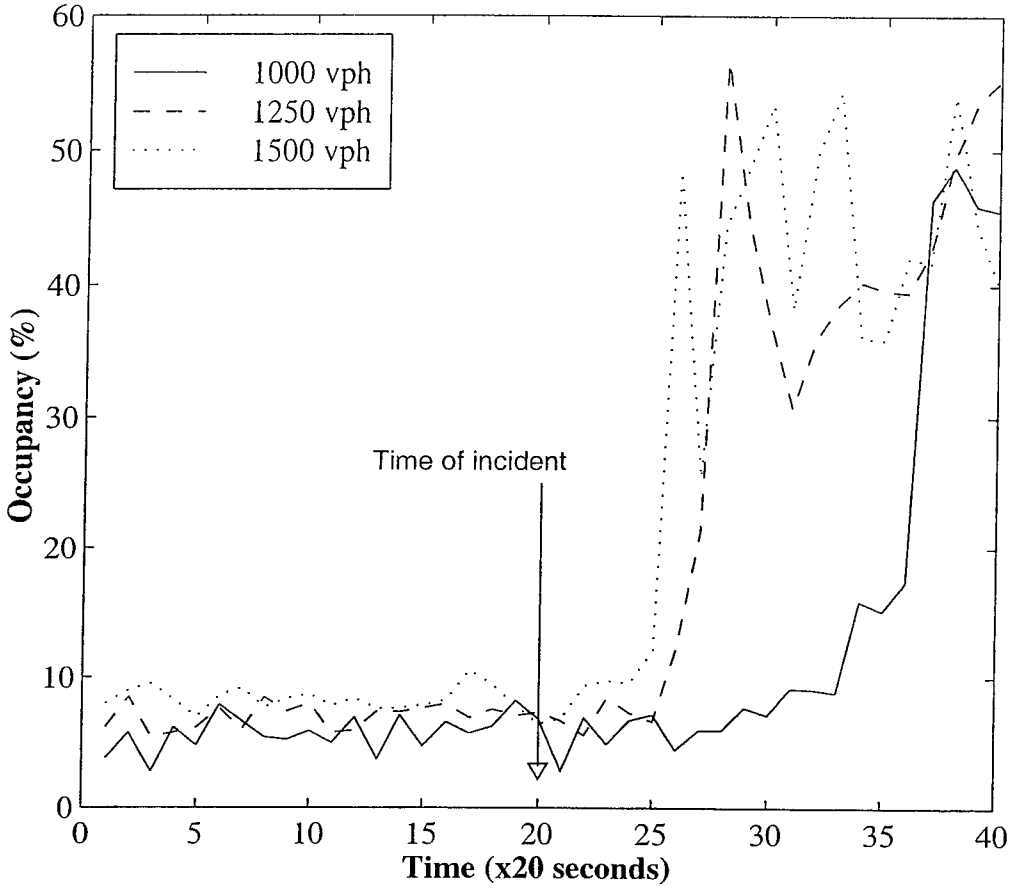

1000 vph
1250 vph
1500 vph
Time of incident
Occupancy (%)
Time (x20 seconds)
0
10
20
30
40
50
60
0
5
10
15
20
25
30
35
40

Figure 1a

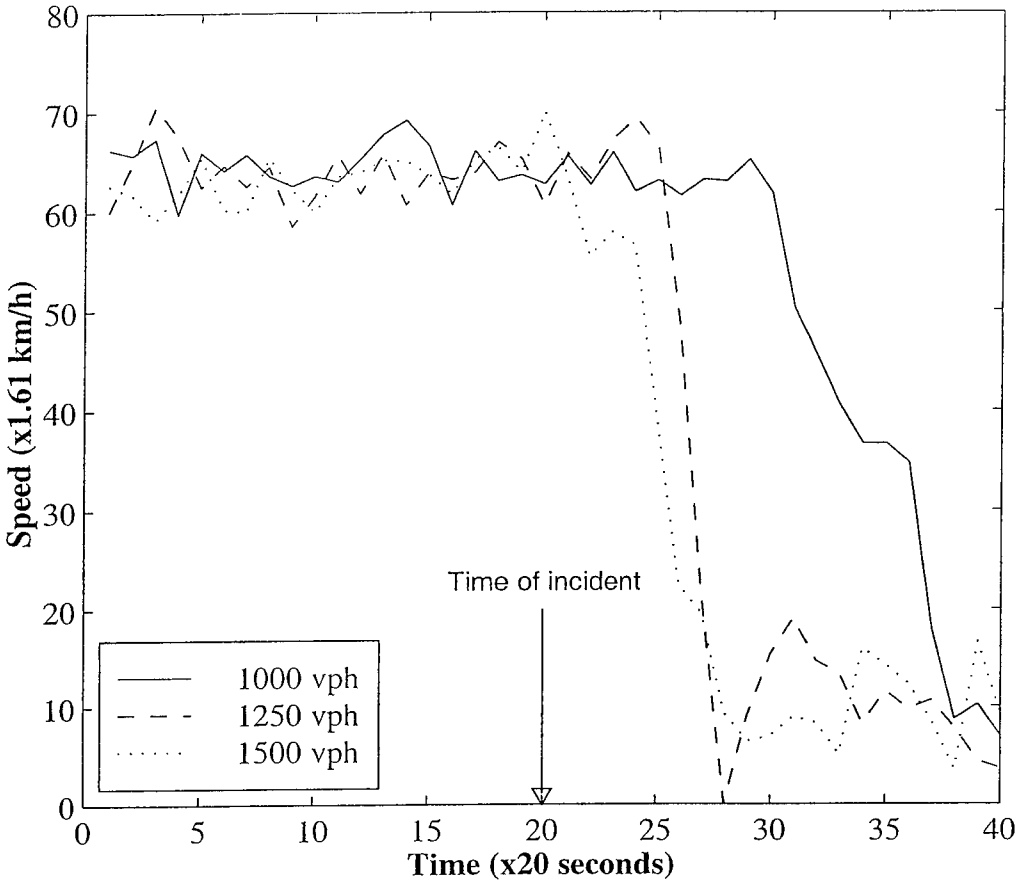

80
70
60
50
40
30
20
10
0
Speed (x1.61 km/h)
0
5
10
15
20
25
30
35
40
Time (x20 seconds)
Time of incident
1000 vph
1250 vph
1500 vph

Figure 1b

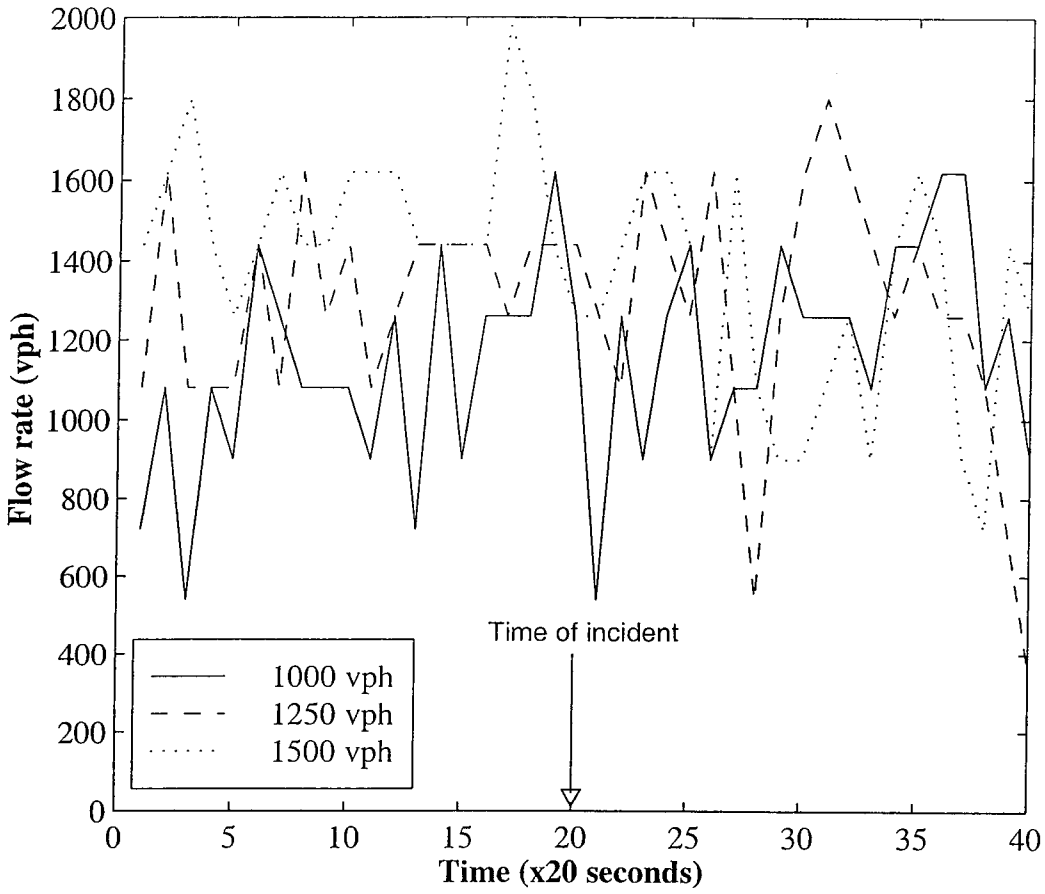

2000
1800
1600
1400
1200
1000
800
600
400
200
0
Flow rate (vph)
0
5
10
15
20
25
30
35
40
Time (x20 seconds)
Time of incident
1000 vph
1250 vph
1500 vph

Figure 1c

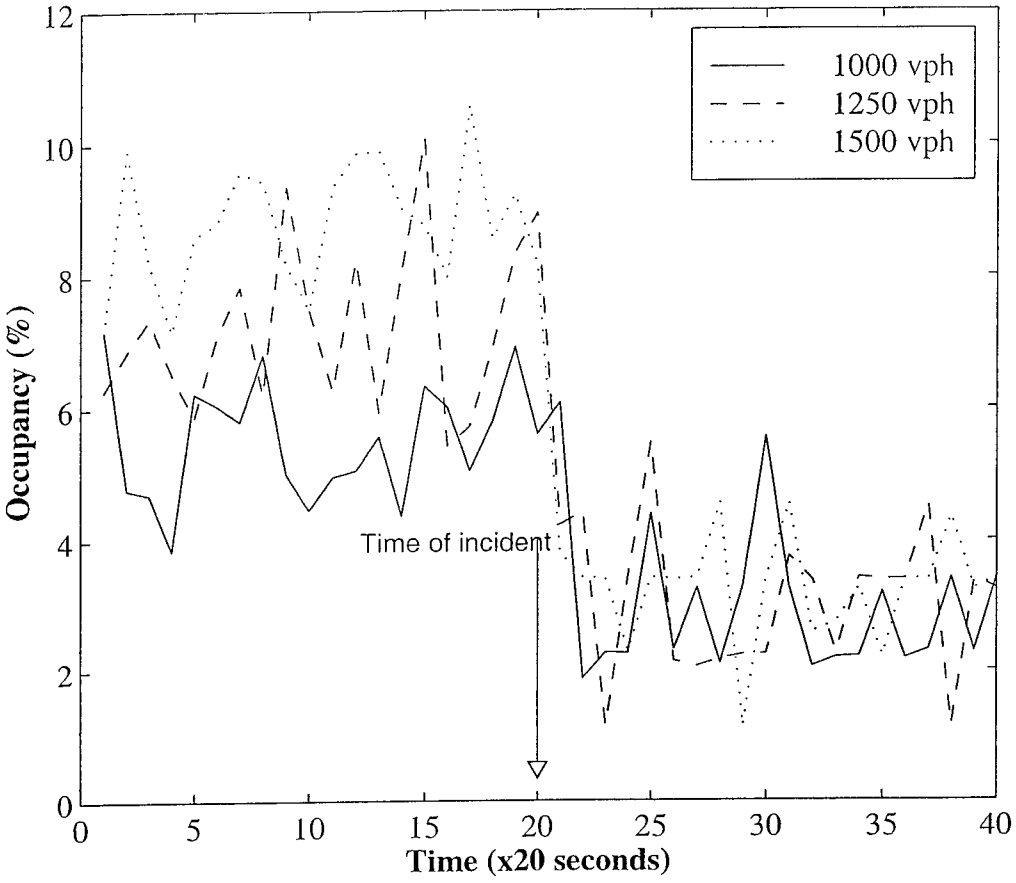

1000 vph
1250 vph
1500 vph
Time of incident
Occupancy (%)
Time (x20 seconds)
0
2
4
6
8
10
12
0
5
10
15
20
25
30
35
40

Figure 2a

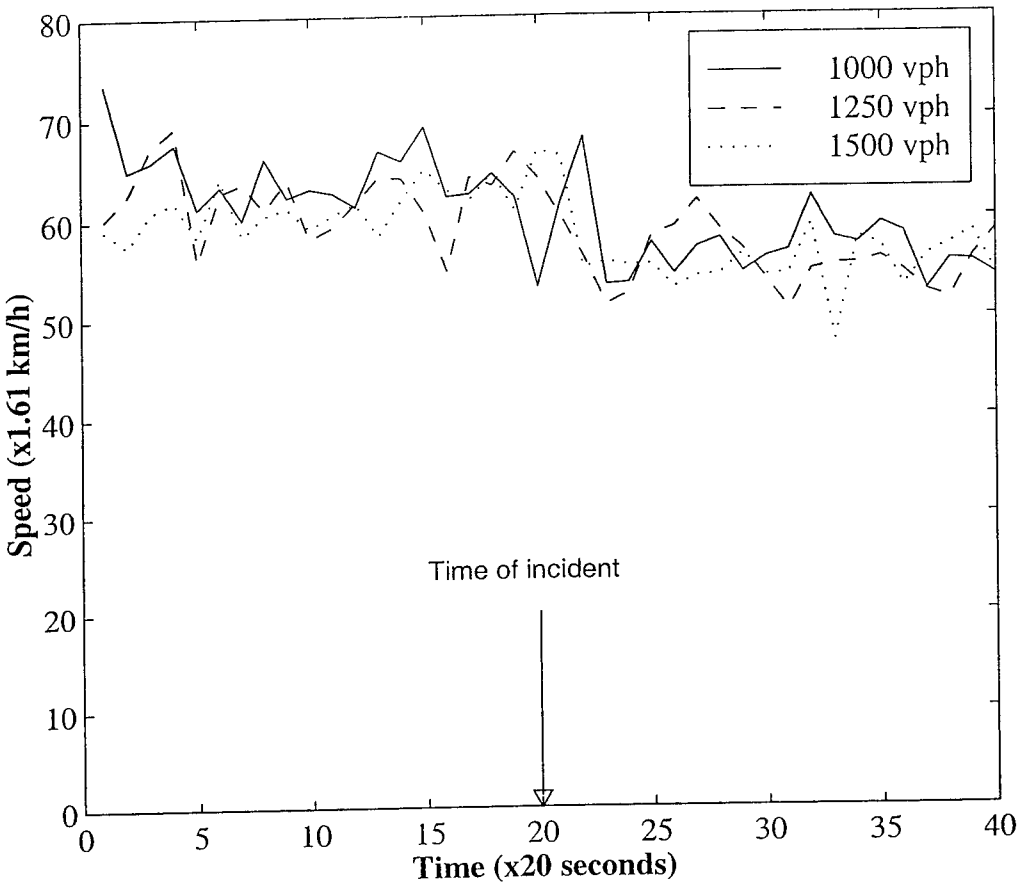

1000 vph
1250 vph
1500 vph
Time of incident
Speed (x1.61 km/h)
Time (x20 seconds)
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25
30
35
40

Figure 2b

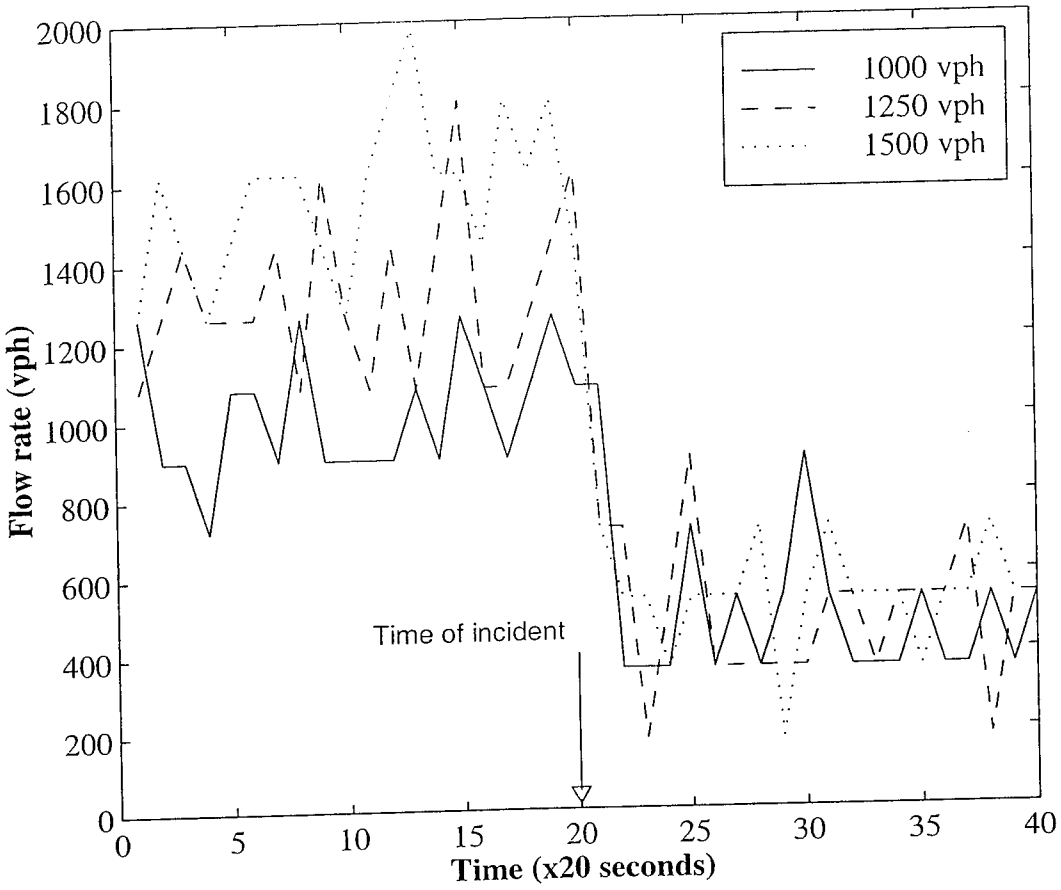

2000
1800
1600
1400
1200
1000
800
600
400
200
0
Flow rate (vph)
0
5
10
15
20
25
30
35
40
Time (x20 seconds)
1000 vph
1250 vph
1500 vph
Time of incident

Figure 2c

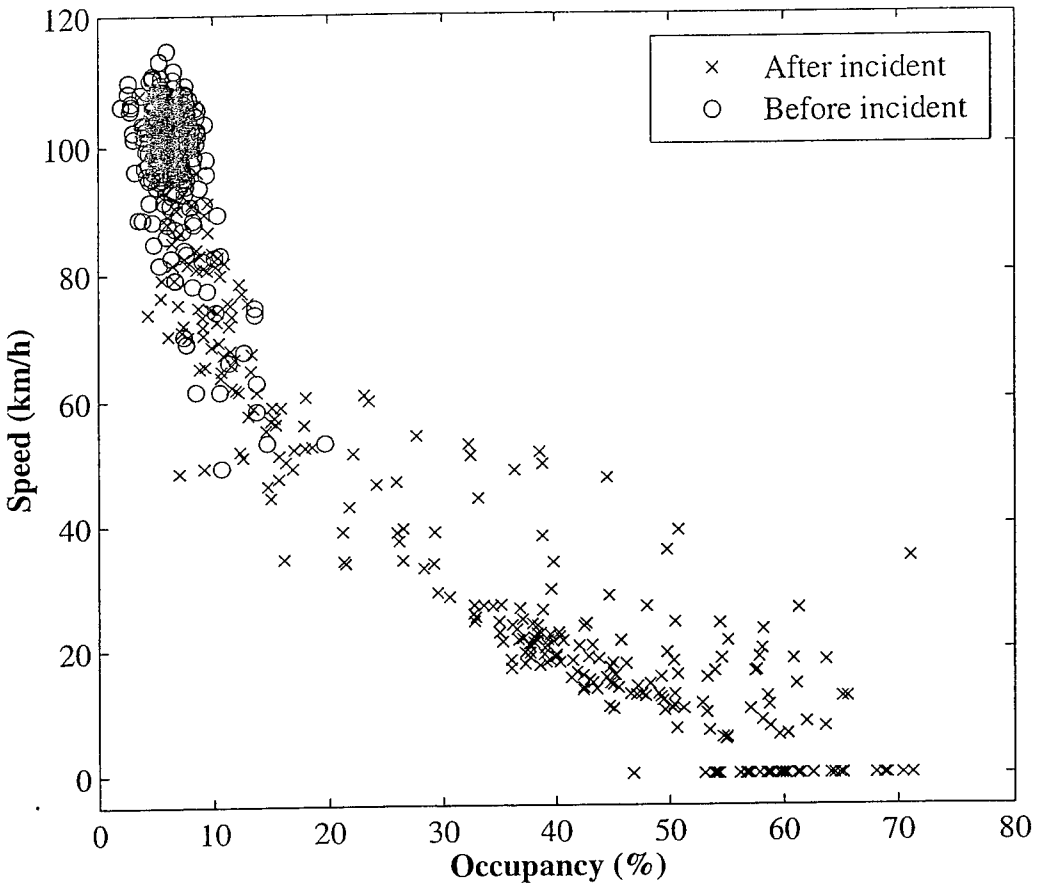

After incident
Before incident
Speed (km/h)
Occupancy (%)
0
20
40
60
80
100
120
0
10
20
30
40
50
60
70
80

Figure 3

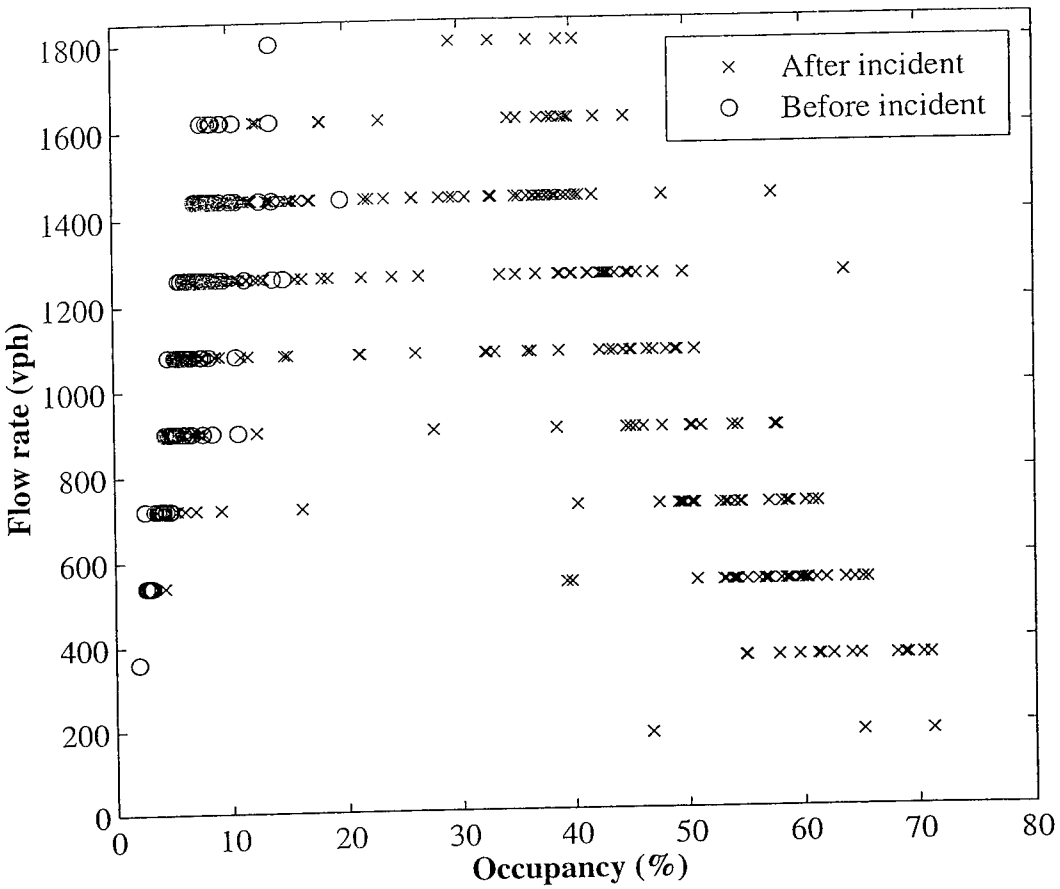

After incident
Before incident
Flow rate (vph)
Occupancy (%)
0
200
400
600
800
1000
1200
1400
1600
1800
0
10
20
30
40
50
60
70
80

Figure 4

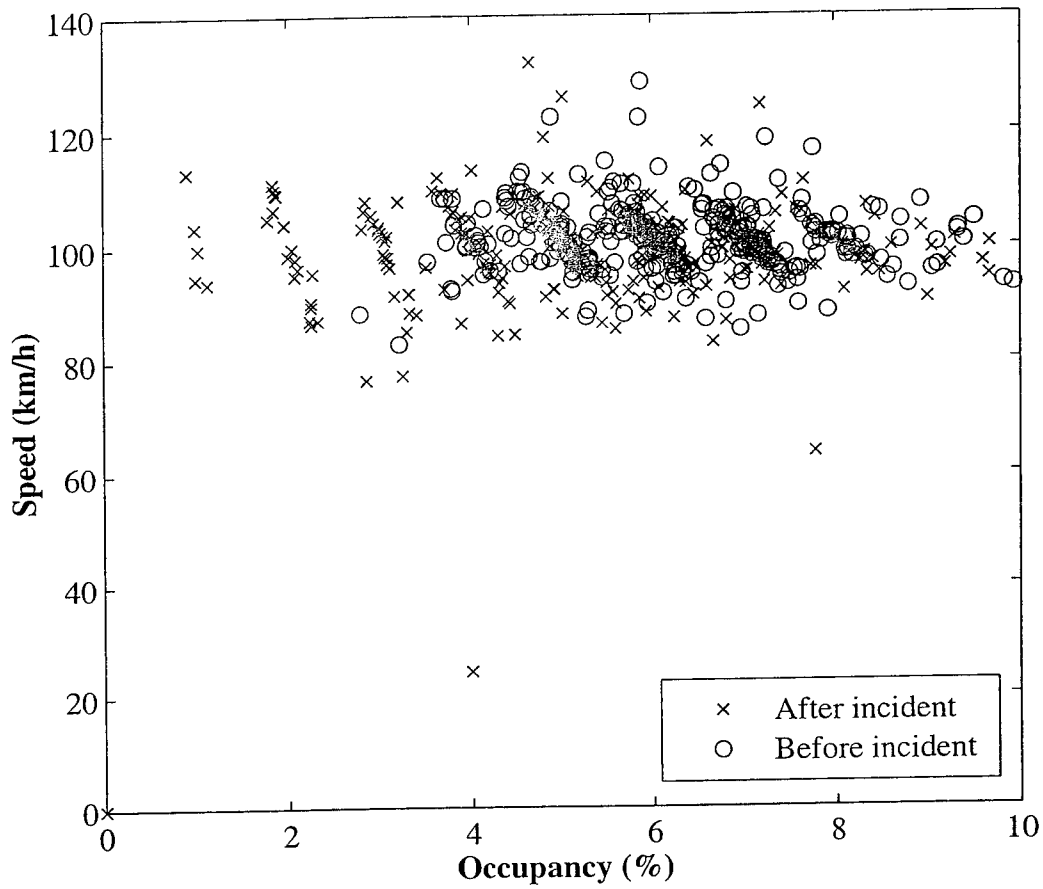

140
120
100
80
60
40
20
0
Speed (km/h)
0
2
4
6
8
10
Occupancy (%)
After incident
Before incident

Figure 5

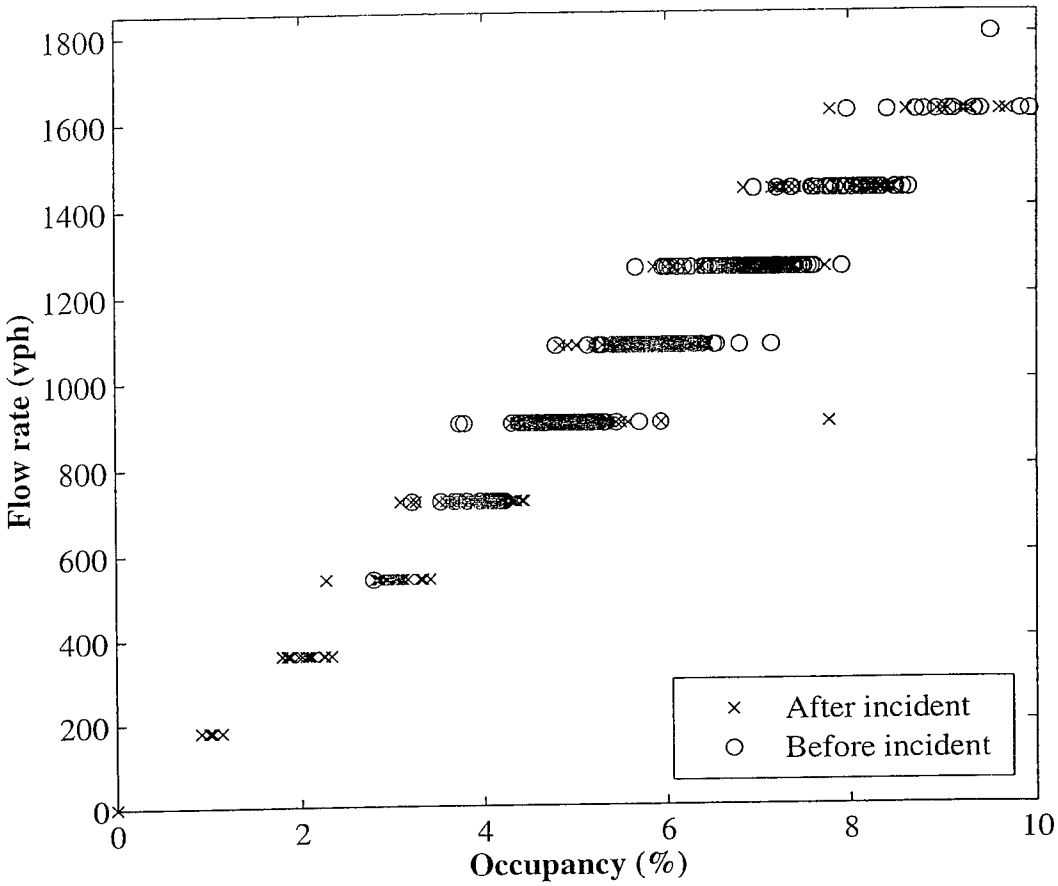

1800
1600
1400
1200
1000
800
600
400
200
0
Flow rate (vph)
0
2
4
6
8
10
Occupancy (%)
After incident
Before incident

Figure 6

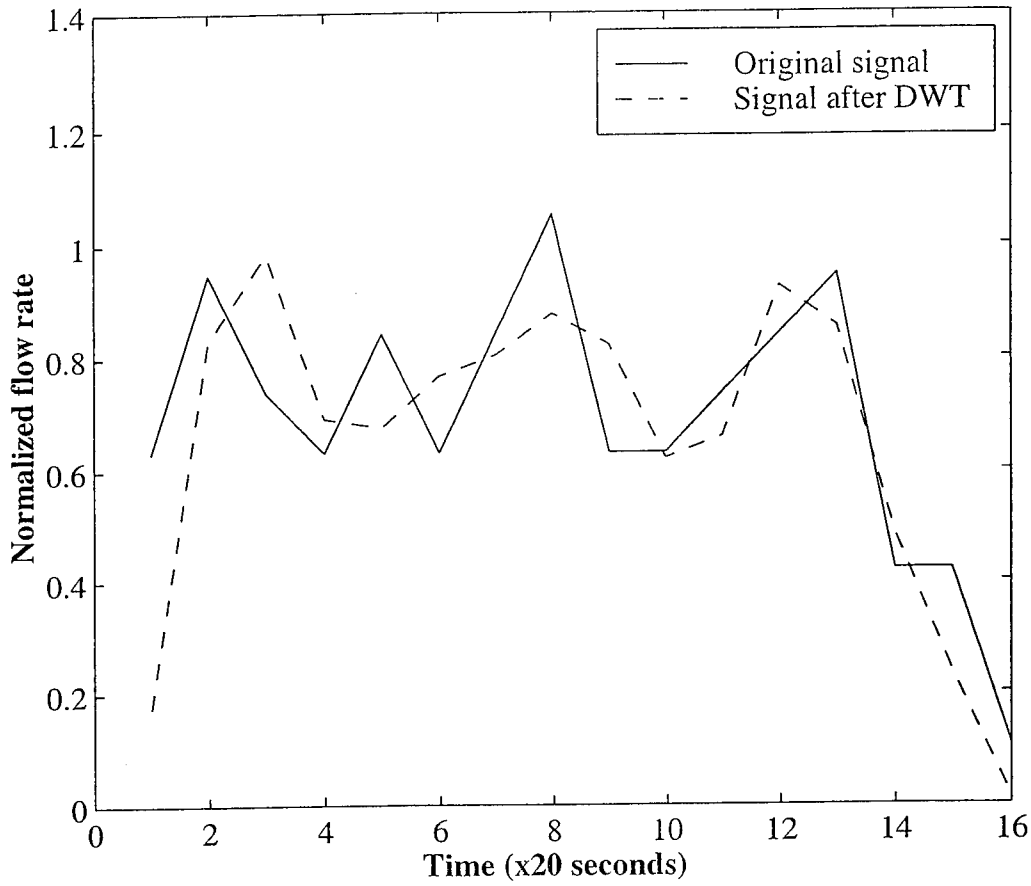

Original signal
Signal after DWT
Normalized flow rate
Time (x20 seconds)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
2
4
6
8
10
12
14
16

Figure 7a

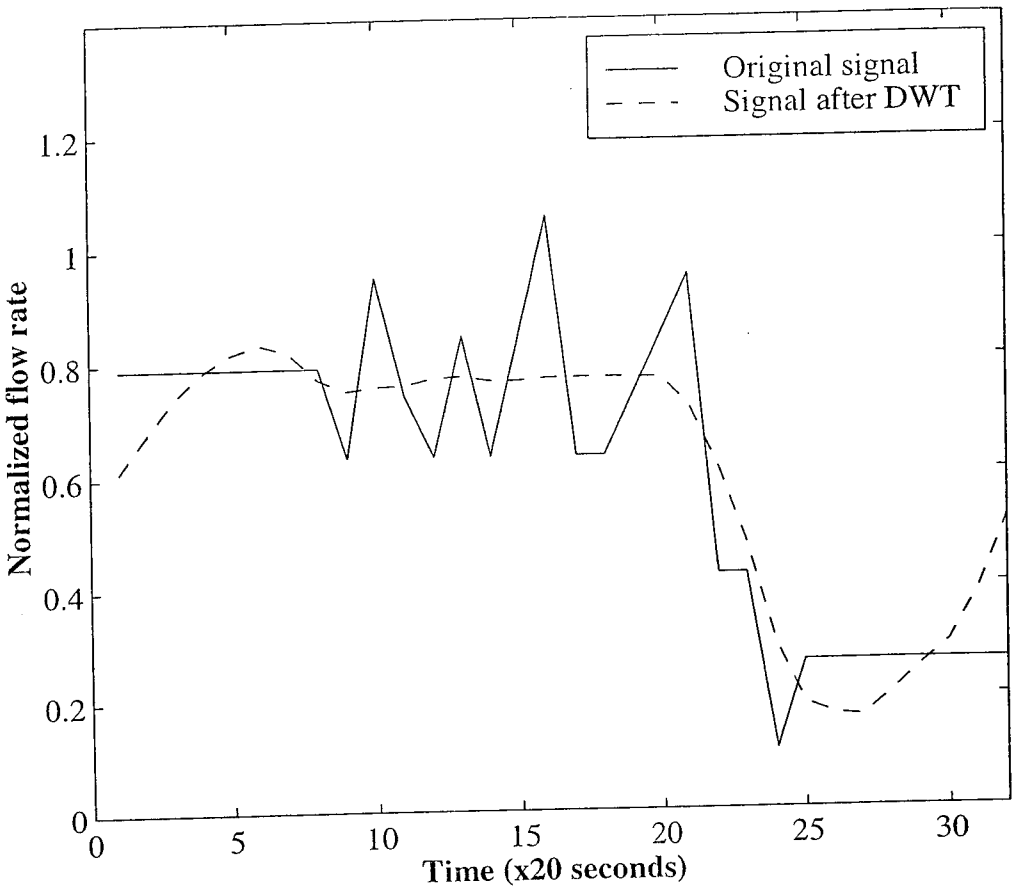

Original signal
Signal after DWT
Normalized flow rate
Time (x20 seconds)
0
0.2
0.4
0.6
0.8
1
1.2
0
5
10
15
20
25
30

Figure 7b

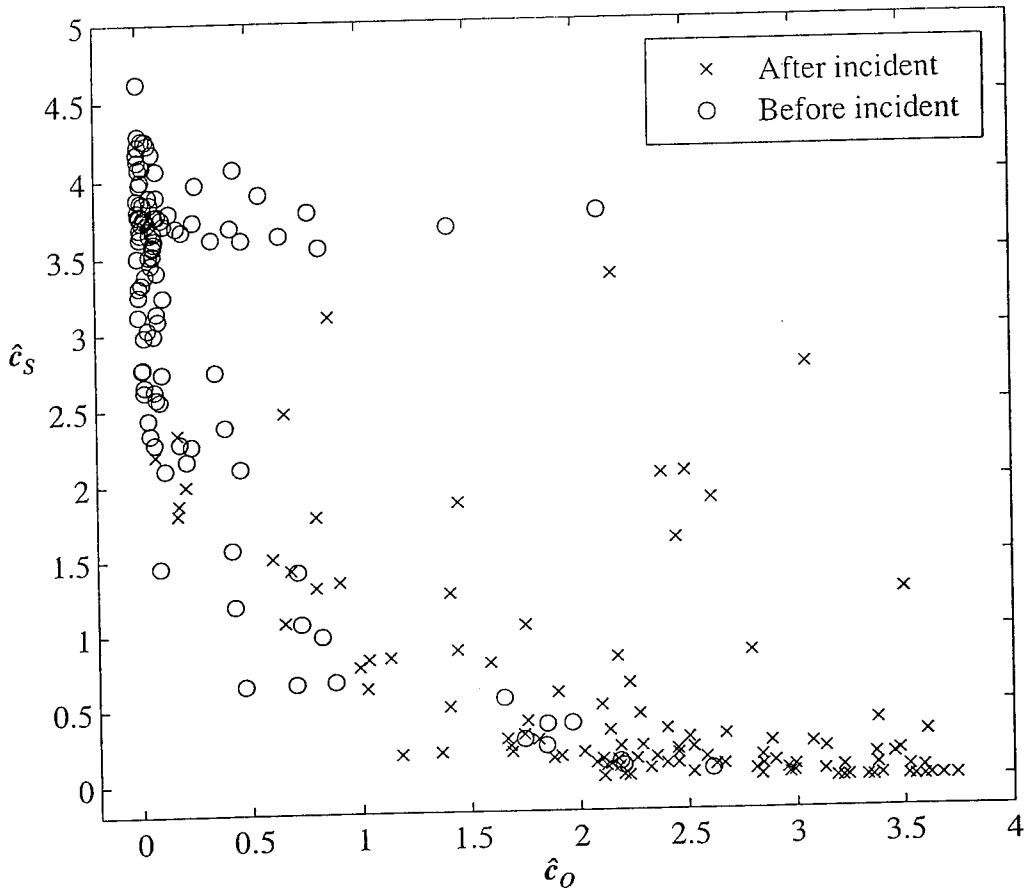

After incident
Before incident
$\hat{c}_S$
$\hat{c}_O$
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

Figure 8

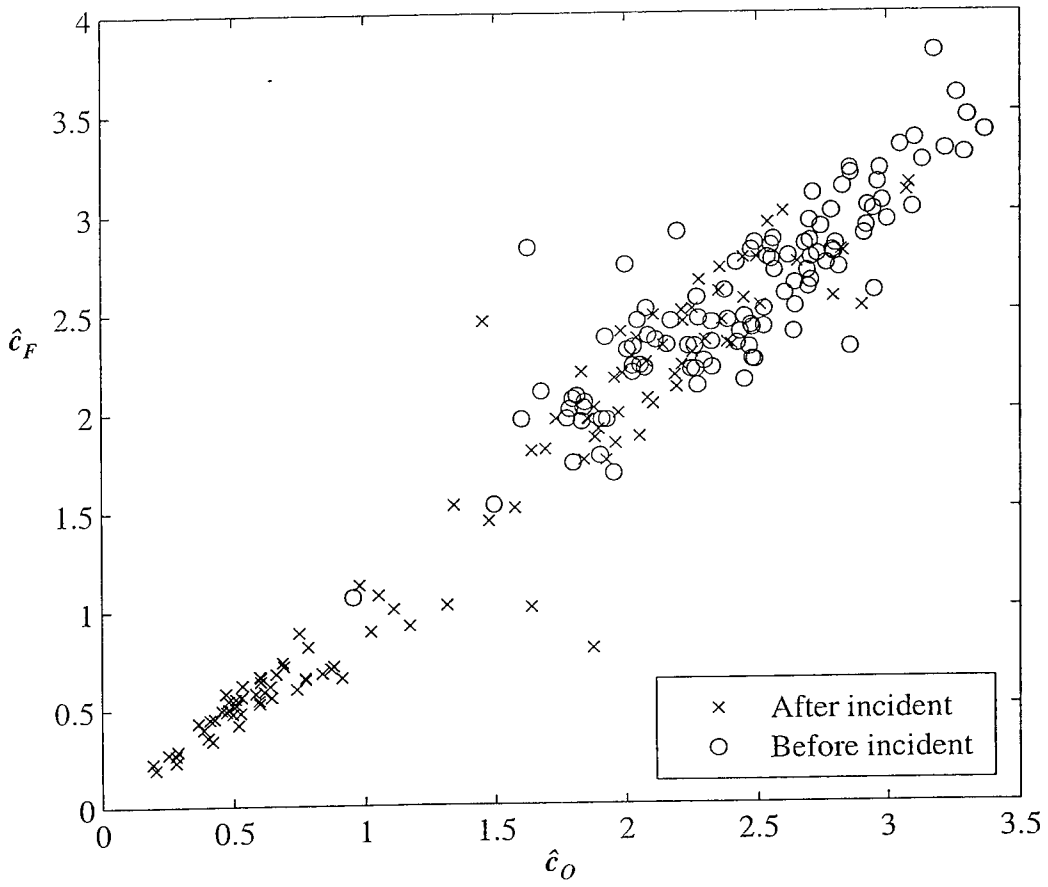

4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3
3.5
$\hat{c}_F$
$\hat{c}_O$
After incident
Before incident

Figure 9

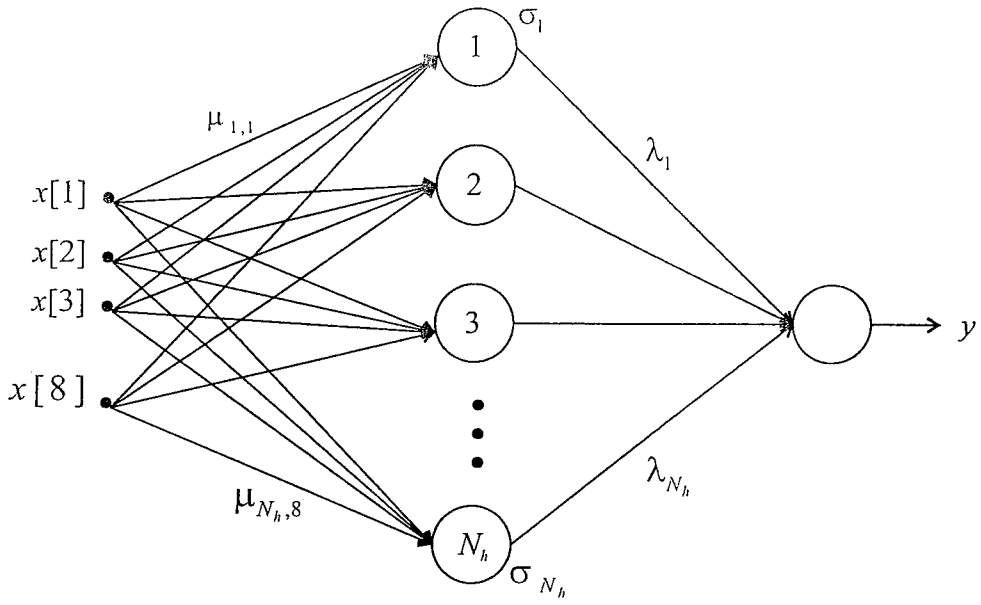

1
2
3
N_h
σ_1
σ_{N_h}
$\mu_{1,1}$
$\mu_{N_h,8}$
λ_1
λ_{N_h}
x[1]
x[2]
x[3]
x[8]
y

Figure 10

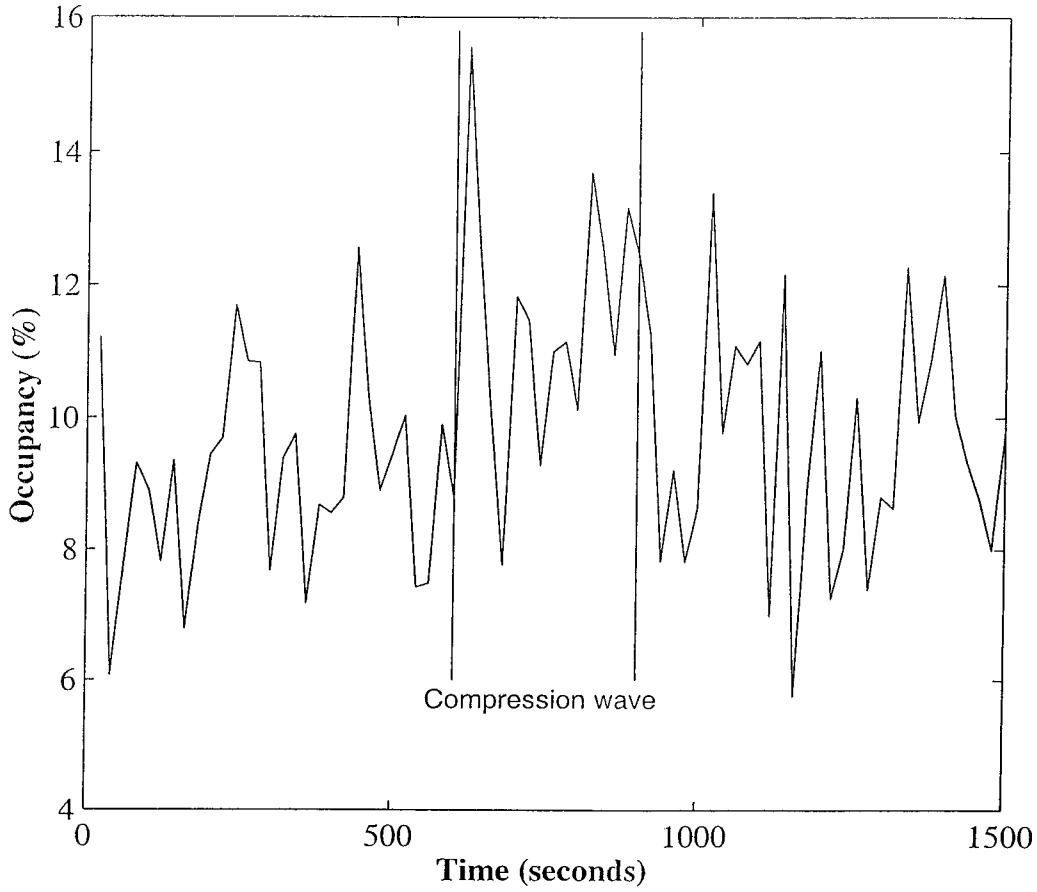

16
14
12
10
8
6
4
Occupancy (%)
Compression wave
0
500
1000
1500
Time (seconds)

Figure 11

Part 6

COMPARISON OF THE FUZZY-WAVELET RBFNN FREEWAY INCIDENT DETECTION MODEL WITH THE CALIFORNIA ALGORITHM

Asim Karim[1] and Hojjat Adeli[2]

ABSTRACT: A multi-paradigm general methodology is advanced for development of reliable, efficient, and practical freeway incident detection algorithms. The performance of the new fuzzy-wavelet radial basis function neural network (RBFNN) freeway incident detection model of Adeli and Karim is evaluated and compared with the benchmark California algorithm #8 using both real and simulated data. The evaluation is based on three quantitative measures of detection rate, false alarm rate, and detection time, and the qualitative measure of algorithm portability. The new algorithm outperformed the California algorithm consistently under various scenarios. False alarms are a major hindrance to the widespread implementation of automatic freeway incident detection algorithms. The false alarm rate ranges from 0 to 0.07 % for the new algorithm and 0.53 to 3.82% for the California algorithm. The new fuzzy-wavelet RBFNN freeway incident detection model is a single-station pattern-based algorithm that is computationally efficient and requires no re-calibration. The new model can be readily transferred without re-training and without any performance deterioration.

INTRODUCTION

In recent years, researchers have investigated neural network based incident detection algorithms with promising performance results. Adeli and Samant (2000) developed an adaptive conjugate gradient neural network pattern recognition model for freeway incident detection that employed data de-noising and enhancement. Discrete wavelet transformation and linear

[1] Graduate Research Associate. Dept. of Civil and Environmental Engineering and Geodetic Science, The Ohio State University.

discriminant analysis is used for data de-noising and enhancement, respectively (Samant and Adeli, 2000). The model is tested using simulated data for several geometric and traffic flow conditions.

Recently, Adeli and Karim (2000) created a new single-station pattern-based freeway incident detection algorithm. The characterizing pattern used is a time-series of the upstream lane occupancy and speed. Wavelet-based de-noising, fuzzy clustering, and neural network classification are used to reliably identify incident and non-incident conditions from the time-series pattern. The algorithm was tested using both simulated and real data producing excellent performance results.

In this article, a general methodology is presented for development of reliable, efficient, and practical freeway incident detection algorithms. Next, the incident detection model of Adeli and Karim (2000) is described briefly followed by a discussion of California algorithm #8. Then, the performance of Adeli and Karim's incident detection model is evaluated and compared with that of California algorithm # 8 on typical urban freeway systems. The emphasis is to evaluate the robustness of the algorithms under various traffic flow and roadway geometry conditions, as a comprehensive indicator of their practical implementation in an area-wide ATMS. Further, the new model is also tested using real incident data from the advanced regional traffic interactive management and information system (ARTIMIS) implemented in Cincinnati, Ohio (http://www.artimis.org/) and the freeway service patrol (FSP) project's I-880 database for the I-880 freeway between Oakland and San Jose, California (http://www.path.berkeley.edu/FSP/).

---

[2] Professor. Dept. of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH 43210, USA.

A NEW TRAFFIC INCIDENT DETECTION METHODOLOGY

A freeway incident detection algorithm must produce consistently reliable results from remotely sensed data of traffic streams. This is a challenging problem especially considering the non-homogenous, turbulent, and often chaotic nature of traffic flow and the limited information available from sensors. This is further complicated by noise introduced in the data during its collection and transmission. This indicates that a wholly model-based approach is less likely to be successful than a model-free, adaptive pattern recognition approach. However, a pattern-based approach must not neglect traffic behavior information that can be used to improve the efficiency and performance of the algorithm. The pattern-based approaches presented in the literature often neglect this aspect and tend to be overly simplistic. To solve the complex freeway incident detection problem effectively, our approach is based on utilizing advanced signal processing, pattern recognition, and classification techniques with appropriate heuristics derived from known traffic flow behavior.

The rationale behind this methodology is:

- Traffic flow is highly complex and not amenable to accurate mathematical modeling. Therefore, reliance must be made on adaptive algorithms that can learn and recognize patterns in an unsupervised manner.
- Traffic data is often corrupted with noise. Noise elimination is essential to improve the performance of any algorithm.
- The algorithm should require little or no calibration for its on-line implementation. That is, the algorithm's performance must be independent of roadway geometry, existence of on- and off-ramps, weather conditions, and changing traffic demand.

- Traffic flow behavior and information from other sources must not be ignored. For example, knowledge of flow behavior should be used wherever possible to simplify the algorithm and improve performance.
- The algorithm must be capable of real-time operation. Therefore, computationally intensive algorithms must be avoided.

Figure 1 presents a schematic view of the new methodology for development of advanced incident detection algorithms. Five sequential stages of processing are identified: (1) preprocessing, (2) de-noising, (3) clustering, (4) classification, and (5) decision-making. In each stage an appropriate technique has to be used to achieve the desired result. These techniques may be unique in each stage, or two or more stages may use the same technique provided that the goals of each stage are achieved. In the following paragraphs, each of these five stages is described briefly.

The preprocessing stage takes the raw traffic data obtained from sensors and transforms the data in the format needed for the algorithm. Common preprocessing approaches include calculating the cumulative values of time-series data and calculating the difference in values obtained from two sensors. The number, type, and format (i.e. the pattern) of traffic data is selected based on the behavior of traffic flow before, during, and after incidents and the performance of the algorithm.

The second stage performs de-noising and enhancement of the signal output obtained from the preprocessing stage. This is an important stage because noise corruption is one of the primary reasons for poor reliability of the incident detection algorithms. Noise is introduced both during data observation and transmission, and depends on random factors such as environmental conditions, sensor calibration errors, and traffic anomalies. The goal of this stage is to produce a clean noise-free signal. Large fluctuations in values over a short period of time due to noise

make it difficult for any algorithm to discriminate between an actual incident pattern and a noise-induced pattern. Noise can be effectively removed from a signal if it can be separated from the true signal. Transform-based techniques, such as discrete wavelet transform, provide the best solution.

The third stage performs a feature extraction process. This stage reduces the dimensionality of the data and improves the performance of the following classification and decision-making stages. Several clustering techniques are available including neural network (Adeli and Hung, 1995; Adeli and Park, 1998), fuzzy logic, and statistical approaches. In general, the statistical discriminant analysis approaches are computationally intensive and require high CPU resources in order to be implemented in real-time, a requirement for effective incident detection algorithms. Fuzzy clustering techniques such as the fuzzy c-means approach are both computationally efficient and capable of handling imprecision.

The classification stage identifies patterns in data into relevant categories. This stage determines whether the data represents an incident or not. Neural network models are most appropriate for this stage of processing. The clustering and classification stages may be combined in an algorithm.

The final decision is made in the decision making stage. This stage can be used to merge information available from other sources such as surveillance cameras before making a decision. Techniques such as fuzzy logic and decision theory may be used in this stage, in addition to heuristics based on human judgement.

FUZZY-WAVELET RBFNN MODEL FOR INCIDENT DETECTION

Recently, Adeli and Karim (2000) developed a new multi-paradigm incident detection model for freeway incident detection. The model is based on the general methodology for the

development of reliable, robust, and efficient incident detection algorithms presented above. The model is self-calibrating once it is trained and does not need to be modified for different roadway geometries and flow conditions. The new incident detection algorithm is described briefly in this section. For complete details, the reader should refer to Adeli and Karim (2000).

This model is a single-station time-series pattern recognition approach that uses advanced de-noising and classification techniques to minimize misclassification of the prevailing traffic flow conditions. Each decision pattern consists of sixteen data points of the upstream lane occupancy and speed. The two time series are normalized by dividing the values in each by the average of all values. This approach reduces the effects of varying flow rates, and thus, improves algorithm portability. The normalized time series data are then de-noised by soft-thresholding the wavelet coefficients. The de-noised data series are then clustered using the fuzzy c-means approach. The de-noised and clustered data represents the essential characteristics of the traffic flow needed to differentiate incident flow conditions from non-incident flow conditions. This pattern is then classified by a trained radial basis function neural network.

The algorithm is shown schematically in Figure 2 and summarized succinctly in the following steps. These steps represent the processing that is needed at each decision interval (equal to the reporting interval for the sensors) and at each detector station.

1. Obtain the most recent 16 data values for the lane occupancy ($x_O[n]$) and the lane speed ($x_S[n]$). When data are available every 20-s, for example, then this process is performed every 20-s by adding the new reading and dropping the last reading in the sequence.
2. For each data sequence $x[n]$ ($n = 1,\ldots,16$) perform the following computations:
   a) Normalize each sequence by dividing their values by the average of the last 16 values. The normalized sequences are denoted by x'.

b) Calculate the discrete wavelet transform (DWT) of the normalized sequence (x') using Daubechies wavelet system of length 8 (D8). The lowest scale resolved is 2. Therefore, the final number of scaling coefficients ($c_{2,k}$) obtained is 4 and the final number of wavelet coefficients ($d_{j,k}$) obtained is 12.

c) Filter the wavelet coefficients ($d_{j,k}$) using the soft-thresholding nonlinearity, $\eta(d) = \text{sgn}(d)(|d| - t)^+$, to remove noise. In this equation $(.)^+$ is equal to (.) when (.) is positive and zero otherwise and the function sgn(.) returns the sign of its argument. The threshold t is given by $t = \sqrt{2\log(N)}$ where N is the total number of data points (equal to 16 in this work). Let $\bar{d}_{j,k}$ denote the filtered wavelet coefficients.

d) Calculate the inverse DWT (denoted by IWT in Figure 2) with $c_{2,k}$ as the scaling coefficients and $\bar{d}_{j,k}$ as the wavelet coefficients to obtain the de-noised normalized sequence $\bar{x}[n]$.

3. Form the traffic pattern matrix $\mathbf{x}_i = \{\bar{x}_O[i], \bar{x}_S[i]\}$ (i = 1, 16). Use the fuzzy c-mean (FCM) algorithm to reduce the dimensionality of $\mathbf{x}$ from 16 x 2 to 4 x 2, denoted by $\mathbf{x}'$. These 8 data points represent the de-noised and clustered pattern that is used in the next classification step.
4. Feed-forward the pattern through the trained radial basis function neural network (RBFNN). If the output y is greater than a pre-selected threshold, then an incident condition is signaled. Otherwise, no incident condition exists.

The RBFNN is trained off-line from representative incident and non-incident patterns. Each pattern is processed by following Steps 1-3 above. Note that the training has to be done only once. The trained RBFNN can then be implemented on all the detector stations in the freeway management system. This portability is possible because the algorithm depends on the shape of a pattern rather than on any magnitude to distinguish between incident and non-incident

conditions. The RBFNN can even be trained using simulated data only and implemented on-line, which is the case in this evaluation.

CALIFORNIA ALGORITHM #8

The California Department of Transportation and its associates developed several algorithms for freeway incident detection in the 1970s that are collectively known as California algorithms. As many as 10 variations of these algorithms were developed. All of these algorithms use the lane occupancy values at one or two adjacent stations as input and compare them with pre-selected thresholds to characterize the state of the traffic flow. In the original California algorithm—also know as California algorithm #1—traffic flow is characterized into either incident or incident-free states based on a sequence of logic tests performed using three occupancy-based traffic patterns. Later algorithms extended this simple logic by increasing the number of logical decisions made and the number of traffic flow states reported by the algorithm.

California algorithm #8 (Payne and Tignor, 1978) incorporates incident persistence and compression wave suppression logic. The algorithm reports an incident only after the incident condition has persisted for a specified number of time periods. Further, it suppresses the signaling of an incident for 5 minutes after a compression wave is detected. California algorithm #8 uses both temporal and spatial occupancy values as input. It can classify traffic into five states: incident-free, compression wave, tentative incident, incident confirmed, and incident in progress. The compression wave state is further classified into 5 states that indicate the presence of a compression wave in the last 1, 2, 3, 4, or 5 minutes. The logic of California algorithm #8 can be described by a binary tree structure where each node, except the leaf (end) nodes, perform a two-way decision made by comparing a traffic pattern (an occupancy-based value) with a pre-selected threshold (Payne and Tignor, 1978; Levin and Krause, 1979). Starting from the root

node a sequence of such decisions are made until a leaf node is reached, which represents a traffic state. This algorithm needs six parameters for calibration. These are defined in Table 1. Five of them (P_1 to P_5) are thresholds for occupancy-based values, while parameter P_6 specifies the number of time periods the algorithm will wait for a compression wave condition to persist before signaling it.

The performance of the algorithm depends on the choice of these parameters. The parameters are determined in a trial-and-error fashion by testing the algorithm on a given data set to obtain the best trade-off between detection rate and false alarm rate. The calibrated parameters are data dependent and may not be optimal for other data sets. This in turn means that the performance of the algorithm will not be optimal at all locations and at all times in a freeway management system. Thus, California algorithms are not readily transferable and need re-calibrations for their effective network wide implementation. Despite this shortcoming the California algorithms—especially algorithms #7 and #8—are the most widely known and accepted algorithms for traffic incident detection. They are often used as benchmarks for the evaluation of new algorithms. Both algorithms #7 and #8 are recognized as the "best" (Levin and Krause, 1979). However, algorithm #8, with its additional compression wave suppression logic, performs better in heavy traffic and produces fewer false alarms as compared to algorithm #7 (Levin and Krause, 1979). For these reasons, we adopt California algorithm #8 for the comparative evaluation of the new fuzzy-wavelet RBFNN incident detection model.

EVALUATION OF THE MODEL

Introduction

In general, there are two approaches to the evaluation of a new computational model. The first approach is to test the model using a standard representative data set and determine its

performance. This data set should be recognized as the benchmark for comparative evaluations of such models. In the second approach, the model is evaluated using non-standard but representative data sets and its performance compared to that of a benchmark model on the same data set. Presently, a standard data set is not available for evaluating freeway incident detection algorithms. Furthermore, real traffic data is not available in sufficiently large and varied quantities to allow any meaningful evaluations. Therefore, freeway incident detection algorithms are usually evaluated using representative simulated data for which the performance of both the new and a benchmark algorithm (such as California algorithm #8) are compared. The use of simulated data has one more advantage not possible with real data: the algorithms can be tested and studied under different freeway traffic flow and geometric conditions.

The fuzzy-wavelet RBFNN freeway incident detection model (also abbreviated as the new algorithm/model in the rest of this article) is tested using both simulated and real data. Simulated data is used for comparative evaluations with California algorithm #8 (also abbreviated as California algorithm), whereas real data is used to test model robustness and portability.

Evaluation Criteria

Three quantitative measures are commonly used to evaluate freeway incident detection algorithms.

- Detection rate: The detection rate is defined as a percentage calculated by dividing the number of incidents correctly signaled by the algorithm to the total number of incidents in the data set. A value of 100 percent represents perfect performance.
- False alarm rate: The false alarm rate is defined as the percentage calculated by dividing the number of incidents incorrectly signaled to the total number of decisions made by the algorithm. A value of zero represents perfect performance. As the ratio is calculated with respect to the total number of decisions made by the algorithm even a small value for the

false alarm rate can represent an unacceptable number of false alarms in practice. For example, a false alarm rate of 0.5% can produce 21.6 false alarms from a single station (that reports every 20 seconds) per day. Urban freeway management systems usually have hundreds of detector stations, thus compounding the problem. Therefore, a very low false alarm rate is of utmost practical importance.

- Detection time: The detection time is defined as the time it takes the algorithm to signal the incident after its occurrence. A consistently short detection time is desirable so that emergency support can be dispatched to the scene and appropriate traffic control measures can be taken quickly. An incident detection algorithm that correctly signals 100 percent of the incidents but takes a long time to do so is of little practical value.

The quantitative measures defined above, however, do not completely describe the performance of an incident detection algorithm in practice. These performance measures are often determined from off-line tests on data for which the algorithm is calibrated. Such calibrations, however, are not practically feasible when an algorithm is implemented on-line in a large freeway management system. Thus, the network wide performance degrades significantly from that reported in the tests. For this reason, the following qualitative measure must also be considered in the evaluation of freeway incident detection algorithms.

- Portability: An algorithm is transferable if it performs at optimal or near optimal levels under different conditions without re-calibration or re-training. This qualitative measure is judged by the performance of the algorithm in terms of the three quantitative measures on different freeway traffic flow and geometric conditions. Ideally, an algorithm should not require any re-calibration for its network wide on-line implementation.

Traffic Data

The new model is tested and evaluated using both simulated and real traffic data. Simulated traffic data is generated from the microscopic stochastic simulation software package TSIS/CORSIM (http://www.fhwa-tsis.com/). More than 110 hours of traffic data is generated representing different freeway geometric and traffic flow conditions. Traffic incidents are simulated by the blockage of one lane and the fifty percent reduction in capacity of the adjacent lane(s). The incidents have a duration of 10 minutes. Coupled loop detectors or sensors are used to obtain lane occupancy, speed, and flow rate at 20-second time intervals. Detector stations are spaced from 610 to 762 m apart. In all, more than 200 separate simulations are conducted with different random number seeds resulting in more than 225,000 reports of lane occupancy, speed, and flow rate from the sensors.

Real traffic data is obtained from two sources: ARTIMIS for the Cincinnati-Northern Kentucky area freeway system, and FSP project's I-880 database for the I-880 freeway between Oakland and San Jose, California. ARTIMIS is an automated freeway management system that monitors and controls 142 km (88 miles) of freeways in the Northern Kentucky/Cincinnati, Ohio, area with 78 closed-circuit TV (CCTV) cameras, 1100 detectors, and numerous changeable message signs. Lane occupancy, speed, and flow rate data are available from the detectors every 30 seconds. Incidents are recorded by CCTV camera monitors and by proprietary incident detection logic. Very limited data were available for incident testing as the archived data period averaged over 15 minutes rather than 30 seconds.

The FSP project's database contains 30-second traffic lane occupancy, speed, and flow rate data from a 14.8 km (9.2 mile) segment of the I-880 freeway between Oakland and San Jose, California. Incidents are recorded by human observers traversing this freeway segment in patrol vehicles and noting incident location, type, and time of occurrence. The freeway has a varied

geometry with 3 to 5 lanes in each direction, one and two lane on- and off-ramps, and lane drop-offs and add-ons.

Training and Calibration

The new model is trained using simulated data. Following the procedure outlined in a previous section 60 incident and 60 incident-free patterns are used for training. These patterns are selected randomly from all the different simulations performed for this evaluation. In particular, the incident-free patterns contain samples from traffic compression waves, stop-and-go traffic, and traffic affected by on- and off-ramps. This selection is done to provide added robustness to the trained network in recognizing incident-free conditions from those caused by incidents. However, it should be noted that the model bases its decision on a pattern that is to a large extent independent of the prevailing traffic and freeway conditions. Once the network is trained and its weights established the model is evaluated without any modifications.

The California algorithm is calibrated with the same 60 incident and 60 incident-free traffic samples used for the training of the fuzzy-wavelet RBFNN model. Threshold calibration is done in a trial-and-error manner whereby the thresholds are modified after each run through the data set based on the determined detection rate, false alarm rate, and detection time. There is a trade-off between the detection rate and the false alarm rate such that an increase in the detection rate results in an increase in the false alarm rate. In the calibration process, a ceiling for the detection rate is achieved and the thresholds are then modified to minimize the false alarm rate. This procedure is identical to that reported by Payne and Tignor (1978) and Levin and Krause (1979). The set of parameters obtained are $P_1 = 13$, $P_2 = -30$, $P_3 = 30$, $P_4 = 15$, $P_5 = 30$, and $P_6 = 2$. Note that compression wave false alarm suppression is done for two time periods (40 or 60 seconds) unlike the 5 minutes used by Payne and Tignor (1978). This low value is chosen to avoid

unacceptably long detection times. This set is used throughout the evaluation without modification.

First Simulation Test - Parametric Evaluation

In this test, the new model is evaluated under different freeway geometric, traffic flow, and detector station location conditions. The general freeway layout and the locations of the detector stations and the incidents are shown in Figure 3. In this evaluation, the number of lanes is varied from 2 to 4, the flow rate is varied from 1000 to 2000 vehicles per hour (vph) per lane, and the location of the incident downstream of a detector station is varied from 152 to 610 m. Detector stations are spaced 762 m apart. An incident is modeled by the blockage of one lane and the fifty percent reduction in capacity of the adjacent lane.

The blockage of a lane produces a bottleneck in the flow of traffic. If the prevailing flow rate is greater than the reduced capacity after the incident, a queue will develop on the upstream side. At some location upstream of the incident the average speed will decrease and the occupancy will increase. This change, however, takes some time to develop and move upstream depending on the prevailing flow rate, the remaining capacity of the freeway at the bottleneck, and the distance of the incident from the upstream detector station. Even when the reduced capacity after an incident is greater than the prevailing flow rate, a change may be noticeable in the upstream speed and occupancy close to the incident location. This change in flow pattern upstream of an incident is the basis for the detection of an incident by the fuzzy-wavelet RBFNN incident detection model.

The performance of the new algorithm and California algorithm on a 2-, 3-, and 4-lane freeway is presented in Tables 2, 3, and 4, respectively. The results include the detection rate, the false alarm rate, and the detection time for each simulated situation. The fuzzy-wavelet RBFNN model is a single-station algorithm, and as described in the previous paragraph, its detection time

depends on the distance of the station from the incident, the prevailing flow rate, and the capacity reduction at the incident location. The detection times for the California algorithm also depend on the same factors. However, because the California algorithm has a two-station logic its detection time variation with distance is less pronounced. This behavior is evident from Figure 4, which shows the variation of detection times for the new and California algorithms with distance of incident from upstream station on a 4-lane freeway with prevailing flow rate of 2000 vph per lane. Notice that the detection time is longer for the California algorithm at shorter distances and shorter at longer distances as compared to the new algorithm. Nonetheless, this difference is not significant and for most practical purposes both algorithms have similar detection time performances. The detection times are long especially when the flow rate is low. When flow rate is high (2000 vph per lane) the detection time varies from 64 to 180 seconds. To shorten the time of response further, which is critical in heavy traffic, the detector stations have to be spaced closer than 762 m.

The detection time (for both the new and California algorithms) does not depend on the number of lanes in the freeway provided the flow rate remains the same. This behavior is evident from Figure 5, which shows the variation of detection times of the new algorithm with distance on a 2-, 3-, and 4-lane freeway with a prevailing flow rate of 2000 vph per lane. As observed from the figure the detection times are practically the same for all freeway lane configurations. The detection times do depend on the flow rate. Figure 6 shows the variation of detection times of the new algorithm with distance on a 4-lane freeway when flow rates are 1000, 1500, and 2000 vph per lane. At a distance of 152 m the detection time varies from 68 to 180 seconds as the flow rate increases from 1000 to 2000 vph per lane. In all these simulations the reduction in capacity is the same and thus does not impact the detection times. The effects of flow rate and

capacity reduction on detection times are inter-related. The detection times would decrease when the capacity is reduced further or when the flow rate is increased.

Both new and California algorithms detected all incidents on a 2-lane freeway (Table 2) yielding a detection rate of 100 percent. On 3- and 4-lane freeways both algorithms failed to detect some incidents for the smallest flow rate of 1000 vph per lane (Tables 3 and 4). This is because the reduced capacity after incident is still greater than the prevailing flow rate, and the impact on traffic on the upstream side is minimal. Both algorithms detected all five incidents when the incident is closest (152 m) to an upstream detector station. The new model, however, performed better on the 4-lane freeway where it also detected some incidents located at distances greater than 305 m (Table 4) yielding an overall detection rate of 83.3% as compared to 75% for the California algorithm.

The fuzzy-wavelet RBFNN model did not signal any false alarms in all the simulated conditions, thus yielding a perfect false alarm rate of zero. The California algorithm, on the other hand, signaled several false alarms especially under heavy traffic conditions. The comparison of false alarm rate on a 4-lane freeway is shown in Figure 7. The new model is thus significantly superior to the California algorithm when it comes to false alarm performance. And this is a very important consideration in the evaluation of freeway incident detection algorithms for network wide implementation.

Second Simulation Test - Freeway with On- and Off-Ramps

In this test, the false alarm rate performance of the new and California algorithms are evaluated on a freeway with on- and off-ramps. The purpose of this test is to determine the portability of the algorithms to conditions of varying flow rates and freeway bottlenecks. These conditions are known to generate false alarms because they create traffic compression waves, stop-and-go traffic, and traffic chaos near on- and off-ramps. The geometry of the freeway, the

location of the detector stations, and the on- and off-ramps are shown in Figure 8 as five contiguous segments identified by the detector station numbers noted at the bottom each segment. It consists of two on-ramps and two off-ramps. There are 3 through lanes and one auxiliary lane of length 244 m for each on- and off-ramp. Detector stations are spaced 610 or 762 m apart, 305 or 610 m upstream of the off-ramps, and 305 m upstream and downstream of the on-ramps. In the simulation model the motorists are warned in advance to the presence of an on- or an off-ramp downstream so that they can make appropriate lane change maneuvers in time.

Four traffic flow scenarios are simulated for this geometric setup as defined in Table 5. Each scenario consists of three time periods each having a different through, on-, and off-ramp flow rate. The second time period in all the scenarios has a larger through-traffic flow rate than the first time period. This simulates sudden spikes in traffic flow. In the third time periods the flow rates drop back to the values in the first time period. Scenarios 1 and 2 simulate moderate to heavy flow conditions with moderate on-ramp traffic, while scenarios 3 and 4 simulate the same with heavy on-ramp traffic.

The presence of on- and off-ramps produces non-homogeneity in traffic flow as vehicles undergo lane change maneuvers either to exit the freeway or to accommodate entering traffic. Traffic flow in the vicinity of ramps is therefore chaotic with frequent congestions and occasional stop-and-go traffic behaviors. This is especially true upstream of an on-ramp where vehicles on the freeway have to accommodate heavy traffic entering from the on-ramp. The lane occupancy and speed downstream of the on-ramp is not significantly affected. Similarly, chaotic traffic flow often occurs upstream of an off-ramp.

The false alarm rate performance of the new and California algorithms for the four simulated scenarios are presented in Table 6. The new fuzzy-wavelet RBFNN model outperformed the California algorithm #8 consistently under various scenarios (Figure 9). The false alarm rate

ranges from 0 to 0.07 % for the new algorithm and 0.53 to 3.82% for the California algorithm. It is observed that the false alarm rate of the California algorithm increases several folds when flow rate is increased. From scenario 1 to 2, the false alarm rate jumped from 0.98 to 2.34 percent, and it jumped from 0.53 to 3.82 percent from scenario 3 to 4. The false alarm rate is larger for scenarios 3 and 4 for both algorithms as compared to scenarios 1 and 2 because of the heavier on-ramp traffic in simulations 3 and 4.

The freeway segment between stations 4 and 5 generated the most false alarms. For example, in scenario 4 the California algorithm signaled 227 false alarms out of 1125 decisions, whereas the new model generated only 4 false alarms. This result highlights the poor portability characteristics of the California algorithm. As Figure 9 shows there are large differences in occupancy values between stations 4 and 5 causing the California algorithm, which has a comparative logic, to generate false alarms. The performance may be improved if the algorithm is re-calibrated using data from this particular location. However, this is not a practical solution to the problem. On the other hand, the fuzzy-wavelet RBFNN model has a single station logic where each traffic pattern is normalized before classification thus eliminating portability problems. Moreover, the new model uses a sufficiently long (5 min 20 seconds for sixteen 20-second time periods) time-series pattern that reduces the impact of sudden changes in traffic flow. As a result, the new model signaled only a few false alarms primarily at detector station 4 due the close proximity of the station to the off-ramp and chaotic traffic situation at that station.

Test Using Real Data

To further evaluate the performance of the new algorithm real traffic data from two sources are used for testing.

ARTIMIS

Sixteen traffic incident data from ARTIMIS were used to evaluate the new model. Each incident data sample consists of 30-second lane occupancy, speed, and flow rate values obtained from the upstream detector station for 10 minutes preceding the time the incident is signaled. The fuzzy-wavelet RBFNN model detected all sixteen incidents resulting in a 100 percent detection rate (Table 7). Moreover, in all cases the algorithm detected the incident before that reported by the on-line incident logic used in ARTIMIS. The exact time of occurrence of the incident is not known; therefore, the detection time cannot be determined. The ARTIMIS incident data contained data for one station (the upstream station) only. Thus, the two-station California algorithm could not be tested using those data.

FSP project's I-880 database

Both incident and incident-free data from the FSP project's I-880 database are used to evaluate the new and California algorithms. Data for 21 incidents that block one or more lanes are used. The times of occurrence of incidents and their locations are only known approximately as this information is recorded by human observers in a subjective manner. Based on this information 20 minutes of 30-second lane occupancy and speed data are extracted from the stations upstream and downstream of the incidents. Four hours of incident-free data are also extracted from the database and tested for false alarms. The performance of the new and California algorithms based on this data set is presented in Table 7. The fuzzy-wavelet RBFNN model outperformed the California algorithm in both detection rate and false alarm rate. The California algorithm signaled 3 false alarms in 480 decisions whereas the new algorithm correctly identified all of them as incident-free conditions. This test again shows the robustness and superior performance of the fuzzy-wavelet RBFNN model as compared to the California algorithm #8. Both new and California algorithms are not re-trained or re-calibrated for the real data test highlighting the superior portability characteristics of the new model.

CONCLUSION

In this article, the performance of the new fuzzy-wavelet RBFNN freeway incident detection model is evaluated and compared with the benchmark California algorithm #8 using both real and simulated data. Three quantitative and one qualitative performance measures are used for comparison. Besides the commonly used measures of detection rate, false alarm rate, and detection time, the qualitative measure of algorithm portability is also evaluated. This additional measure is of utmost practical importance because re-training and/or re-calibration is not a practically feasible solution to poor algorithm performance under varying conditions. Therefore, in all the tests performed in this evaluation no re-calibration or re-training is done, and the algorithms were compared based on the three quantitative measures.

More than 110 hours of simulated data is generated on various freeway geometries and with different flow rates for testing. Results indicate the clear superiority of the new model over the California algorithm #8. Both the new and California algorithms detected all incidents in moderate to heavy traffic. However, in light traffic (flow rate of 1000 vph per lane) on a 4-lane freeway, the new model performed better than the California algorithm, detecting incidents even when they are more than 305 m downstream of the detector station. The detection times for both algorithms are identical for practical purposes. For a freeway segment with no on- and off-ramps the new model signaled no false alarms while the California algorithm reported several false alarms especially in heavy traffic.

False alarms are a major hindrance to the widespread implementation of automatic freeway incident detection algorithms. They are not only a nuisance but also costly in the freeway management system. As a result, the false alarm rate performance of an algorithm is of utmost practical importance especially on congested urban freeways with on- and off-ramps. In such

simulated situations, it is found that the new model performed much better than the California algorithm. For example, on a 3-lane freeway segment with two on- and off-ramps and heavy flow rates (scenario 4) the new model produced a false alarm rate of 0.07% as compared to 3.82% for the California algorithm.

To further evaluate the robustness and portability of the new model, real data from ARTIMIS and FSP project's I-880 database is also used for testing. Again, the new algorithm outperformed the California algorithm in both detection rate and false alarm rate performance. The new fuzzy-wavelet RBFNN freeway incident detection model is a single-station pattern-based algorithm that is computationally efficient and requires no re-calibration. It consistently outperformed the California algorithm #8, which is considered the benchmark algorithm for freeway incident detection and the most widely used. This shows the promise of the new model to solve the decades long quest for reliable automatic freeway incident detection on urban freeways. This research shows that the new model can be readily transferred without re-training and without any performance deterioration.

ACKNOWLEDGMENT

This manuscript is based on a research project sponsored by the Ohio Department of Transportation and Federal Highway Administration.

APPENDIX I. REFERENCES

Adeli, H. and Hung, S.L. (1995), *Machine Learning - Neural Networks, Genetic Algorithms, and Fuzzy Sets*, John Wiley and Sons, New York.

Adeli, H. and Park, H.S. (1998), *Neurocomputing for Design Automation*, CRC Press, Boca Raton, Florida.

Adeli, H. and Karim, A. (2000), "Fuzzy-Wavelet RBFNN Model for Freeway Incident Detection," *Journal of Transportation Engineering*, ASCE, Vol. 126, No. 6, pp. 464-471.

Adeli, H. and Samant, A. (2000), "An Adaptive Conjugate Gradient Neural Network-Wavelet Model for Traffic Incident Detection," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 15, pp. 251-260.

Levin, M. and Krause, G. M. (1979), "Incident-Detection Algorithms: Part 1. Off-Line Evaluation," *Transportation Research Record*, No. 722, pp. 49-58.

Payne, H. J. and Tignor, S. C. (1978), "Freeway Incident-Detection Algorithms Based on Decision Trees With States," *Transportation Research Record*, No. 682, pp. 30-37.

Samant, A. and Adeli, H. (2000), "An Adaptive Conjugate Gradient Neural Network-Wavelet Model for Traffic Incident Detection," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 15, No. 4, pp. 241-250.

Table 1. Definition of parameters used in California algorithm #8

Parameter	Definition
P_1	Threshold of occupancy difference between consecutive stations
P_2	Threshold of percent occupancy change at downstream station
P_3	Threshold of percent occupancy difference between consecutive stations
P_4	Threshold of occupancy at downstream station
P_5	Another threshold of occupancy at downstream station
P_6	Number of compression wave suppression periods

Table 2. Performance of the new incident detection model and California algorithm #8 on a two-lane freeway

Flow rate (vph per lane)	Location (m)	New Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	**Detection time** (s)
1000	152	5/5	0/150	172	5/5	0/150	164
	305	5/5	0/150	300	5/5	0/150	252
	457	5/5	0/150	368	5/5	0/150	384
	610	5/5	0/150	500	5/5	0/150	480
1500	152	5/5	0/150	72	5/5	0/150	92
	305	5/5	0/150	152	5/5	0/150	132
	457	5/5	0/150	164	5/5	0/150	176
	610	5/5	0/150	240	5/5	1/150	228
2000	152	5/5	0/150	64	5/5	0/150	96
	305	5/5	0/150	88	5/5	2/150	84
	457	5/5	0/150	128	5/5	1/150	116
	610	5/5	0/150	140	5/5	0/150	132
Totals		60/60 100%	0/1800 0%		60/60 100%	4/1800 0.22%	

Table 3. Performance of the new incident detection model and California algorithm #8 on a three-lane freeway

Flow rate (vph per lane)	Location (m)	New Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	**Detection time** (s)
1000	152	5/5	0/150	156	5/5	0/150	248
	305	0/5	0/150	-	0/5	0/150	-
	457	0/5	0/150	-	0/5	0/150	-
	610	0/5	0/150	-	0/5	0/150	-
1500	152	5/5	0/150	80	5/5	0/150	96
	305	5/5	0/150	124	5/5	1/150	132
	457	5/5	0/150	244	5/5	0/150	208
	610	5/5	0/150	280	5/5	0/150	264
2000	152	5/5	0/150	64	5/5	1/150	76
	305	5/5	0/150	96	5/5	0/150	92
	457	5/5	0/150	136	5/5	0/150	136
	610	5/5	0/150	160	5/5	0/150	148
Totals		45/60 75%	0/1800 0%		45/60 75%	2/1800 0.11%	

Table 4. Performance of the new incident detection model and Californial algorithm #8 on a four-lane freeway

Flow rate (vph per lane)	Location (m)	New Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	**Detection time** (s)
1000	152	5/5	0/150	180	5/5	0/150	168
	305	2/5	0/150	390	2/5	0/150	440
	457	2/5	0/150	250	0/5	0/150	-
	610	1/5	0/150	320	0/5	0/150	-
1500	152	5/5	0/150	76	5/5	0/150	96
	305	5/5	0/150	124	5/5	1/150	132
	457	5/5	0/150	208	5/5	0/150	188
	610	5/5	0/150	272	5/5	0/150	268
2000	152	5/5	0/150	68	5/5	0/150	84
	305	5/5	0/150	84	5/5	2/150	96
	457	5/5	0/150	136	5/5	1/150	128
	610	5/5	0/150	144	5/5	1/150	140
Totals		50/60 83.3%	0/1800 0%		45/60 75%	5/1800 0.28%	

Table 5. Definition of the four simulation scenarios evaluated for the three-lane freeway with ramps

Scenario #	Time period #	Entry flow rate (vph)	On-ramp flow rate (vph)		Off-ramp flow rate (vph)	
			A	B	A	B
1	1	4500	300	500	225	450
	2	4800	300	500	240	480
	3	4500	300	300	225	450
2	1	5250	300	500	260	525
	2	5500	300	500	275	550
	3	5259	300	300	260	525
3	1	4000	600	600	200	400
	2	4500	600	600	225	450
	3	4000	600	600	200	400
4	1	5500	600	600	275	550
	2	6000	600	600	300	600
	3	5500	600	600	275	550

Table 6. False alarm performance of the new incident detection model and California algorithm #8 for the three-lane freeway with ramps

Station #	False alarms (out of 1125 decisions)							
	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	New	Cal.	New	Cal.	New	Cal.	New	Cal.
1	0		0		0		0	
2	0	0	0	0	0	0	1	0
3	0	3	0	1	0	1	0	0
4	2	0	0	1	2	0	4	5
5	0	51	0	130	1	27	0	207
6	0	1	0	0	0	2	0	3
	0.03%	0.98%	0%	2.34%	0.04%	0.53%	0.07%	3.82%

Table 7. Performance of the new incident detection model and California algorithm #8 using real traffic data

ARTIMIS	FSP Project			
Detections	Detections		False alarms	
New	New	Cal.	New	Cal.
16/16	20/21	19/21	0/480	3/480
100%	95.2%	90.5%	0%	0.63%

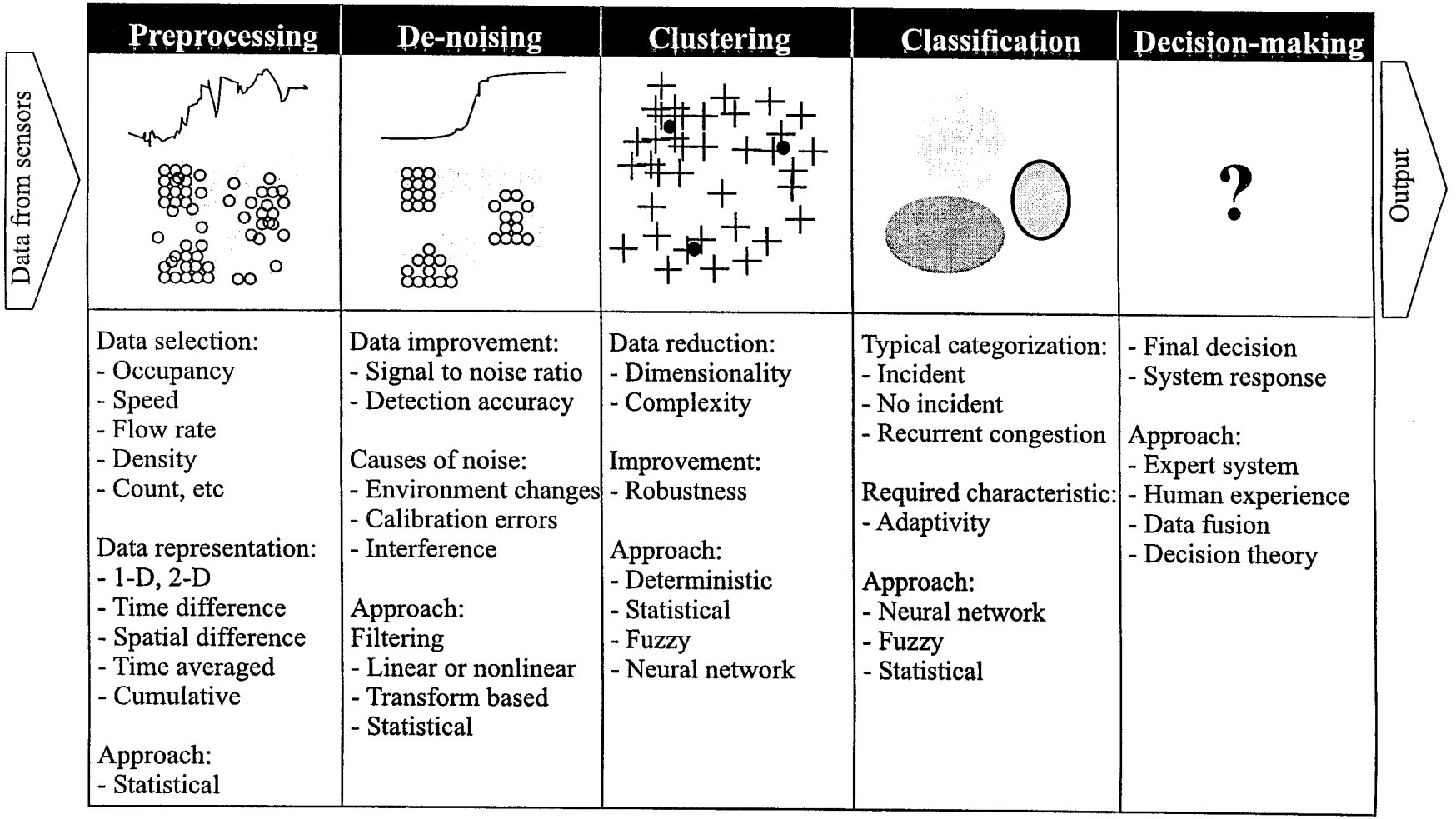

Data from sensors
Preprocessing
De-noising
Clustering
Classification
Decision-making
?
Output
Data selection:
- Occupancy
- Speed
- Flow rate
- Density
- Count, etc
Data representation:
- 1-D, 2-D
- Time difference
- Spatial difference
- Time averaged
- Cumulative
Approach:
- Statistical
Data improvement:
- Signal to noise ratio
- Detection accuracy
Causes of noise:
- Environment changes
- Calibration errors
- Interference
Approach:
Filtering
- Linear or nonlinear
- Transform based
- Statistical
Data reduction:
- Dimensionality
- Complexity
Improvement:
- Robustness
Approach:
- Deterministic
- Statistical
- Fuzzy
- Neural network
Typical categorization:
- Incident
- No incident
- Recurrent congestion
Required characteristic:
- Adaptivity
Approach:
- Neural network
- Fuzzy
- Statistical
- Final decision
- System response
Approach:
- Expert system
- Human experience
- Data fusion
- Decision theory

Figure 1

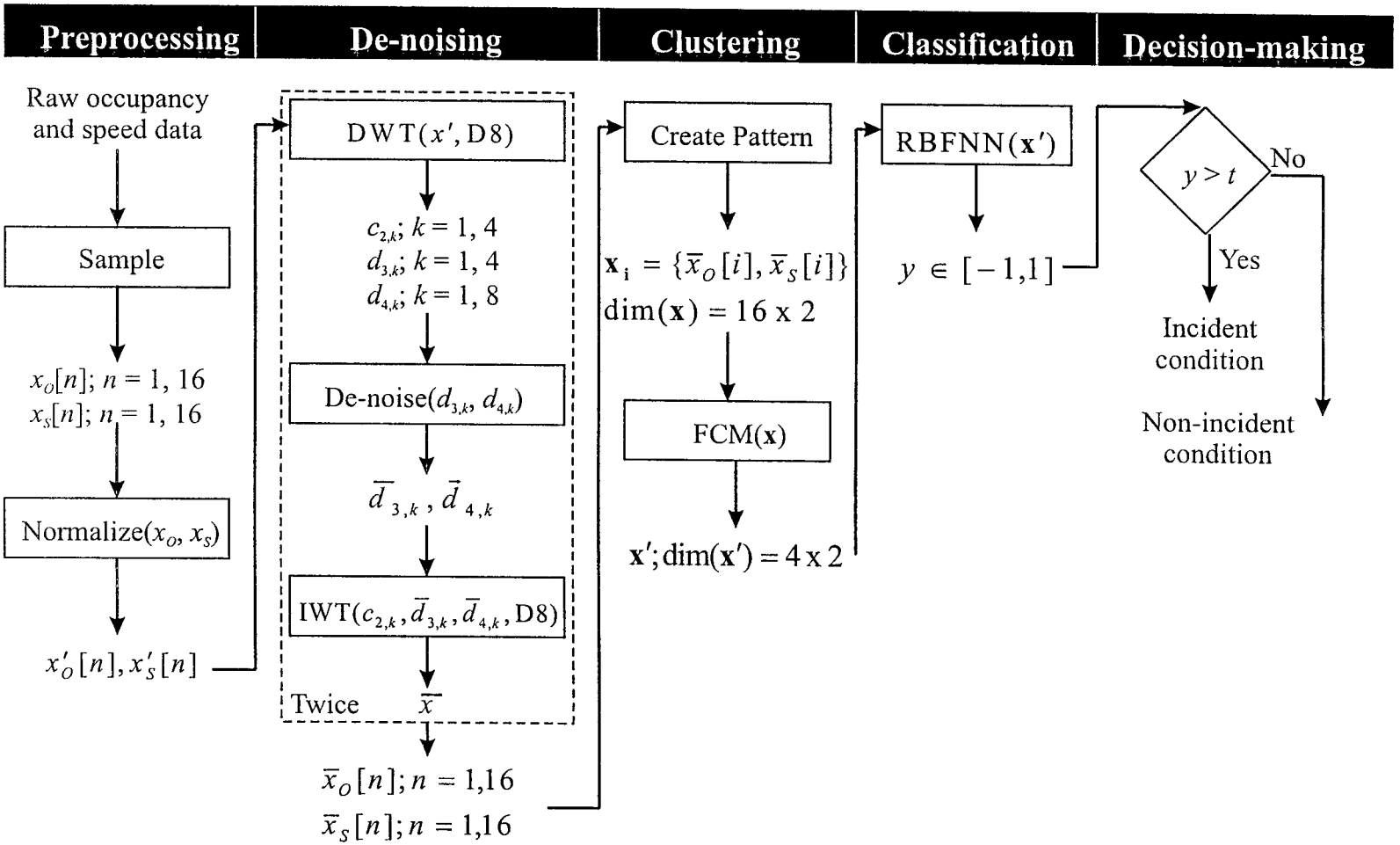

Preprocessing
De-noising
Clustering
Classification
Decision-making
Raw occupancy and speed data
Sample
$x_O[n]; n = 1, 16$
$x_S[n]; n = 1, 16$
Normalize(x_O, x_S)
$x'_O[n], x'_S[n]$
DWT(x', D8)
$c_{2,k}; k = 1, 4$
$d_{3,k}; k = 1, 4$
$d_{4,k}; k = 1, 8$
De-noise($d_{3,k}$, $d_{4,k}$)
$\bar{d}_{3,k}$, $\bar{d}_{4,k}$
IWT($c_{2,k}$, $\bar{d}_{3,k}$, $\bar{d}_{4,k}$, D8)
Twice
$\bar{x}$
$\bar{x}_O[n]; n = 1,16$
$\bar{x}_S[n]; n = 1,16$
Create Pattern
$\mathbf{x}_i = \{\bar{x}_O[i], \bar{x}_S[i]\}$
$\dim(\mathbf{x}) = 16 \times 2$
FCM($\mathbf{x}$)
$\mathbf{x}'; \dim(\mathbf{x}') = 4 \times 2$
RBFNN($\mathbf{x}'$)
$y \in [-1,1]$
$y > t$
No
Yes
Incident condition
Non-incident condition

Figure 2

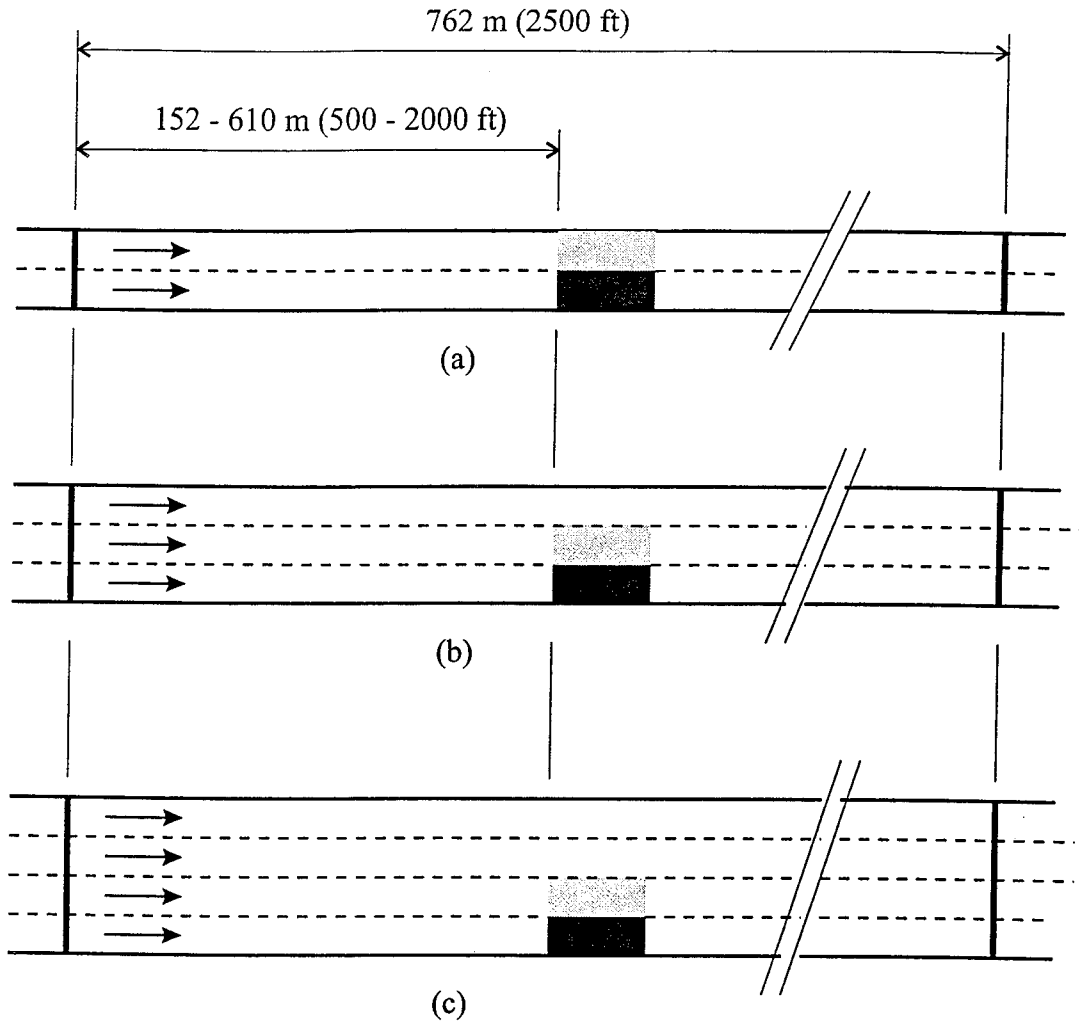

762 m (2500 ft)
152 - 610 m (500 - 2000 ft)
(a)
(b)
(c)
Legend
Detector station
Blocked lane
Lane with reduced capacity

Figure 3

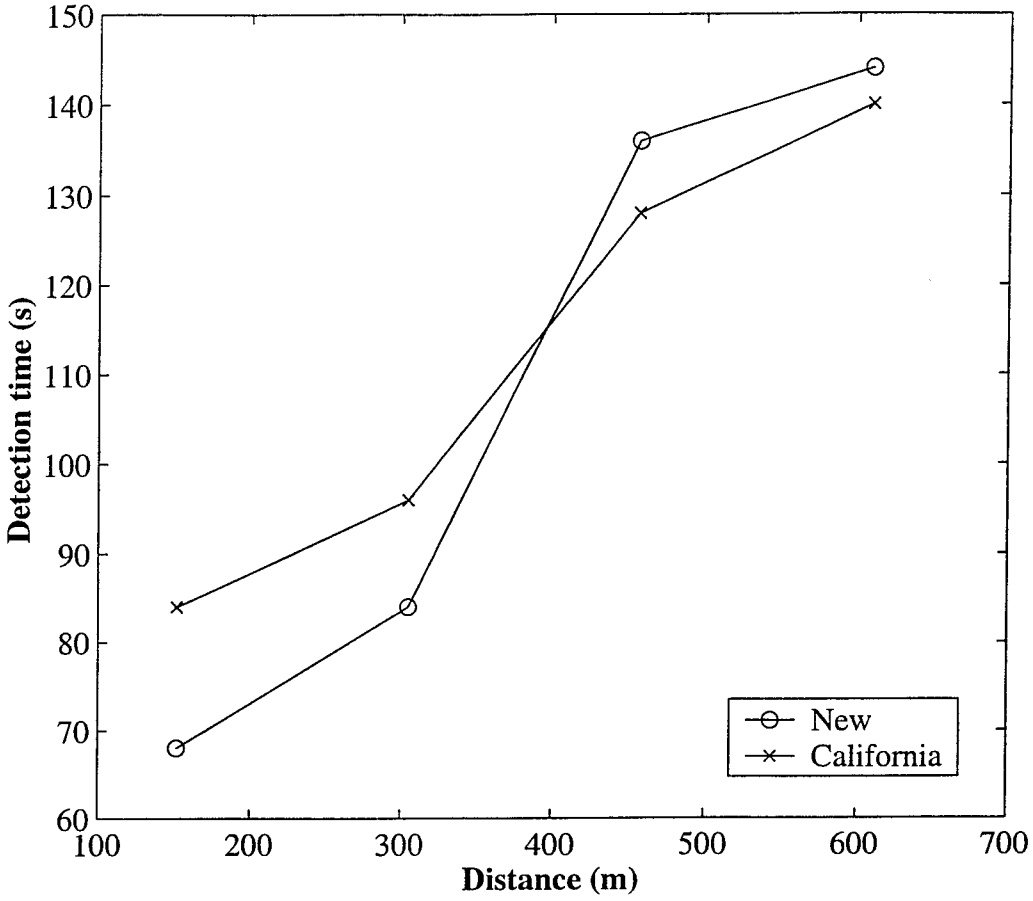

150
140
130
120
110
100
90
80
70
60
100
200
300
400
500
600
700
Detection time (s)
Distance (m)
New
California

Figure 4

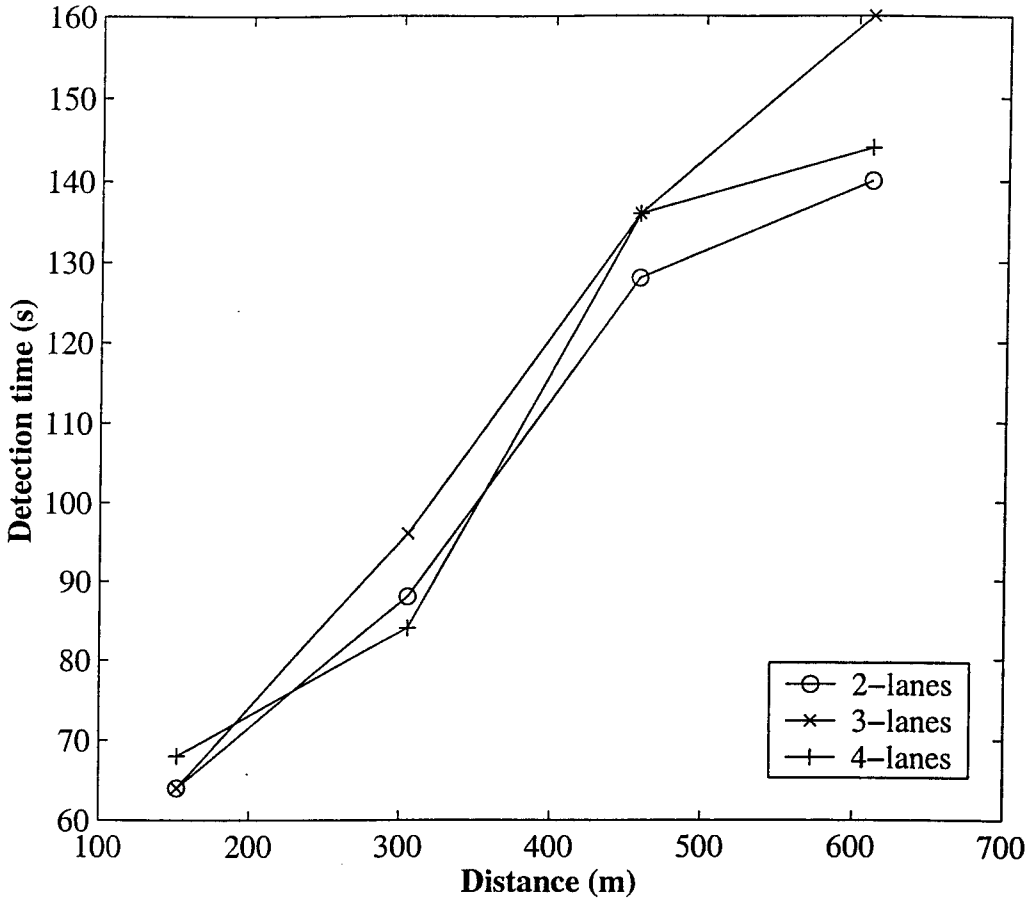

160
150
140
130
120
110
100
90
80
70
60
100
200
300
400
500
600
700
Detection time (s)
Distance (m)
2–lanes
3–lanes
4–lanes

Figure 5

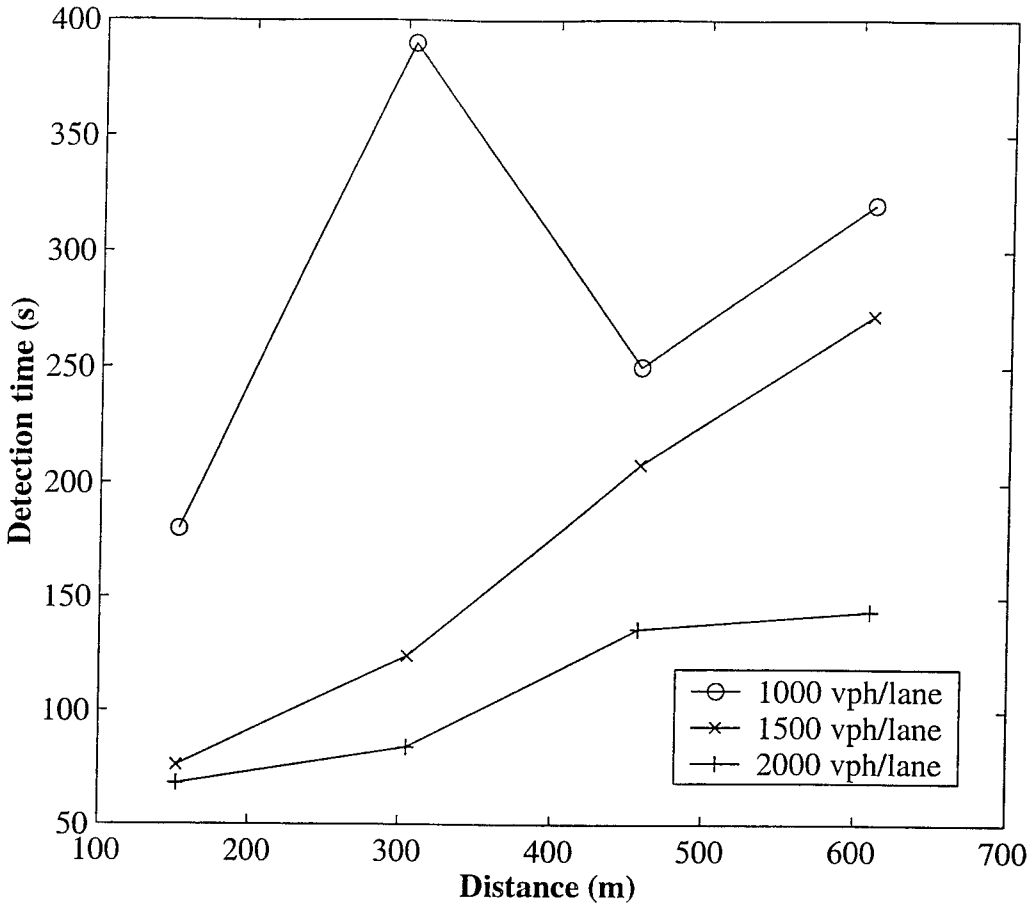

400
350
300
250
200
150
100
50
Detection time (s)
100
200
300
400
500
600
700
Distance (m)
1000 vph/lane
1500 vph/lane
2000 vph/lane

Figure 6

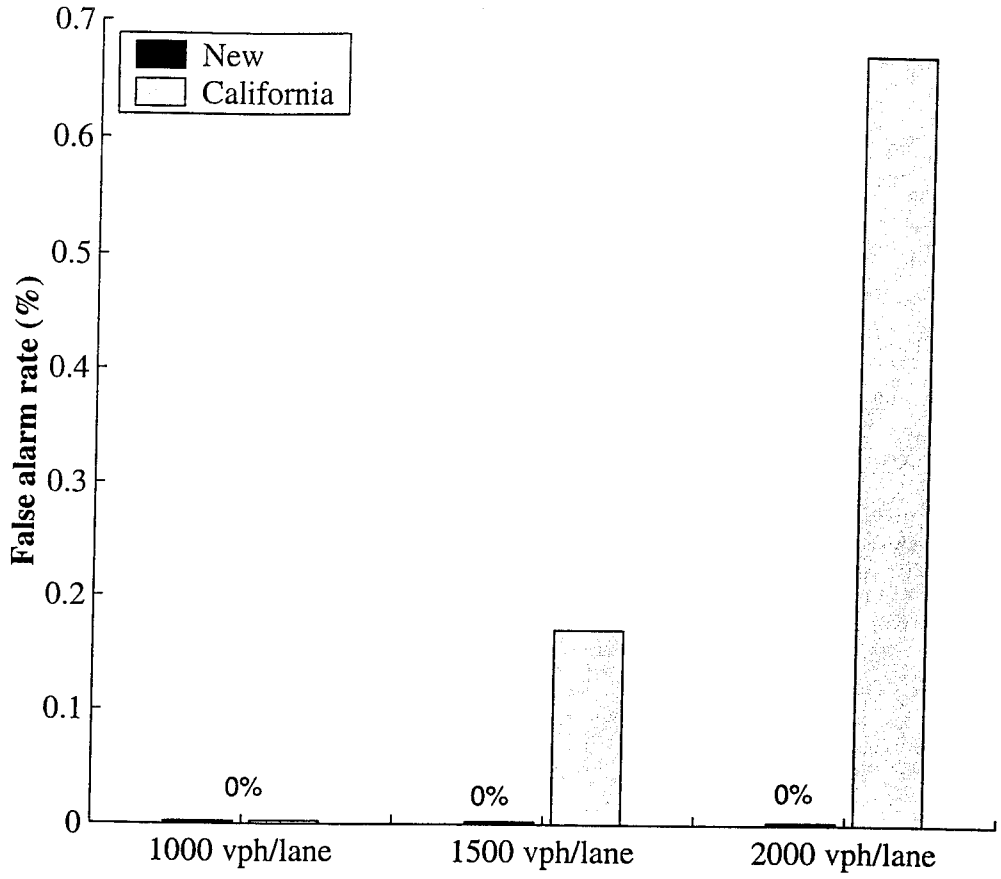

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
False alarm rate (%)
New
California
0%
0%
0%
1000 vph/lane
1500 vph/lane
2000 vph/lane

Figure 7

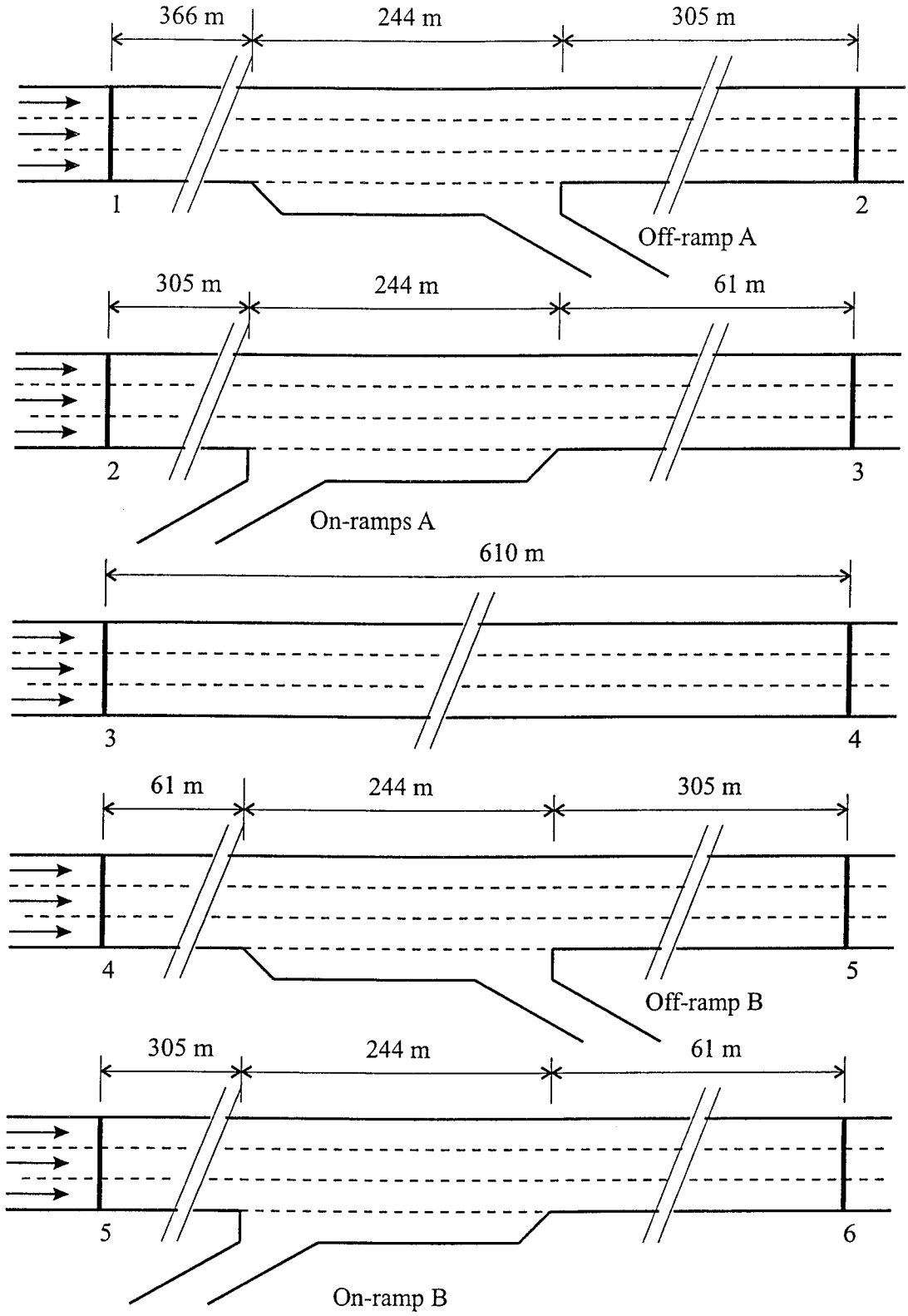

366 m
244 m
305 m
1
2
Off-ramp A
305 m
244 m
61 m
2
3
On-ramps A
610 m
3
4
61 m
244 m
305 m
4
5
Off-ramp B
305 m
244 m
61 m
5
6
On-ramp B
Legend
Detector station

Figure 8

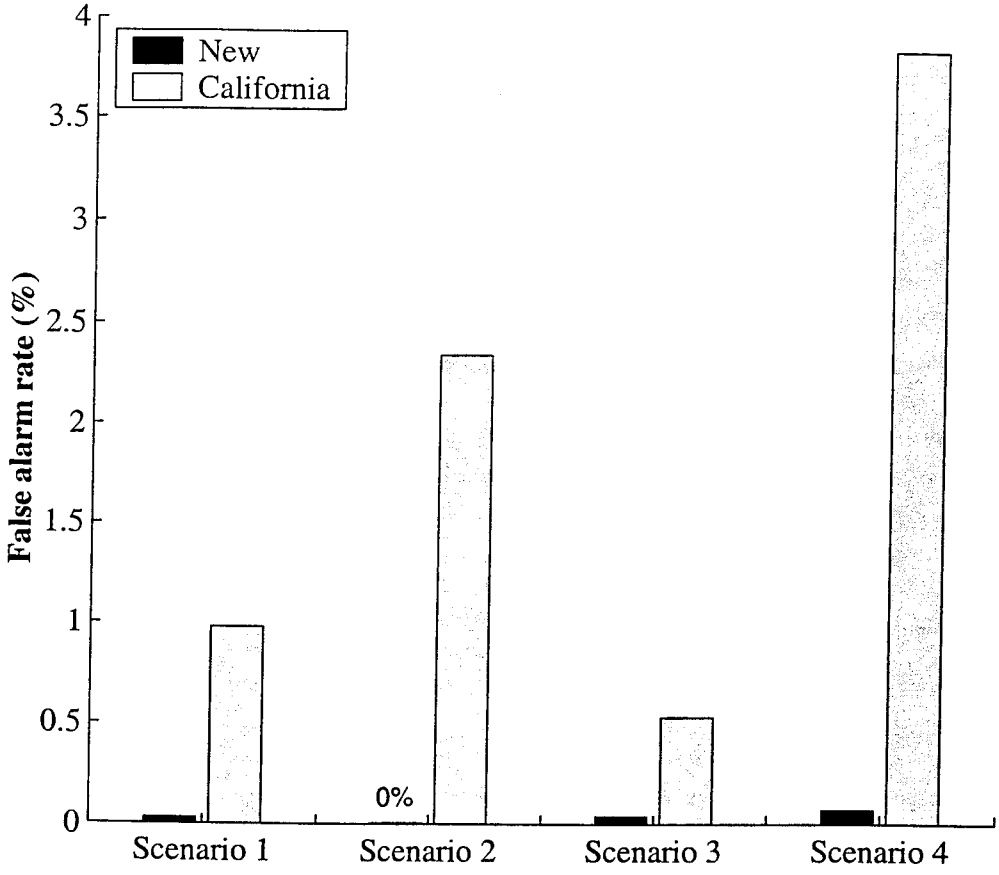

New
California
False alarm rate (%)
4
3.5
3
2.5
2
1.5
1
0.5
0
0%
Scenario 1
Scenario 2
Scenario 3
Scenario 4

Figure 9

Part 7

FAST AUTOMATIC INCIDENT DETECTION ON URBAN AND RURAL FREEWAYS USING THE WAVELET ENERGY ALGORITHM

Asim Karim[1] and Hojjat Adeli[2]

ABSTRACT: A comprehensive evaluation of the single-station wavelet energy neural network freeway incident detection algorithm of Karim and Adeli is presented. Quantitative performance measures of detection rate, false alarm rate, and detection time as well as the qualitative measure of portability are investigated for both urban and rural freeway conditions. Further, the performance of the algorithm is compared with that of the California algorithm #8. This research demonstrates the portability of the wavelet energy algorithm and its excellent performance for urban freeways across a wide range of traffic flow and roadway geometry conditions regardless of the density of the loop detectors. Rural freeways present additional challenges in that flow rates are low and detector stations are spaced further apart. Considering the difficulty in automatic detection of incidents on rural freeways, the new wavelet energy algorithm performs well on such freeways. The algorithm is fast as it detects an incident on urban freeways in less than two minutes and on rural freeways in less than three minutes.

INTRODUCTION

There are two major uses of automatic incident detection in an advanced traffic management system (ATMS). First, it is used to signal the dispatch of emergency crews to the site for prompt medical support, obstruction removal, and general maintenance of

[1] Graduate Research Associate. Dept. of Civil and Environmental Engineering and Geodetic Science, The Ohio State University.

motorists' safety. Second, it provides useful information to the routing control system to maintain and optimize system wide performance. For the best performance, the incident detection system must provide quick and reliable information. The traffic incident detection system is a main component of an ATMS (Figure 1). The other components that make up the advanced traffic management system include the traffic routing and control system, the data archiving system, and the pre- and post-processing systems. Traffic sensors provide the main source of data for analysis. Additionally, information may be obtained from the news media, special traffic probe vehicles, and motorists' call-ins. The goal of an ATMS is to maximize the system throughput. This is currently achieved by means of traffic control devices such as entry ramp access control and changeable message signs that guide and control traffic.

Recently, Adeli and Karim (2000) presented a new multi-paradigm intelligent system approach to the solution of the freeway incident detection problem employing advanced signal processing, neural network pattern recognition (Adeli and Hung, 1995; Adeli and Park, 1998), and classification techniques. This is a single-station algorithm that uses loop detector data upstream of the incident A wavelet-based de-noising technique is employed to eliminate undesirable fluctuations in observed data from traffic sensors (Samant and Adeli, 2000). Fuzzy c-mean clustering is used to extract significant information from the observed data and to reduce its dimensionality. A radial basis function neural network (RBFNN) is developed to classify the de-noised and clustered observed data. The performance of the model is evaluated and compared with the benchmark California algorithm #8 using both real and simulated data (Karim and Adeli,

---

[2] Professor. Dept. of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH 43210, USA.

2001a). The new algorithm outperformed the California algorithm consistently under various scenarios. The false alarm rate ranges from 0 to 0.07 % for the new algorithm and 0.5 to 3.8% for the California algorithm. The incident detection time ranged from 64 seconds for larger flow rates and shorter distances to the detector station to 480 seconds for lower flow rates and longer distances to the detector station.

In order to reduce the incident detection time to the range of one-to-two minutes on urban freeways, Karim and Adeli (2001b) developed a new single-station pattern recognition algorithm for freeway incident detection using data obtained from loop detectors downstream of the incident. The algorithm uses an innovative energy representation of the traffic data in the wavelet domain to de-noise and enhance desirable features before classifying them by a radial-basis function neural network. The algorithm is based on a new methodology for the development of freeway incident detection algorithms that emphasizes de-noising, feature enhancement, and the selection of a traffic pattern independent of the roadway geometry and traffic flow conditions.

The purpose of evaluating a new freeway incident detection algorithm is to determine its robustness under different traffic flow and roadway geometry conditions, and thus to assess its cost-effectiveness for practical network-wide implementation. Three quantitative performance measures are commonly used for this purpose. They are the detection rate (percentage of number of correctly detected incidents to the total number of incidents in the data set), the false alarm rate (percentage of the number of false alarms signaled by the algorithm to the total number of decisions made), and the detection time (the time it takes for the algorithm to signal the incident after its occurrence).

These three quantitative measures, however, do not provide a complete picture of an algorithm's performance in practice. The qualitative measure of portability without re-calibration must also be considered in conjunction with the quantitative measures. This is because the cost of maintaining and re-calibrating the algorithm to perform acceptably at all locations in a large freeway system can make its network-wide implementation economically infeasible. There is a cost associated with every missed detection and every false alarm, the time taken to detect an incident, and the efforts exerted to maintain and calibrate the algorithm. These costs ultimately determine the success or failure of the algorithm in practice. As reported by Abdulhai and Ritchie (1999), traffic control centers place differing cost premiums on each performance measure whenever a trade-off is sought. In any case, a higher detection rate, a lower false alarm rate, and a shorter detection time is always desirable. Moreover, an algorithm that is readily portable is often preferred over one that performs excellently only at a given location.

All freeway incident detection algorithms reported in the literature have been developed and evaluated for urban freeway systems. This is understandable because of the negative impacts incidents create on congested urban freeways and the need to remove them as soon as possible. However, there is also a need to develop and evaluate incident detection algorithms for rural freeways. The vehicle-miles of rural freeways in the United States is much larger than that for urban freeways and there is indeed a need for automatic and rapid detection of incidents so that emergency/medical support can be dispatched in time. Challenges such as low flow rates and long distance between loop detectors have hampered the development of algorithms that work effectively in rural freeway environments.

In this article, first a comprehensive parametric evaluation of the new wavelet energy freeway incident detection algorithm of Karim and Adeli (2001b) is presented using both real and simulated data. Several urban freeway scenarios are simulated for evaluation by varying the flow rate, the number of lanes, and the distance of the incident from detector station. The effects of on- and off-ramps are also considered. Next, the algorithm is evaluated on rural freeway scenarios where flow rates are low and detector stations are spaced far apart. For comparison, the performance of the California algorithm #8 is also presented.

In the following section, factors to consider in rural freeway incident detection are delineated. Then, the wavelet energy freeway incident detection algorithm is described step-by-step, followed by a comprehensive evaluation of the algorithm and discussions of the test results.

FACTORS TO CONSIDER IN RURAL FREEWAY INCIDENT DETECTION

Traffic on urban freeways is characterized by high demand and periodic congestion that reduces the level of service expected by motorists. Because of the high demand and insufficient capacity the level of service degrades dramatically when an obstructing incident occurs. Therefore, quick and reliable identification and localization of such incidents is essential to prevent unacceptable backups and delays caused by obstructions that are not cleared quickly. As such, an effective incident detection algorithm must be both reliable and fast in detecting an incident.

Traffic on rural freeways, on the other hand, is usually congestion-free under normal operating conditions. Furthermore, the impact of an obstructing incident is often less

severe because traffic demands on rural freeways usually do not exceed the capacity. Nevertheless, the need for reliable automatic incident detection still exists. Incidents in rural areas, unlike in urban areas, may go unreported for several minutes. Furthermore, the transit of emergency and medical support to rural locations can take more time. Therefore, rapid automatic notification of an incident condition is very valuable. Automatic incident detection on rural freeways is challenging because of low flow rates and large distances between detectors. Most of the incident detection algorithms developed so far have not been evaluated under such conditions, and, in general, perform poorly under low flow rate conditions.

Several factors have to be considered in the development and evaluation of an automatic rural freeway incident detection algorithm. These considerations are in general more stringent and demanding than those required for reliable detection on urban freeways.

- Density of detectors: It is practically infeasible to have closely spaced loop detectors on rural freeway segments. Thus, the algorithm must work reliably under situations where detectors are spaced 2-3 km apart. The cost-effectiveness of the solution improves dramatically with an increase in the distance between detectors at which the algorithm can produce reliable results.
- Detection time: The detection time on rural freeways is important not for traffic management purposes but for emergency medical support reasons. Often a serious congestion may not develop as a result of a rural incident. However, rapid identification and localization of the incident is still necessary to ensure that emergency support can arrive on the scene at the earliest possible time. There is a

tradeoff between the detection time and the distance between detectors. In general, the closer the spacing between detectors, the shorter the detection time; however, reducing the spacing between detectors significantly increases the number of detectors that have to be installed and maintained on long stretches of rural freeways.

- Low prevailing flow rates: Traffic incident detection algorithms normally depend on the change in traffic pattern that results from an incident to identify its occurrence. However, when the prevailing flow rate is low and the incident does not reduce freeway capacity significantly the change in traffic pattern can be minor. This poses a serious challenge in the design of reliable algorithms.
- Calibration and maintenance: Because of the huge mileage of rural freeways calibration and maintenance of algorithms at all locations can become extremely costly. Therefore, algorithms for rural freeway incident detection should require minimal maintenance for acceptable operational performance. Custom calibration of the algorithm at each location is practically infeasible.

An algorithm that is cost-effective for implementation on an urban freeway system may be impractical for implementation on rural freeways. In general, a lower performance should be expected for an algorithm on rural freeways than on urban freeways because of the constraints on detector spacing and flow rates. The goal is to have an algorithm that requires no re-calibration with acceptable performance. Note that these considerations apply to passive techniques for incident detection only where traffic data obtained from loop detectors embedded in the pavement are analyzed to identify characterizing patterns. Active techniques, such as in-vehicle transponders, may be more

effective in rural settings but require more investment and are often perceived as intrusive by the public.

WAVELET ENERGY MODEL FOR FREEWAY INCIDENT DETECTION

The new single-station incident detection algorithm developed by Karim and Adeli (2001b) takes as inputs a time-series of lane occupancy and lane speed at the upstream detector station or a time-series of lane occupancy and lane flow rate at the downstream detector station. Each time series consists of 16 data values averaged over and obtained at every 20- or 30-second interval. The patterns at both upstream and downstream detector stations are transformed and represented in the wavelet domain as an energy functional. This representation makes it possible to de-noise, enhance, and reduce the dimensionality of the patterns effectively and efficiently. The processed patterns are then classified into one of two states representing either an incident or incident-free condition by a radial basis function neural network. The key ideas are described in Karim and Adeli (2001b) in general terms. A complete detailed step-by-step algorithm is presented in this section.

Only the downstream station logic is implemented and tested in this evaluation. It was found that the upstream logic produced results almost identical—and in the case of detection time, slightly inferior—to those produced by the downstream logic. Therefore, the wavelet energy algorithm consists of the collection, processing, and classification of the downstream lane occupancy and flow rate time-series data. In a freeway management system, this algorithm is implemented at every detector station and reports on the presence or absence of an incident upstream of the station. The algorithm is shown schematically in Figure 2 and described in the following steps.

1. Obtain the last 16 lane occupancy and lane flow rate readings and form the sequences $f_O[i]$ and $f_F[i]$, respectively, where $i = 1,\ldots,16$. When readings are available every 20-s, for example, this process is performed every 20 seconds by adding the new reading and dropping the last reading in the sequence.
2. For each data sequence $f[i]$ perform the following computations:
   a) Sort the elements in the sequence $f[i]$ to create a new sequence $g[i]$ such that $g[i] \geq g[i+1];\ i = 1,\ldots,15$
   b) Normalize $f[i]$ by dividing all its elements by the average of the two largest values:

$$\bar{f}[i] = \frac{f[i]}{0.5(g[1]+g[2])} \qquad i = 1,\ldots,16 \tag{1}$$

   c) Extend the normalized sequence $\bar{f}[i]$ by 8 elements on each side, as follows:

$$\hat{f}[i] = \begin{cases} 0.5\left(\bar{f}[1]+\bar{f}[2]\right) & 1 \leq i \leq 8 \\ \bar{f}[i-8] & 9 \leq i \leq 24 \\ 0.5\left(\bar{f}[15]+\bar{f}[16]\right) & 25 \leq i \leq 32 \end{cases} \tag{2}$$

   The sequence $\hat{f}[i]$ now has 32 elements.

   d) Perform a two-stage low-pass filter of the sequence $\hat{f}[i]$, as follows:

$$c_4[k] = \sum_i h_0[i-2k]\hat{f}[i] \tag{3}$$

$$c_3[k] = \sum_i h_0[i-2k]c_4[i] \tag{4}$$

   where $h_0[i]$ is the 8-coefficient low-pass filter for the Daubechies wavelet system of length 8 (Daubechies, 1992). The sequence $c_3[i]$ ($i = 1,\ldots,8$), called the scaling

coefficients, represents a lower scale or resolution (scale 3) of the original 32-element sequence $\hat{f}[i]$ (scale 5).

e) Enhance the sequence $c[i]$

$$c[i-2] = |c_3[i]|^2 \qquad i = 3,4,5,6 \tag{5}$$

The sequence $c[i]$ has 4 elements representing the squared scaling coefficients (a measure of energy in the wavelet domain) for the middle 16 elements of $\hat{f}[i]$. These elements correspond to the input traffic data before it is extended for processing. Let the processed lane occupancy and speed data be denoted as $c_O[i]$ and $c_F[i]$, respectively.

3. Form the feature pattern by concatenating the processed lane occupancy and flow rate sequences:

$$x[i] = c_O[i],\ x[i+4] = c_F[i] \qquad i = 1,\ldots,4 \tag{6}$$

The 8-element sequence $x[i]$ represents the de-noised, clustered, and enhanced pattern that is used in the subsequent step for classification

4. Feed-forward the feature pattern $x[i]$ through a trained radial-basis function neural network. The neural network has 8 input nodes, 12 hidden nodes with Gaussian transfer functions, and one output node with a linear transfer function. If the output is greater than a pre-selected threshold (a small positive value such as 0.2) then an incident is signaled; otherwise, the pattern represents an incident-free condition.

The RBFNN is trained with incident and incident-free patterns to determine the weights of the links connecting the input layer to the hidden layer and the links connecting the hidden layer to the output node. Training is done iteratively to minimize the output error.

Once the network is trained no further training is necessary. For further details, refer to Karim and Adeli (2001b).

EVALUATION AND PARAMETRIC INVESTIGATION

Goals

A comprehensive evaluation of the wavelet energy freeway incident detection algorithm is presented in this section. The goals of the evaluation are:

1. To determine the quantitative performance measures (detection rate, false alarm rate, and detection time) for typical urban freeway conditions;
2. To determine the quantitative performance measures for typical rural freeway conditions;
3. To assess the transferability or portability of the algorithm, that is, to compare the algorithm's performance under different roadway geometry and traffic flow conditions without re-calibration;
4. To perform a parametric evaluation of the algorithm, that is, to determine the sensitivity of the algorithm to variations in roadway geometry and traffic flow conditions.
5. To compare the performance of the algorithm with that of California algorithm #8 (Payne and Tignor, 1978).

The roadway geometry conditions evaluated are the number of lanes (2, 3 and 4), the distance of the incident from detector station (152 to 2744 m), and proximity to on- and off-ramps. Traffic flow is varied from 500 to 2000 vehicles per hour (vph) per lane. An incident is modeled as the blockage of one lane and the 50 or 40 percent reduction in

capacity of the adjacent lane(s). The time of blockage is varied from 3 minutes to 10 minutes.

Data

The majority of the traffic data used in the evaluation are generated using the simulation software TSIS (http://www.fhwa-tsis.com/). TSIS is a microscopic simulation tool that considers each vehicle as a separate entity in a stochastic model of vehicles and their environment (roadway geometry, pavement conditions, proximity to other vehicles, etc).

In addition to simulated data, real data from the San Francisco Bay area freeway service patrol project's I-880 database is also used for evaluation. This database is a collection of binary files of loop detector outputs collected over a period of about 2 months. A software program is used to process this database and extract selected information in a readable format for further processing. The database contains basic information such as lane occupancy, flow rate, and speed. The information on the location and time of incidents is recorded by human observers and has to be correlated to the loop data for analysis. Because this information is recorded by humans, it is not reliable and has to be verified by visual observation of the loop detector data. In all, data for 21 single-lane blocking incidents and four hours of incident-free conditions are extracted for evaluation in this research.

Training and Calibration

The wavelet energy freeway incident detection algorithm is trained with 60 incident and 60 incident-free patterns. These patterns are chosen randomly from all the simulated data generated for the evaluation. No real data is used in the training phase of the

network. The training determines the weights for the RBFNN. Once the algorithm is trained no further training is done as it is evaluated using different sets of data.

The California algorithm #8 (Payne and Tignor, 1978) is a well-known two-station comparative algorithm for freeway incident detection that uses lane occupancy data as input. The algorithm logic consists of a sequence of decisions where occupancy-based input values are compared with pre-selected thresholds to characterize traffic flow into one of five major states. California algorithm #8 is one of several variations that were developed in the 1970s. It incorporates an incident persistence test and a compression wave suppression test to reduce the generation of false alarms. Six parameters or thresholds have to be calibrated for the algorithm. Employing the same 60 incident and 60 incident-free patterns used for the wavelet energy algorithm, calibration of the California algorithm is done in a trial-and-error fashion until the misclassification error is minimized. The threshold values used in this evaluation are as follows (these values produced the best overall calibration results for the data used):

Threshold of occupancy difference between consecutive stations = 13%,

Threshold of percent occupancy change at downstream station over the time interval =30,

Threshold of percent occupancy difference between consecutive stations = 30,

Threshold of occupancy at downstream station = 15%,

Second threshold of occupancy at downstream station = 30%, and

Number of compression wave suppression periods = 2.

The same set of parameters is used throughout the evaluation without re-calibration. This is done to test the portability property of the algorithm and compare it with that of the new wavelet energy algorithm.

Parametric Evaluation Using Simulated Data on Typical Urban Freeways

Figure 3 shows the freeway layouts simulated for the parametric evaluation. These layouts represent typical urban freeway segments with 2, 3, and 4 lanes with detectors spaced 762 m apart. The location of the incident, which consists of the blockage of one lane and the 50 percent reduction in capacity of the adjacent lane, is varied from 152 to 610 m from the downstream (or upstream) detector station. The flow rates considered are 1000, 1500, and 2000 vph per lane. The data set used for this evaluation is identical to that used for the parametric evaluation of the earlier fuzzy-wavelet RBFNN model (Karim and Adeli, 2001a).

The performance of the new wavelet energy algorithm is compared with that of the California algorithm #8 on 2, 3 and 4 lane freeways in Tables 1, 2, and 3, respectively. The wavelet energy algorithm performs perfectly in all scenarios in terms of producing an overall detection rate of 100 percent and a false alarm rate of zero. The California algorithm, on the other hand, failed to detect 25 percent of the incidents on 3- and 4-lane freeways. This result demonstrates the excellent performance of the new wavelet energy algorithm in difficult-to-detect situations such as the closure of just one lane on a multiple lane freeway when prevailing flow rate is low. In general, whenever the prevailing flow rate is less than the reduced capacity after the incident, incident detection algorithms like California algorithm #8 are less likely to detect an incident because a significant queue does not develop in a short period of time (say, a few minutes). This characteristic also exists in other incident detection algorithms that utilize only the upstream occupancy to detect the presence of an incident condition.

The detection times reported by the new wavelet energy algorithm varies from 56 to 116 seconds. The detection time generally increases with an increase in the distance of the incident from the downstream detector station. However, this variation of the detection time with location of incident is substantially less pronounced than that for the California algorithm. This is evident from Figure 4, which compares the detection times for the wavelet energy and California algorithms on a 2-lane freeway. The detection time for California algorithm is a lot longer, varying from 76 to 480 seconds; it increases substantially with a decrease in flow rate and distance of incident from downstream detector station. This is because the California algorithm is based on the formation of congestion on the upstream side of the incident, which takes more time to develop when the prevailing flow rate is low. The wavelet energy algorithm, on the other hand, does not exhibit this behavior as seen in Figure 4. The performance of the wavelet energy algorithm is also not greatly effected by changes in geometry such as the number of lanes as noted in Figure 5. The relative independence of the wavelet energy algorithm to changes in flow rate and roadway geometry demonstrates its superior portability property as compared to the California algorithm.

False alarms generated by automatic freeway incident detection algorithms are often a major source of excessive operational costs. Traffic control centers would often prefer an algorithm that generates fewer false alarms over another one with better detection rate but higher false alarm rate. On urban freeway segments, the wavelet energy algorithm generated no false alarms, thus producing an overall false alarm rate of zero. In contrast, the California algorithm produced false alarm rates of 0.22, 0.11, and 0.28 percent, on 2-,

3-, and 4-lane freeways, respectively. These false alarms are generated during moderate and heavy traffic flow conditions.

False Alarm Performance in the Vicinity of On- and Off-Ramps

Traffic flow in the vicinity of on- and off-ramps is often chaotic and marked by large fluctuations in occupancy, speed, and flow rate as vehicles maneuver to enter and exit the freeway. This is especially true for urban freeways where ramps are usually spaced closely apart and the entering and exiting flow rates are high. On- and off-ramps are thus geometric bottlenecks that create non-homogeneities in traffic flow, and are responsible for generating a large number of false alarms from existing automatic freeway incident detection algorithms. To test the false alarm performance of the algorithms in such situations a 3-lane urban freeway segment with two on- and off-ramps is modeled for simulation (Figure 6). For this freeway geometry four traffic flow scenarios are evaluated, as described in Table 4. Each scenario consists of three time periods of different mainline, on-, and off-ramp traffic flow rates. This is done to simulate sudden changes in entering and exiting flows on heavy traffic freeways that often cause automatic freeway incident detection algorithms to produce false alarms.

The false alarm performance of the wavelet energy algorithm and California algorithm #8 in the vicinity of on- and off-ramps is given in Table 5. The remarkable false alarm performance of the wavelet energy algorithm is evident; it produced no false alarms at all six detector station locations and in 27000 (4X6X1125) decisions. The California algorithm, on the other hand, produced numerous false alarms, ranging from 0.5% to 3.8%, especially for the roadway segment between detectors 4 and 5 (Figure 6).

Note that both algorithms are not re-calibrated or retrained for this and all other evaluations. This is done to ascertain the portability property of the algorithms. The California algorithm #8 may be re-calibrated for each segment to produce fewer false alarms. However, this procedure is time consuming and expensive on a large urban freeway management system. Furthermore, this procedure may be required on a regular basis to ensure optimal performance with changing traffic flow conditions. The wavelet energy algorithm, on the other hand, performed excellently without any need for retraining and thus is readily transferable and portable for implementation on urban freeway systems.

Evaluation on Rural Freeways

Rural freeways present a challenge for passive automatic freeway incident detection algorithms that use loop detector data. As discussed earlier, it is economically infeasible to have closely spaced loop detectors on the large network of rural freeways in the U.S. Thus, incident detection algorithms can only rely on sparse information to arrive at a decision. This is further complicated by the often low flow rates on rural freeways that are impacted little by an incident. As a result, passive automatic incident detection algorithms often perform poorly on rural freeways making them impractical for traffic agencies to implement. Traffic agencies also desire algorithms that require little maintenance and no site-specific calibrations for their optimal performance on rural freeways.

To the best of the authors' knowledge, no automatic freeway incident detection algorithm has been evaluated for rural freeway conditions. In this section, the new wavelet energy algorithm and California algorithm #8 are evaluated on a simulated 2-lane

rural freeway segment with loop detectors spaced 3048 m (10,000 ft) apart. The performance of the algorithms is determined for flow rates of 500, 1000, 1500, and 2000 vph per lane. The distance of the incident from the downstream detector station is varied from 152 to 2744 m. A lane-blocking incident is modeled as the closure of one lane and the 40 percent reduction in capacity of the adjacent lane. A shoulder incident is modeled by the 40 percent reduction in capacity of both lanes. Incidents of 5- and 10-minute durations are evaluated.

The performance of the wavelet energy algorithm and California algorithm #8 on a 2-lane rural freeway with a lane-blocking incident of 10 minutes duration is given in Table 6. Results are categorized by prevailing flow rates (500, 1000, 1500, and 2000 vph/lane) and distance of the incident from the downstream detector station (152-2744 m). The wavelet energy algorithm performed much better overall than the California algorithm #8. When the prevailing flow rate is a low 500 vph per lane, the wavelet energy algorithm detected 18 percent of the incidents as compared to zero for the California algorithm. At this low flow rate, there is little or no impact of the incident on traffic patterns upstream and downstream of the incident. A change in the upstream traffic pattern is usually non-existent because any shock wave created dissipates within 50 to 100 m of the incident. On the downstream side, the shock wave travels much faster and is less likely to be masked by oncoming traffic flow. However, because of the natural variation inherent in traffic flow and the fact that the change in pattern is small, this pattern often cannot be distinguished from normal traffic flow patterns.

This is evident from Figure 7, which shows a typical lane occupancy time-series plot at the downstream detector station. An incident occurs at time 900 seconds and persists

for 600 seconds; however, no visible change in the occupancy *pattern* such as a persistent reduction in the occupancy during and after the incident is noticeable from the plot (the spike in the figure is an outlier due to an extraneous factor such as noise in the data and is not an indicator of any change in the occupancy pattern). The wavelet energy algorithm is able to detect some incidents because it considers both occupancy and flow rate readings to create an enhanced and de-noised pattern before classifying it. The increased sensitivity of the algorithm, however, does come with a higher false alarm rate. The number of false alarms can be reduced by increasing the threshold t (see Figure 2) used in the wavelet energy algorithm. This can be done easily and in real-time by an appropriate logic in the algorithm.

A flow rate of 1000 vph per lane is typical on many rural freeways under normal operational conditions. Under these conditions the wavelet energy algorithm detected 88 percent of the incidents with a false alarm rate of 0.08 percent. The California algorithm, on the other hand, produced detection and false alarm rates of 20 percent and zero, respectively. The California algorithm failed to detect any incident that is less than 2479 m from the downstream station. The wavelet energy algorithm is able to detect 85% of incidents for such distances from the downstream station. The California algorithm will require the detector stations to be spaced at about 610 m apart for its performance to be at par with the wavelet energy algorithm. Such a high density of loop detectors is economically infeasible for rural freeways. Furthermore, the wavelet energy algorithm required an average time of 151 seconds to detect the incidents, which is acceptable for rural incident management applications. These results show the superiority of the wavelet energy algorithm on rural freeways.

At flow rates of 1500 and 2000 vph per lane the wavelet energy algorithm detected all incidents producing a detection rate of 100 percent, while the California algorithm produced a detection rate of 72 and 100 percent, respectively. The California algorithm again failed to detect incidents at distances of less than 600 m from the downstream detector station at the lower flow rate of 1500 vph per lane highlighting its unsuitability for implementation on rural freeways. It also had a false alarm rate of 0.56% at the higher flow rate of 2000 vph per lane compared with 0% for the wavelet energy algorithm. The detection times for the wavelet energy and California algorithms varied from 44 to 160 and 148 to 500 seconds, respectively. Except when flow rate is 500 vph per lane the detection time for the wavelet energy algorithm on rural freeway is less than three minutes.

Often an incident results in the blockage of a lane for only a short duration of time. For example, a disabled vehicle may block one lane for a few minutes before it is moved onto the shoulders. Detecting such incidents are often more challenging for incident detection algorithms as the impact of the incident lasts just for a shorter period of time. In all the previous evaluations, the incident duration is equal to 10 minutes. Table 7 shows the performance of the wavelet energy algorithm and California algorithm #8 on a 2-lane rural freeway when the lane blockage lasts for 5 minutes only. The detection rate, false alarm rate, and detection times produced by the two algorithms for this scenario are similar to those produced for 10-minute incidents recorded in Table 6. This is because the maximum detection time for the energy wavelet algorithm in all cases is 160 seconds which is substantially less than the 5-minute duration of the incident. As long as the duration of an incident is greater that the detection time it does not affect the performance

of the algorithm in any significant way. The same does not hold true for the California algorithm because its detection time is as large as 430 seconds. Consequently, as is the case for the 10-minute duration incidents, the performance of the wavelet energy algorithm is superior to that of California algorithm #8.

Sometimes incidents produce no lane blockage but only reduction in the capacity of the lanes. This situation may occur when, for example, a disabled truck is parked on a shoulder reducing the capacity of the lanes. To study such scenarios on rural freeways a 40 percent reduction in capacity of both lanes that lasts for 10-minutes is modeled for evaluation. The performance of the wavelet energy and California algorithms under such scenario are given in Table 8. The detection rates produced by both wavelet energy and California algorithms dropped slightly as compared to the case when one lane is blocked (Table 7). This is because an incident that does not block any lanes produces a less severe disruption in traffic flow than an incident that blocks at least one lane. This is especially true when the flow rate is low (1000 vph per lane). For the same reason also, the average detection time by California algorithm is longer as it takes more time for the congestion to develop and be detected by the algorithm. The detection time of the wavelet energy algorithm is in the range of 40-145 seconds while that of the California algorithm is in the range of 252-580 seconds.

Evaluation Using Real Data

Limited usable real traffic data was available to the authors. Real traffic flow and incident data are extracted from the San Francisco bay area freeway service patrol project's I-880 database for evaluation of the wavelet energy and California algorithms. Data for 21 incidents that block at least one lane are used to determine detection rate

performance, while 4 hours of incident-free data are used to ascertain the false alarm rate performance. The time of incident information in the database is inaccurate and therefore cannot be used to determine detection times. The performance of the wavelet energy and California algorithms using real data is shown in Table 9. The wavelet energy algorithm outperformed the California algorithm in both detection and false alarm rate. In particular, the wavelet energy algorithm did not signal any false alarm at all. In contrast, the California algorithm produced false alarm rate of 0.63% for this small real data set. It should be noted that this evaluation was also done without re-calibrating or re-training the algorithms. Also, note that the algorithms have been trained/calibrated using simulated data only. The detection rate of the wavelet energy incident detection algorithm can be improved when a good amount of real data is available.

PERFORMANCE SUMMARY AND CONCLUSION

Transferability or portability is a qualitative property of a freeway incident detection algorithm that determines how well the algorithm performs across various traffic flow and roadway geometry conditions. In all the tests performed in this evaluation the algorithms are not re-calibrated or retrained. Thus, a good way to assess the algorithms' portability is to compare their *performance vectors* across different test scenarios. A *performance vector* is defined as a vector with three performance elements: the percentage of missed detections (equal to 100 minus the detection rate), the false alarm rate, and the detection time. The smaller the value of each element the better the performance. Table 10 gives the performance vectors for the wavelet energy and California algorithms for the various scenarios evaluated in this research (extracted from Tables 1 through 3 and 6 through 8). The wavelet energy algorithm performed

consistently well across all scenarios including typical rural and urban freeway conditions. Furthermore, for any given scenario the wavelet energy algorithm outperformed the California algorithm #8. This result establishes the portability of the wavelet energy algorithm and demonstrates its excellent performance for urban freeways across a wide range of traffic flow and roadway geometry conditions regardless of the density of the loop detectors.

To the best of the authors' knowledge, no systematic evaluation of any existing incident detection algorithm has ever been published in the literature before. This paper presented the first investigation of this kind. Considering the difficulty in automatic detection of incidents on rural freeways, the new wavelet energy algorithm performs well on such freeways with detectors being placed a large 3 km apart, except when the flow rate is lower than 500 vph per lane. It is unlikely that a passive incident detection algorithm based on loop detector data can perform better than the wavelet energy algorithm in such low flow rate conditions; the traffic is just not affected enough to be detected reliably.

It is concluded that the new wavelet energy algorithm is not only highly robust and suitable for practical implementation on large urban freeway systems but also suitable and cost-effective for implementation on most rural freeways.

ACKNOWLEDGMENT

This manuscript is based on a research project sponsored by the Ohio Department of Transportation and Federal Highway Administration.

APPENDIX I. REFERENCES

Abdulhai, B. and Ritchie, S. G. (1999), "Enhancing the Universality and transferability of Freeway Incident Detection Using a Bayesian-based Neural Network," *Transportation Research Part C*, Vol. 7, pp. 261-280.

Adeli, H. and Hung, S.L. (1995), *Machine Learning - Neural Networks, Genetic Algorithms, and Fuzzy Sets*, John Wiley and Sons, New York.

Adeli, H. and Park, H.S. (1998), *Neurocomputing for Design Automation*, CRC Press, Boca Raton, Florida.

Adeli, H. and Karim, A. (2000), "Fuzzy-Wavelet RBFNN Model for Freeway Incident Detection," *Journal of Transportation Engineering*, ASCE, Vol. 126, No. 6, pp. 464-471.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA.

Karim, A. and Adeli, H (2001a), "Comparison of the Fuzzy-Wavelet RBFNN Freeway Incident Detection Model with the California Algorithm," *Journal of Transportation Engineering*, ASCE, accepted.

Karim, A. and Adeli, H. (2001b), "Incident Detection Algorithm Using Wavelet Energy Representation of Traffic Patterns," *Journal of Transportation Engineering*, ASCE, accepted.

Payne, H. J. and Tignor, S. C. (1978), "Freeway Incident-Detection Algorithms Based on Decision Trees With States," *Transportation Research Record*, No. 682, pp. 30-37.

Samant, A. and Adeli, H. (2000), "An Adaptive Conjugate Gradient Neural Network-Wavelet Model for Traffic Incident Detection," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 15, No. 4, pp. 241-250.

Table 1 Performance of the new wavelet energy algorithm and California algorithm #8 on a two-lane freeway

Flow rate (vph per lane)	Location (m) *	Wavelet energy Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	Detection time (s)
1000	152	5/5	0/150	80	5/5	0/150	480
	305	5/5	0/150	96	5/5	0/150	384
	457	5/5	0/150	68	5/5	0/150	252
	610	5/5	0/150	112	5/5	0/150	164
1500	152	5/5	0/150	68	5/5	1/150	228
	305	5/5	0/150	80	5/5	0/150	176
	457	5/5	0/150	92	5/5	0/150	132
	610	5/5	0/150	96	5/5	0/150	92
2000	152	5/5	0/150	68	5/5	0/150	132
	305	5/5	0/150	92	5/5	1/150	116
	457	5/5	0/150	92	5/5	2/150	84
	610	5/5	0/150	124	5/5	0/150	96
Totals		60/60 100%	0/1800 0%		60/60 100%	4/1800 0.22%	

* Location of the incident from the downstream detector station. The distance between detector stations is 762 m.

Table 2 Performance of the new wavelet energy algorithm and California algorithm #8 on a three-lane freeway

Flow rate (vph per lane)	Location (m) *	Wavelet energy Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	Detection time (s)
1000	152	5/5	0/150	56	0/5	0/150	-
	305	5/5	0/150	68	0/5	0/150	-
	457	5/5	0/150	80	0/5	0/150	-
	610	5/5	0/150	72	5/5	0/150	248
1500	152	5/5	0/150	56	5/5	0/150	264
	305	5/5	0/150	76	5/5	0/150	208
	457	5/5	0/150	76	5/5	1/150	132
	610	5/5	0/150	88	5/5	0/150	96
2000	152	5/5	0/150	88	5/5	0/150	148
	305	5/5	0/150	116	5/5	0/150	136
	457	5/5	0/150	100	5/5	0/150	92
	610	5/5	0/150	96	5/5	1/150	76
Totals		60/60 100%	0/1800 0%		45/60 75%	2/1800 0.11%	

* Location of the incident from the downstream detector station. The distance between detector stations is 762 m.

Table 3 Performance of the new wavelet energy algorithm and California algorithm #8 on a four-lane freeway

Flow rate (vph per lane)	Location (m) *	Wavelet energy Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	Detection time (s)
1000	152	5/5	0/150	60	0/5	0/150	-
	305	5/5	0/150	68	0/5	0/150	-
	457	5/5	0/150	72	2/5	0/150	440
	610	5/5	0/150	84	5/5	0/150	168
1500	152	5/5	0/150	72	5/5	0/150	268
	305	5/5	0/150	96	5/5	0/150	188
	457	5/5	0/150	92	5/5	1/150	132
	610	5/5	0/150	84	5/5	0/150	96
2000	152	5/5	0/150	68	5/5	1/150	140
	305	5/5	0/150	84	5/5	1/150	128
	457	5/5	0/150	84	5/5	2/150	96
	610	5/5	0/150	108	5/5	0/150	84
Totals		60/60 100%	0/1800 0%		45/60 75%	5/1800 0.28%	

* Location of the incident from the downstream detector station. The distance between detector stations is 762 m.

Table 4 Description of the four simulation scenarios used for evaluating the false alarm performance on a three-lane freeway with ramps

Scenario #	Time period #	Entry flow rate (vph)	On-ramp flow rate (vph)		Off-ramp flow rate (vph)	
			A	B	A	B
1	1	4500	300	500	225	450
	2	4800	300	500	240	480
	3	4500	300	300	225	450
2	1	5250	300	500	260	525
	2	5500	300	500	275	550
	3	5259	300	300	260	525
3	1	4000	600	600	200	400
	2	4500	600	600	225	450
	3	4000	600	600	200	400
4	1	5500	600	600	275	550
	2	6000	600	600	300	600
	3	5500	600	600	275	550

Table 5 False alarm performance of the wavelet energy and California algorithm #8 for the three-lane freeway with ramps

Station #	False alarms (out of 1125 decisions for each station in a senario)							
	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	WE	Cal.	WE	Cal.	WE	Cal.	WE	Cal.
1	0		0		0		0	
2	0	0	0	0	0	0	0	0
3	0	3	0	1	0	1	0	0
4	0	0	0	1	0	0	0	5
5	0	51	0	130	0	27	0	207
6	0	1	0	0	0	2	0	3
	0%	0.98%	0%	2.34%	0%	0.53%	0%	3.82%

WE = Wavelet energy algorithm; Cal. = California algorithm #8

Table 6 Performance of the wavelet energy algorithm and California algorithm #8 on a two-lane rural freeway (incident duration is 10 minutes; 1 lane is blocked, the other lane's capacity is reduced by 40%)

Flow rate (vph per lane)	Location (m) *	Wavelet energy Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	Detection time (s)
500	152	0/5	3/125	-	0/5	0/125	-
	305	0/5	2/125	-	0/5	0/125	-
	610	2/5	6/125	240	0/5	0/125	-
	915	0/5	1/125	-	0/5	0/125	-
	1220	2/5	0/125	280	0/5	3/125	-
	1524	1/5	0/125	20	0/5	0/125	-
	1829	1/5	0/125	20	0/5	0/125	-
	2134	2/5	0/125	130	0/5	0/125	-
	2439	0/5	1/125	-	0/5	0/125	-
	2744	1/5	0/125	180	0/5	0/125	-
Totals		9/50 18%	13/1250 1.04%		0/50 0%	3/1250 0.24%	
1000	152	4/5	0/125	150	0/5	0/125	-
	305	5/5	1/125	80	0/5	0/125	-
	610	4/5	0/125	125	0/5	0/125	-
	915	3/5	0/125	153	0/5	0/125	-
	1220	5/5	0/125	156	0/5	0/125	-
	1524	3/5	0/125	153	0/5	0/125	-
	1829	5/5	0/125	164	0/5	0/125	-
	2134	5/5	0/125	186	0/5	0/125	-
	2439	5/5	0/125	188	5/5	0/125	452
	2744	5/5	0/125	152	5/5	0/125	244
Totals		44/50 88%	1/1250 0.08%		10/50 20%	0/1250 0%	

Table 6 – continued

Flow rate (vph per lane)	Location (m) *	Wavelet energy Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	Detection time (s)
1500	152	5/5	0/125	92	0/5	0/125	-
	305	5/5	0/125	76	0/5	0/125	-
	610	5/5	1/125	68	3/5	0/125	246
	915	5/5	0/125	44	3/5	0/125	406
	1220	5/5	0/125	120	5/5	0/125	500
	1524	5/5	0/125	120	5/5	0/125	428
	1829	5/5	0/125	120	5/5	0/125	332
	2134	5/5	0/125	116	5/5	0/125	236
	2439	5/5	0/125	160	5/5	0/125	180
	2744	5/5	0/125	156	5/5	0/125	152
Totals		50/50 100%	1/1250 0.08%		36/50 72%	0/1250 0%	
2000	152	5/5	0/125	52	5/5	2/125	160
	305	5/5	0/125	60	5/5	1/125	232
	610	5/5	0/125	64	5/5	0/125	228
	915	5/5	0/125	84	5/5	0/125	168
	1220	5/5	0/125	68	5/5	0/125	164
	1524	5/5	0/125	112	5/5	1/125	212
	1829	5/5	0/125	100	5/5	0/125	176
	2134	5/5	0/125	136	5/5	1/125	160
	2439	5/5	0/125	156	5/5	2/125	148
	2744	5/5	0/125	140	5/5	0/125	148
Totals		50/50 100%	0/1250 0%		50/50 100%	7/1250 0.56%	

* Location of the incident from the downstream detector station. The distance between detector stations is 3048 m.

Table 7 Performance of the wavelet energy algorithm and California algorithm #8 on a two-lane rural freeway (incident duration is 5 minutes; 1 lane is blocked, the other lane's capacity is reduced by 40%)

Flow rate (vph per lane)	Location (m) *	Wavelet energy Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	Detection time (s)
1000	152	5/5	0/125	104	0/5	0/125	-
	305	5/5	0/125	120	0/5	0/125	-
	610	3/5	0/125	160	0/5	0/125	-
	915	5/5	1/125	120	0/5	0/125	-
	1220	4/5	0/125	85	0/5	0/125	-
	1524	3/5	0/125	146	0/5	0/125	-
1500	152	5/5	0/125	68	0/5	0/125	-
	305	5/5	0/125	80	1/5	0/125	100
	610	5/5	0/125	80	0/5	0/125	-
	915	5/5	0/125	112	1/5	0/125	120
	1220	5/5	0/125	96	0/5	0/125	-
	1524	5/5	0/125	88	4/5	0/125	430
2000	152	5/5	0/125	44	5/5	0/125	204
	305	5/5	0/125	60	1/5	0/125	80
	610	5/5	0/125	72	5/5	0/125	184
	915	5/5	0/125	80	5/5	0/125	132
	1220	5/5	0/125	72	5/5	1/125	168
	1524	5/5	0/125	112	5/5	0/125	192
Totals		85/90 94.4%	1/2250 0.04%		32/90 35.6%	1/2250 0.04%	

* Location of the incident from the downstream detector station. The distance between detector stations is 3048 m.

Table 8 Performance of the wavelet energy algorithm and California algorithm #8 on a two-lane rural freeway (incident duration is 10 minutes; no lane is blocked, the capacity of each lane is reduced by 40%)

Flow rate (vph per lane)	Location (m) *	Wavelet energy Algorithm			California Algorithm #8		
		Detections	False alarms	Detection time (s)	Detections	False alarms	Detection time (s)
1000	152	0/5	0/125	-	0/5	0/125	-
	305	0/5	0/125	-	0/5	0/125	-
	610	1/5	0/125	80	0/5	0/125	-
	915	0/5	0/125	-	0/5	0/125	-
	1220	2/5	0/125	60	0/5	0/125	-
	1524	3/5	0/125	113	0/5	0/125	-
1500	152	3/5	0/125	80	0/5	0/125	-
	305	2/5	0/125	120	0/5	0/125	-
	610	2/5	0/125	60	0/5	0/125	-
	915	2/5	0/125	145	0/5	0/125	-
	1220	3/5	0/125	127	0/5	0/125	-
	1524	5/5	0/125	128	0/5	0/125	-
2000	152	5/5	0/125	40	4/5	0/125	580
	305	5/5	0/125	60	5/5	0/125	508
	610	5/5	0/125	68	5/5	0/125	444
	915	5/5	0/125	72	5/5	0/125	444
	1220	5/5	0/125	80	5/5	0/125	252
	1524	5/5	0/125	116	5/5	0/125	276
Totals		53/90 58.9%	0/2250 0%		29/90 32.2%	0/2250 0%	

* Location of the incident from the downstream detector station. The distance between detector stations is 3048 m.

Table 9 Performance of the wavelet energy and California algorithms using real traffic data from the San Francisco bay area freeway service patrol project's I-880 database

Detection rate		False alarms	
WE	Cal.	WE	Cal.
20/21	19/21	0/480	3/480
95.2%	90.5%	0%	0.63%

WE = Wavelet energy algorithm; Cal. = California algorithm #8

Table 10 Performance vector for assessment of algorithm portability

Wavelet energy algorithm	California algorithm #8
0, 0, 89	0, 0, 320
0, 0, 84	0, 0.17, 157
0, 0, 94	0, 0.5, 102
0, 0, 69	75, 0, 248
0, 0, 74	0, 0.17, 175
0, 0, 100	0, 0.17, 113
0, 0, 71	65, 0, 304
0, 0, 86	0, 0.17, 171
0, 0, 86	0, 0.67, 112
82, 1.04, 145	100, 0.24, inf
12, 0.08, 151	80, 0, 348
0, 0.08, 107	28, 0, 310
0, 0, 97	0, 0.56, 180
17, 0.13, 122	100, 0, inf
0, 0, 87	80, 0, 217
0, 0, 73	13, 0.13, 160
80, 0, 84	100, 0, inf
60, 0, 110	100, 0, inf
0, 0, 73	3, 0, 417

inf = No incidents are detected and the detection time is theoretically equal to infinity.

LIST OF CAPTIONS FOR FIGURES

1. Information processing in an advanced traffic management system
2. The wavelet energy freeway incident detection algorithm
3. Layout of urban freeway segments simulated for parametric evaluation
4. Variation of detection time with distance of incident from downstream detector station on a 2-lane urban freeway for the wavelet energy algorithm (denoted by WE) and the California Algorithm #8 (denoted by Cal)
5. Variation of detection time with distance for the wavelet energy algorithm on 2-, 3-, and 4-lane urban freeway segments when flow rate is 2000 vph per lane
6. Layout of urban freeway with ramps evaluated for false alarm performance
7. Lane occupancy plot at downstream detector station on a 2-lane rural freeway when flow rate is 500 vph per lane

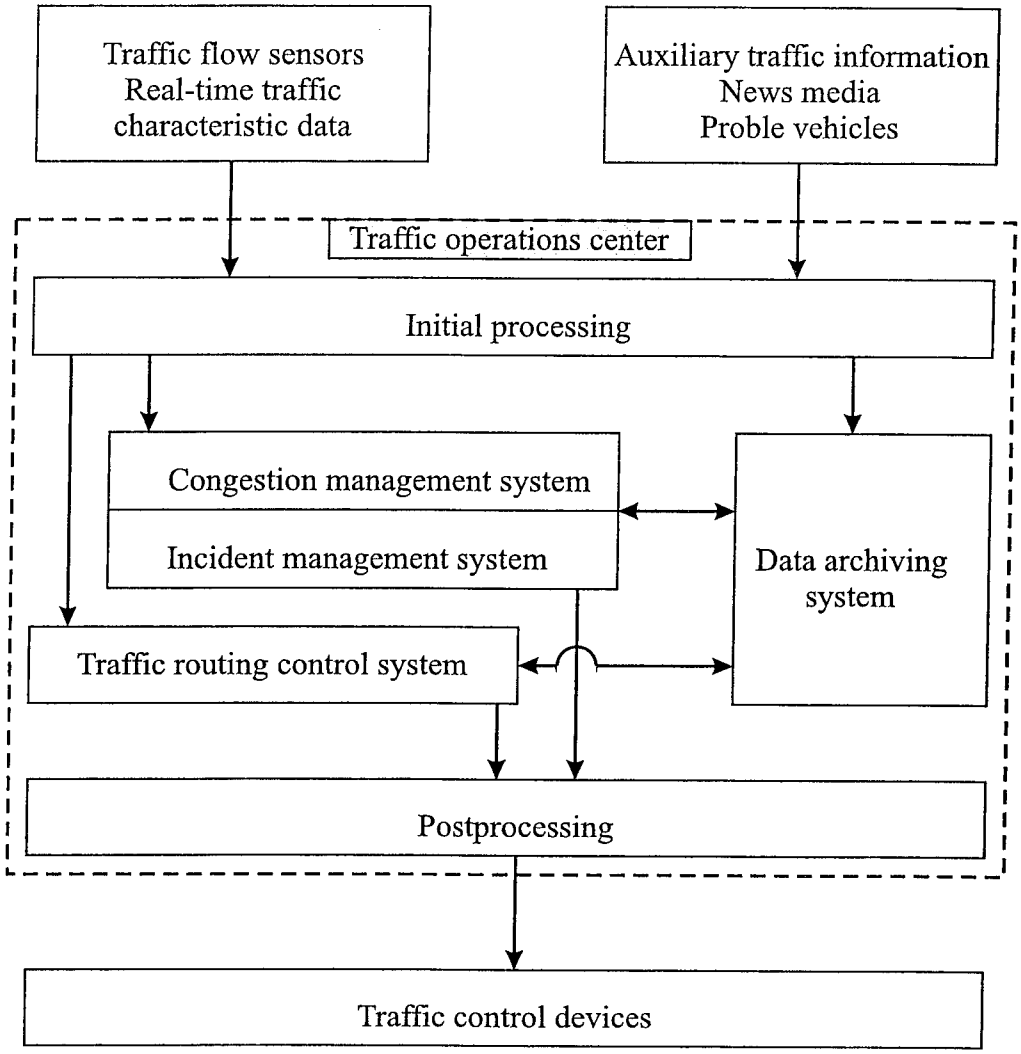

Traffic flow sensors
Real-time traffic
characteristic data
Auxiliary traffic information
News media
Proble vehicles
Traffic operations center
Initial processing
Congestion management system
Incident management system
Data archiving
system
Traffic routing control system
Postprocessing
Traffic control devices

Figure 1

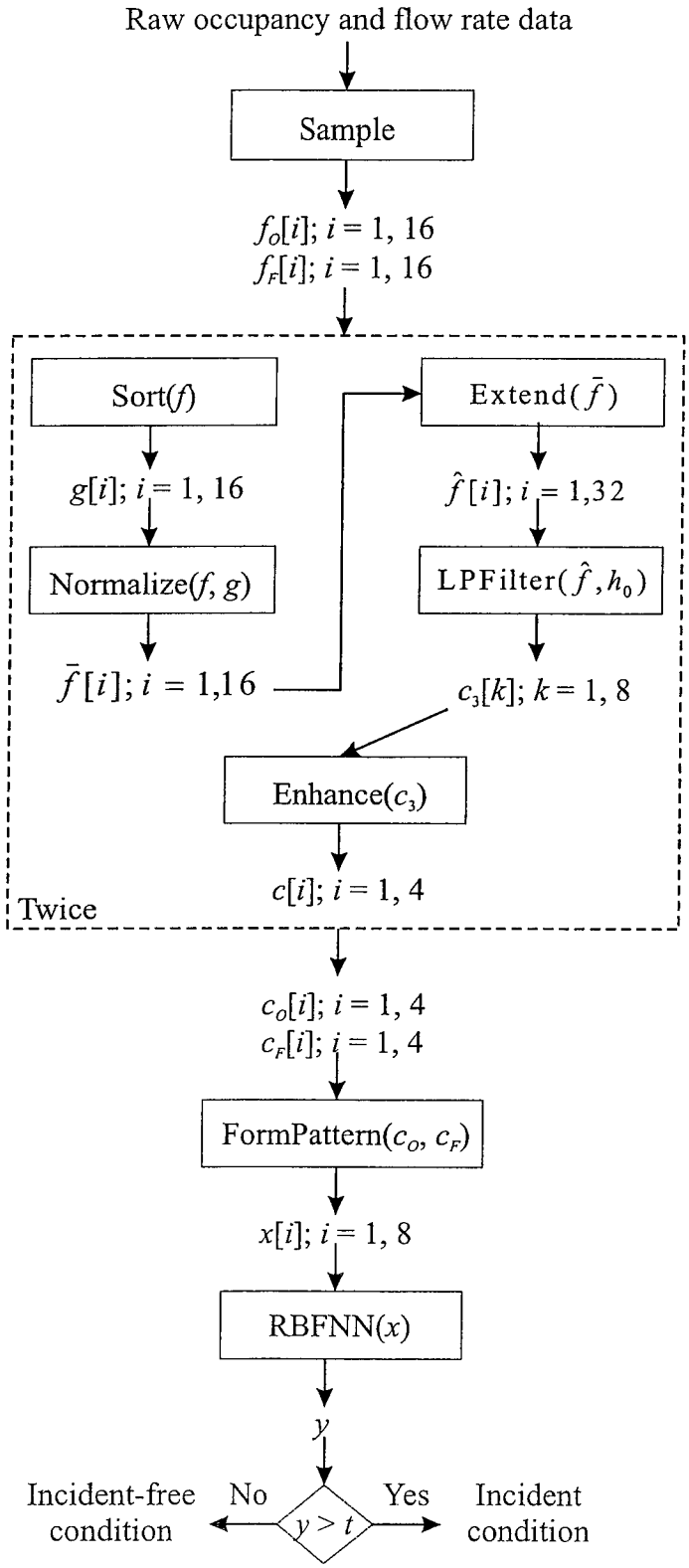

Raw occupancy and flow rate data
Sample
$f_O[i]; i = 1, 16$
$f_F[i]; i = 1, 16$
Sort(f)
Extend($\bar{f}$)
$g[i]; i = 1, 16$
$\hat{f}[i]; i = 1,32$
Normalize(f, g)
LPFilter($\hat{f}, h_0$)
$\bar{f}[i]; i = 1,16$
$c_3[k]; k = 1, 8$
Enhance(c_3)
$c[i]; i = 1, 4$
Twice
$c_O[i]; i = 1, 4$
$c_F[i]; i = 1, 4$
FormPattern(c_O, c_F)
$x[i]; i = 1, 8$
RBFNN(x)
y
Incident-free condition
No
$y > t$
Yes
Incident condition

Figure 2

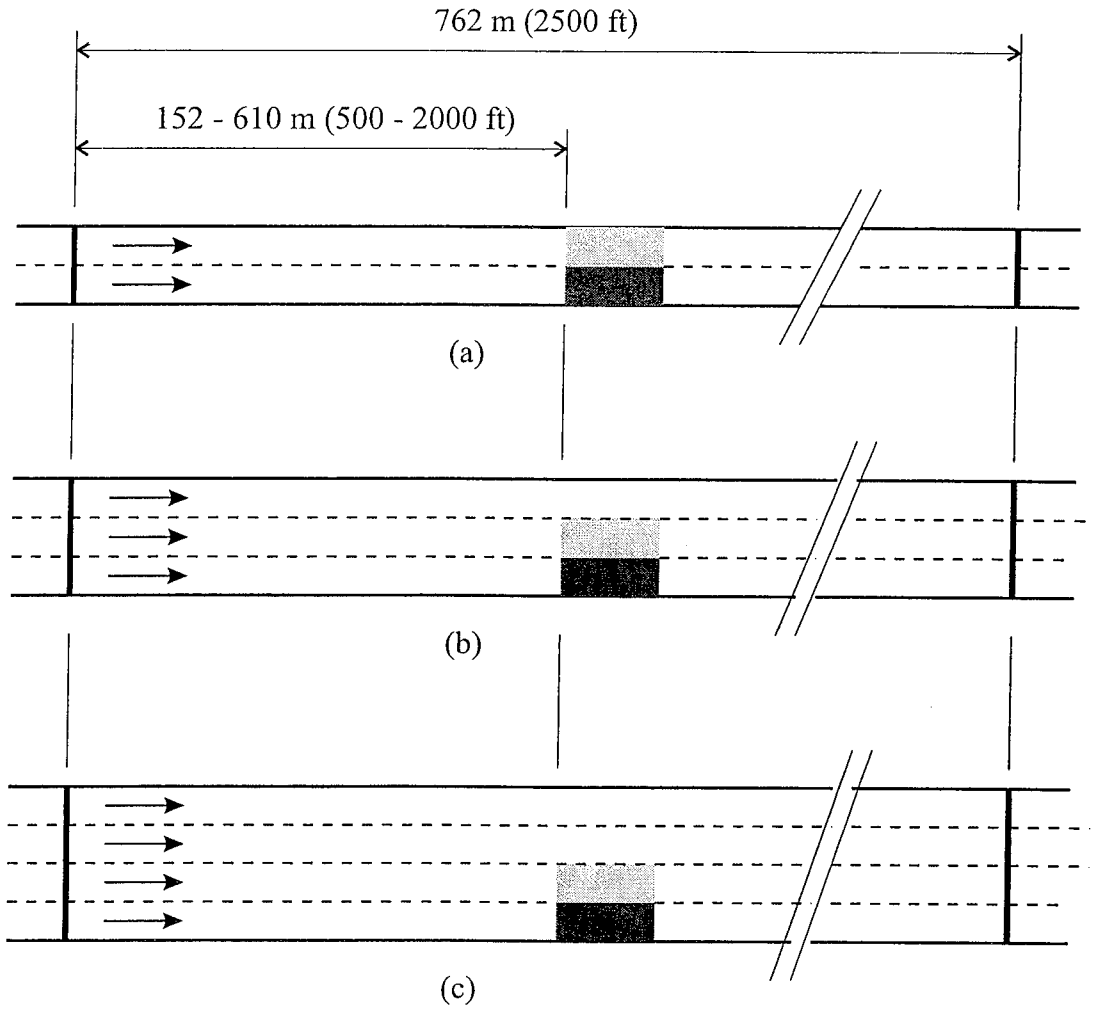

762 m (2500 ft)
152 - 610 m (500 - 2000 ft)
(a)
(b)
(c)
Legend
Detector station
Blocked lane
Lane with reduced capacity

Figure 3

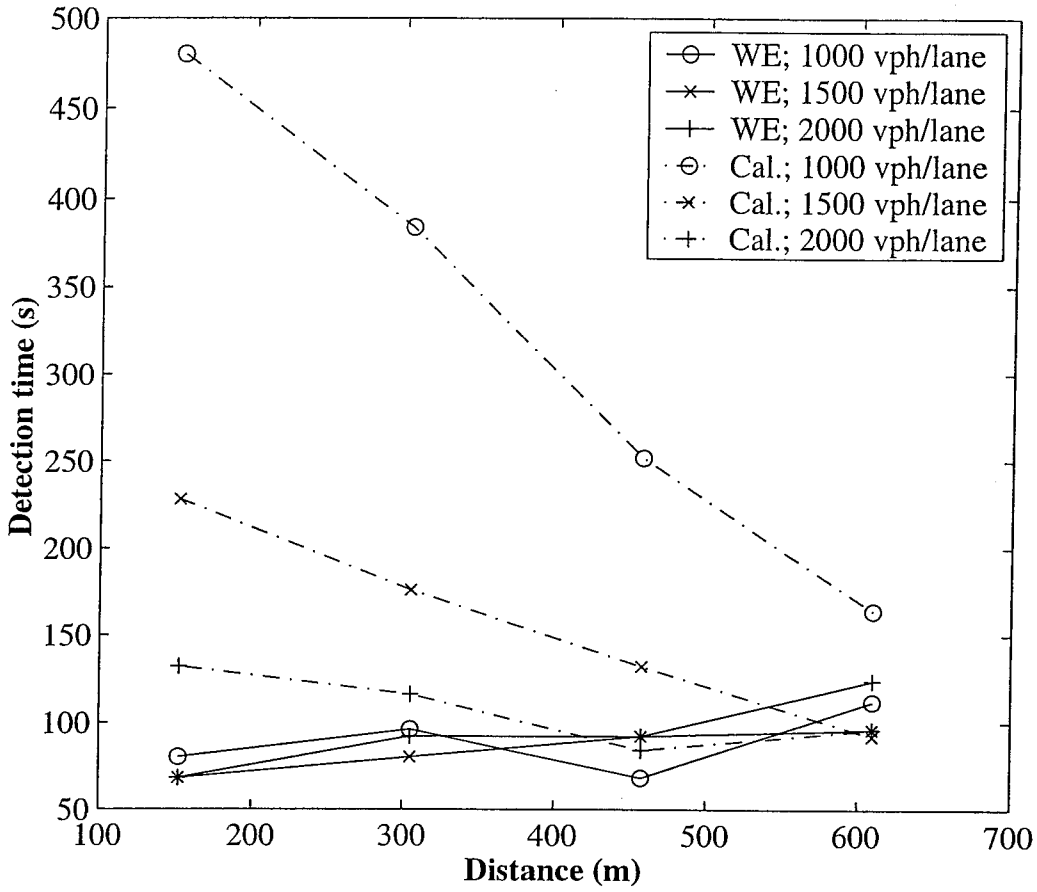

WE; 1000 vph/lane
WE; 1500 vph/lane
WE; 2000 vph/lane
Cal.; 1000 vph/lane
Cal.; 1500 vph/lane
Cal.; 2000 vph/lane
Detection time (s)
Distance (m)
500
450
400
350
300
250
200
150
100
50
100
200
300
400
500
600
700

Figure 4

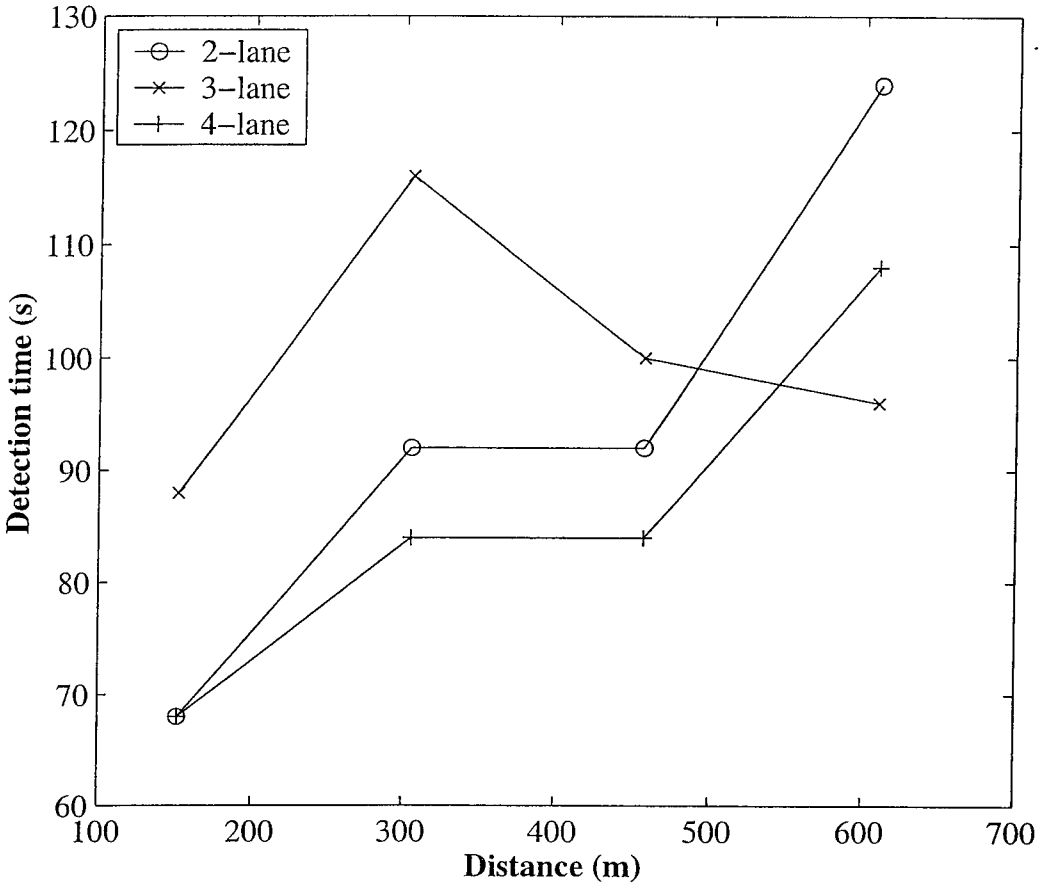

2–lane
3–lane
4–lane
Detection time (s)
Distance (m)
130
120
110
100
90
80
70
60
100
200
300
400
500
600
700

Figure 5

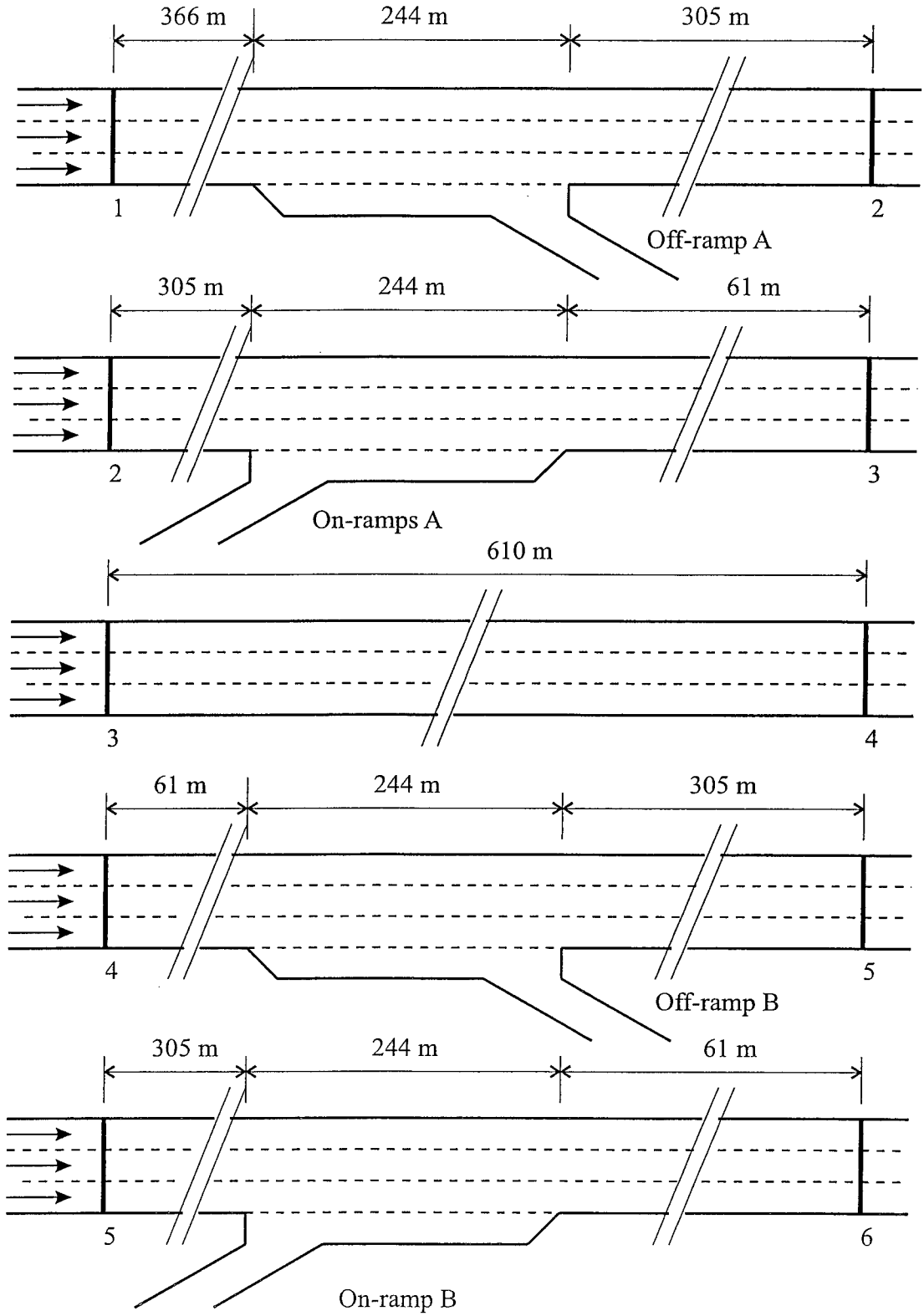

366 m
244 m
305 m
1
2
Off-ramp A
305 m
244 m
61 m
2
3
On-ramps A
610 m
3
4
61 m
244 m
305 m
4
5
Off-ramp B
305 m
244 m
61 m
5
6
On-ramp B
Legend
Detector station

Figure 6

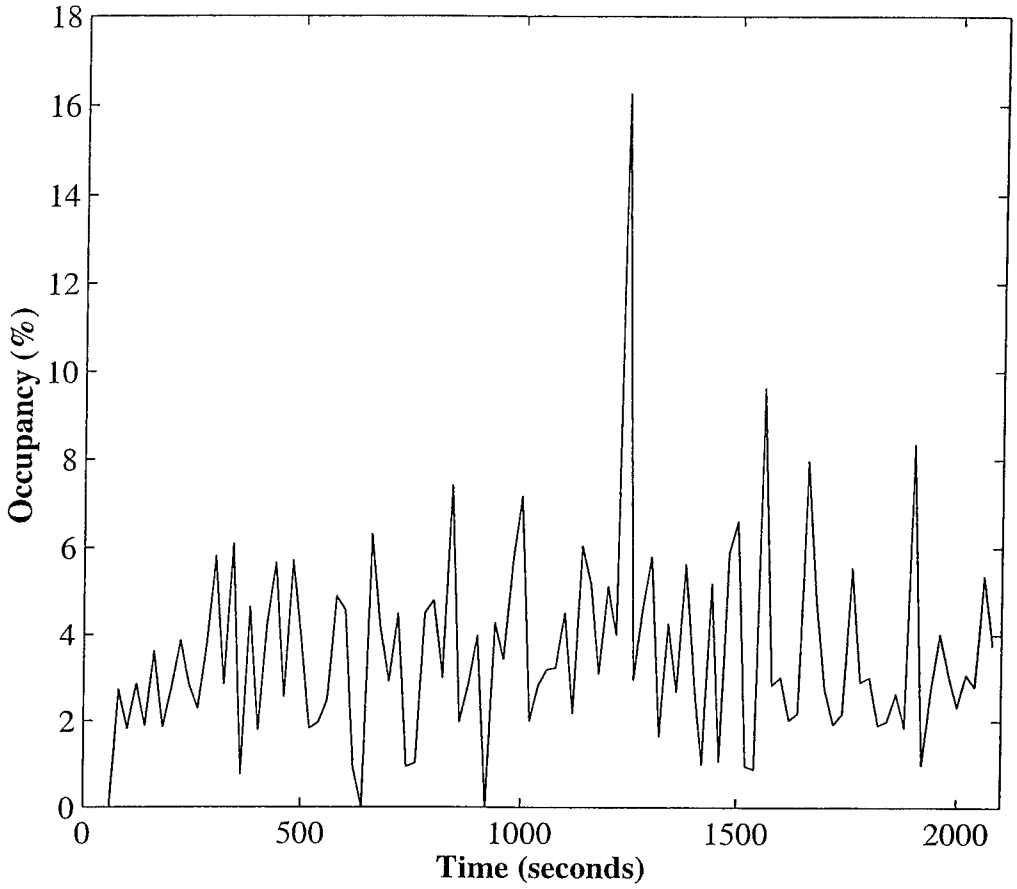

Occupancy (%)
Time (seconds)
0
2
4
6
8
10
12
14
16
18
0
500
1000
1500
2000

Figure 7