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 (ATMS) that provides information for emergency relief and traffic control and management purposes. Earlier algorithms for 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 denoise, 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 nonincident 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 online in any ATMS without any need for recalibration.

DOI: 10.1061/(ASCE)0733-947X(2002)128:3(232)

CE Database keywords: Traffic management; Traffic accidents; Algorithms; Highways.

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-s 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 nonincident 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 effectively with erroneous data from malfunctioning traffic sensors.

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 a significant reduction in lane speed and a 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 direction 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.

[1]Graduate Research Associate, Dept. of Civil and Environmental Engineering and Geodetic Science, Ohio State Univ., 2070 Neil Ave., Columbus, OH 43210.

[2]Professor, Dept. of Civil and Environmental Engineering and Geodetic Science, Ohio State Univ., 470 Hitchcock Hall, 2070 Neil Ave., Columbus, OH 43210.

Note. Discussion open until October 1, 2002. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on August 22, 2000; approved on September 18, 2001. This paper is part of the ***Journal of Transportation Engineering***, Vol. 128, No. 3, May 1, 2002. ©ASCE, ISSN 0733-947X/2002/3-232–242/$8.00+$.50 per page.

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, upstream lane occupancy and speed time series data are adopted as the characterizing pattern for traffic state classification. The raw data is first denoised by soft thresholding in the wavelet domain. Subsequently, the denoised 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 denoised 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 traffic and incident characteristics, can be as large as 5 min.

In this work, 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 denoises the traffic patterns, which are then classified using a radial basis function (RBF) neural network. The new algorithm is developed, described, and evaluated in 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. When selecting appropriate patterns for an effective incident detection algorithm three goals were set.

- 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; and
- Third, the patterns should indicate an incident condition in less than 1 min 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.fhwatsis.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; whereas, 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-ramps and off-ramps, and the prevailing flow conditions.

Figs. 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 1,000, 1,250, and 1,500 vehicles per hour (vph) per lane. The incident occurs at time 400 s. Note that the time at which the upstream traffic occupancy and speed change [Figs. 1(a and b)] depends on the preincident flow rate. The formation of a queue, which produces significant changes in traffic occupancy and speed patterns, also depends on the reduction in the capacity and roadway conditions (not presented in the figures). Fig. 1(c) 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 traffic occupancy and flow rate [Figs. 2(a and c)] but no significant change in traffic speed [Fig. 2(b)]. 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 (Cook and Cleveland 1974; Persaud and Hall 1989) 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, the computational model relies on single-station patterns. This model can handle patterns from both up-

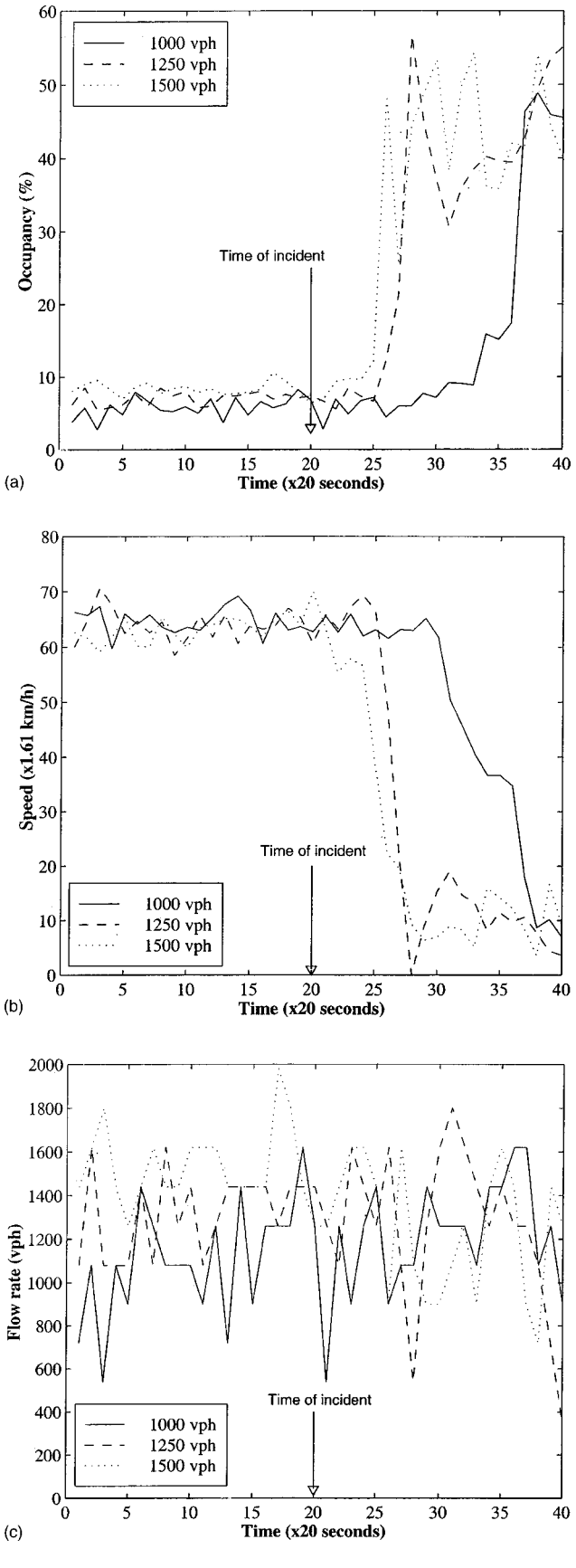

Fig. 1. Time-series plots of upstream traffic data on two-lane freeway with three prevailing flow rates of 1,000, 1,250, and 1,500 vph per lane before and after incident (a) lane occupancy plot; (b) lane speed plot; (c) lane flow rate plot

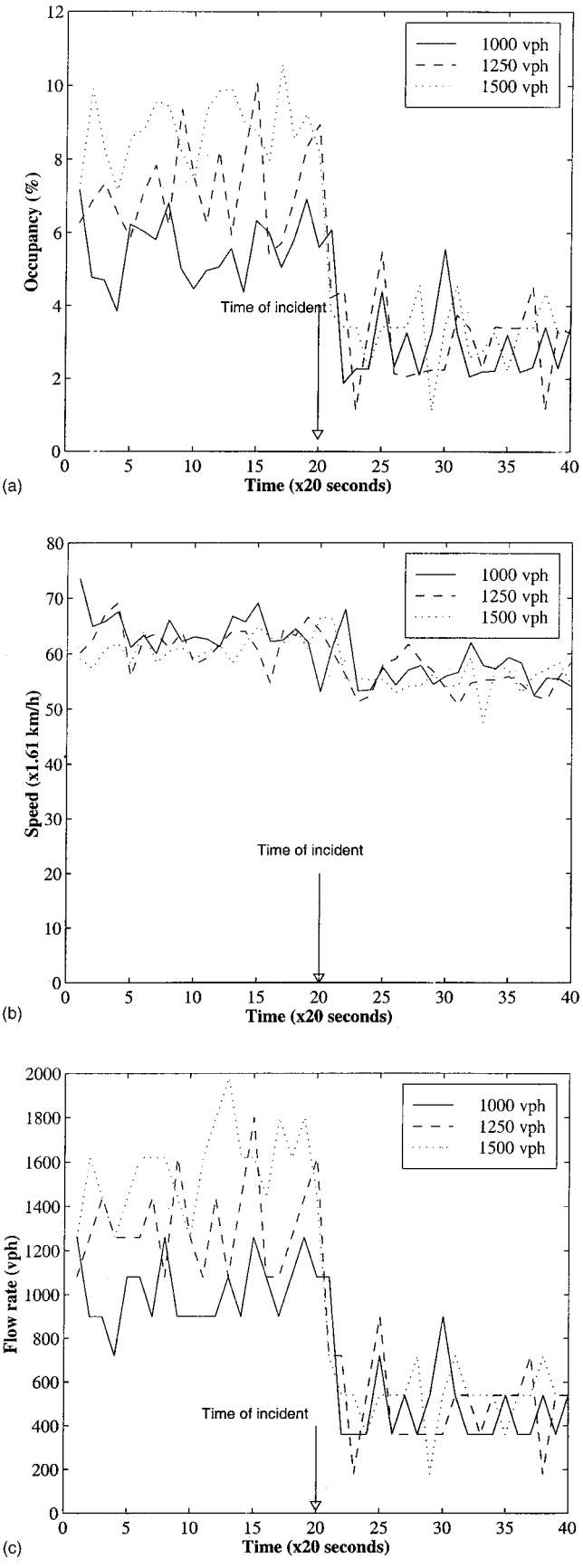

Fig. 2. Time-series plots of downstream traffic data on two-lane freeway with three prevailing flow rates of 1,000, 1,250, and 1,500 vph per lane before and after incident (a) lane occupancy plot; (b) lane speed plot; (c) lane flow rate plot

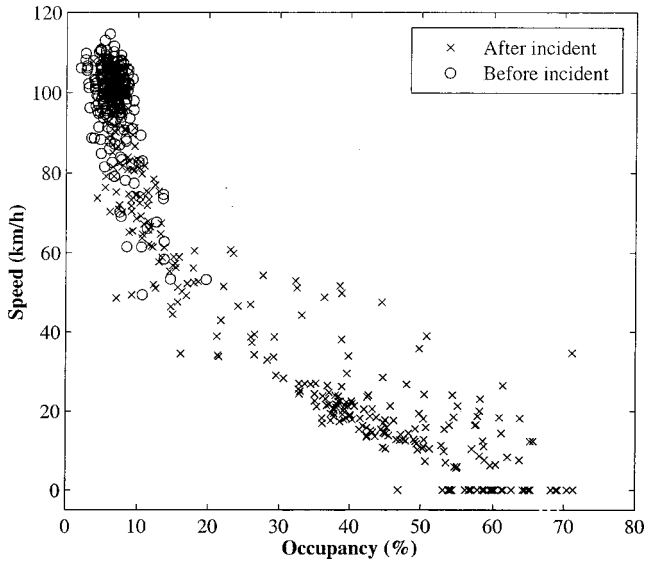

Fig. 3. Scatter plot of upstream lane occupancy and speed before and after incidents

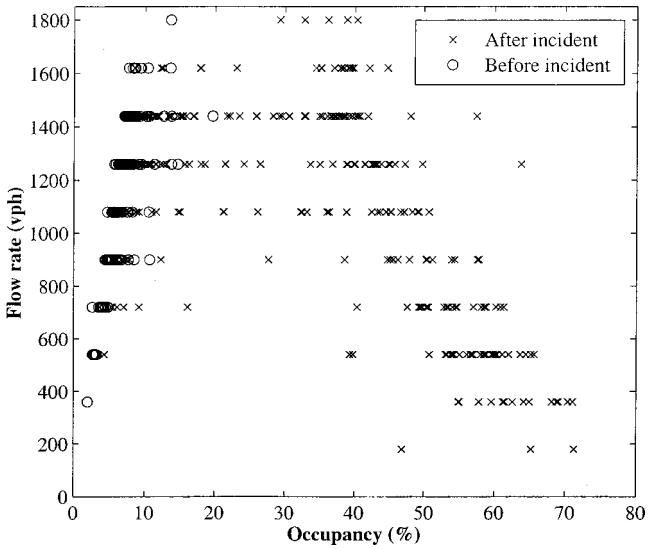

Fig. 4. Scatter plot of upstream lane occupancy and flow rate before and after incidents

stream 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 Figs. 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 occupancy and the speed. A pattern based on the upstream time histories of lane occupancy and speed is therefore most appropriate for reliable incident detection purposes. This conclusion is confirmed by Fig. 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 denoising and feature enhancement.) On the other hand, the scatter plot of occupancy and flow rate (Fig. 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 [Figs. 2(a–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 Figs. 2(a 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 Fig. 2(b). 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 (Fig. 5) and occupancy and flow rate (Fig. 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 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.

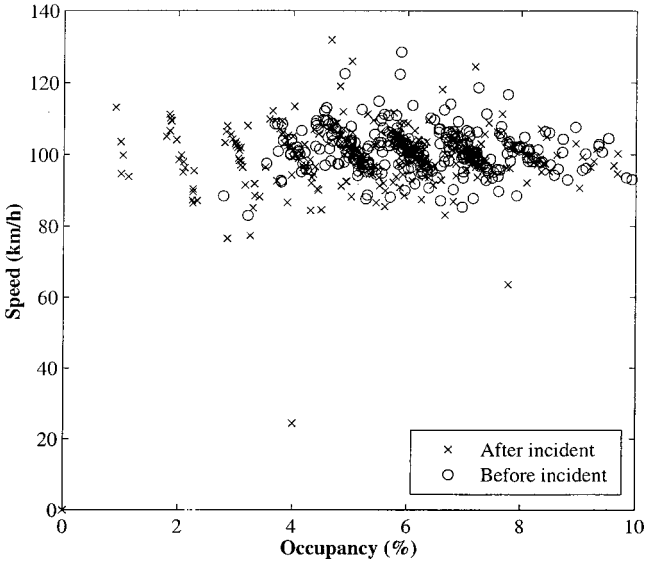

Fig. 5. Scatter plot of downstream lane occupancy and speed before and after incidents

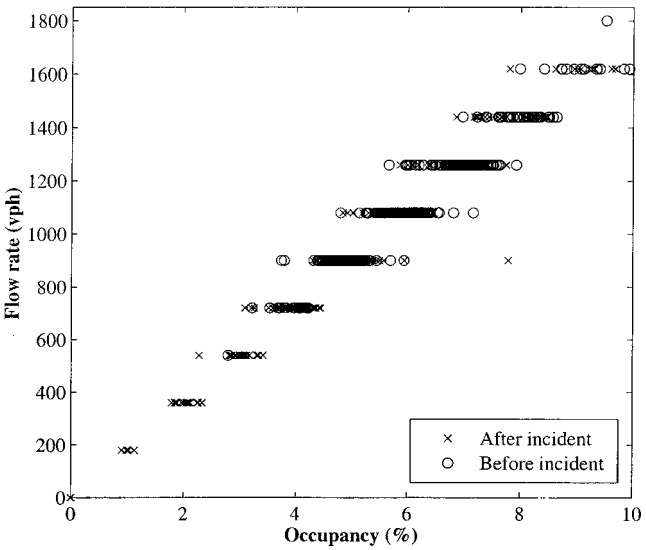

Fig. 6. Scatter plot of downstream lane occupancy and flow rate before and after incidents

On the basis of 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 denoising approach is adopted to process the data. False alarms from compression waves are avoided by using a sufficiently long time series (>5 min) 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, denoising, and compressing signals. In particular, the DWT of a signal breaks it down into several time-frequency components that enables the extraction of features desirable for signal identification and recognition. The DWT and wavelet theory in general have been developed rapidly in the past 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 1D 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)

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

where $L^2(R)$ is the space of all square integrable functions defined in the 1D 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:

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

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

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

where h_0 and h_1=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 a finite 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

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

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

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.

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 $f[k]$. The use of 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 Fig. 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 Denoising

In this traffic incident detection model, we process the time-series traffic data (lane occupancy, speed, and flow rate) obtained at each detector station, with the objectives of reducing noise and enhancing the desirable features. This processing is essential to ensure that no incidents go undetected and no false alarms are triggered. Upstream lane occupancy ($f_O[i]$) and speed ($f_S[i]$) form one pattern for identifying incident conditions. 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 16 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 s (which is the minimum available from current detector stations), 16 data points constitute 5 min and 20 s of data. Compression waves are usually temporary conditions and not very likely to exist for as long as 5 min. 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.

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 recalibration 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 $\overline{f}_O[i]$, $\overline{f}_S[i]$, and $\overline{f}_F[i]$, respectively.

The normalized data series are extended by eight points at each end before their DWT's are calculated. All eight data points in the extension have a magnitude equal to the average of the previous two (first two for the extension on one end and last two for the extension on the other end) values in the series. The extended normalized data series are given by

$$\hat{f}[i] = \begin{cases} 0.5(\overline{f}[1] + \overline{f}[2]) & 1 \leq i \leq 8 \\ \overline{f}[i-8] & 9 \leq i \leq 24 \\ 0.5(\overline{f}[15] + \overline{f}[16]) & 25 \leq i \leq 32 \end{cases} \qquad (11)$$

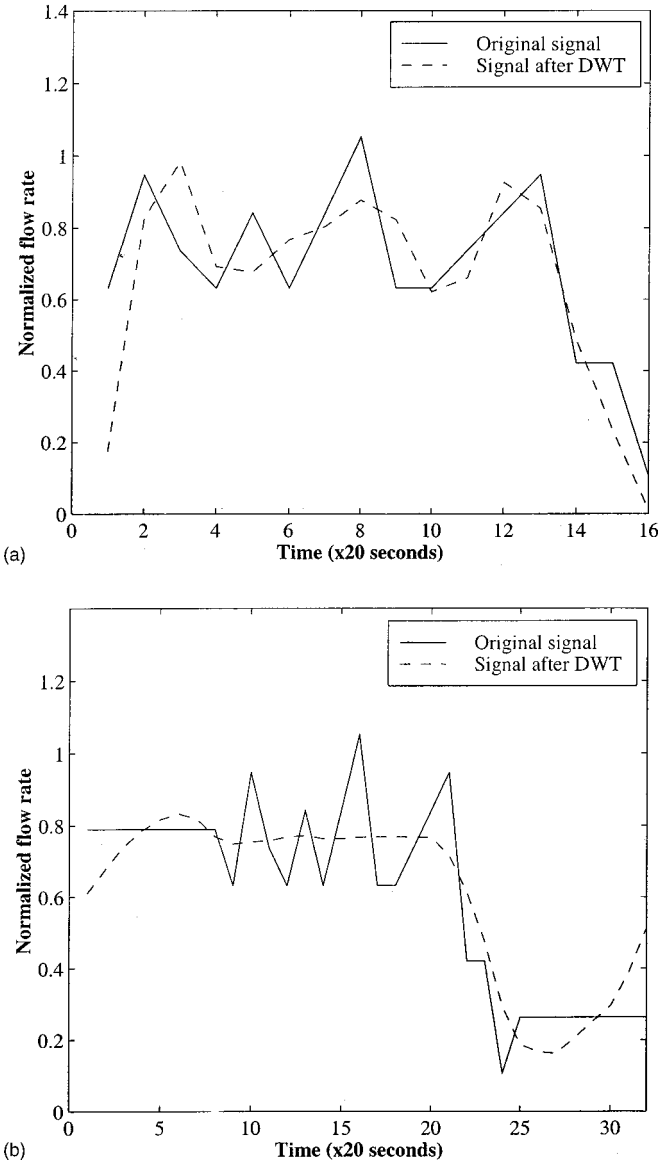

Fig. 7. (a) Discrete wavelet transform (DWT) of 16-point flow rate traffic pattern; (b) DWT of an extended 32-point flow rate traffic pattern [based on the data in (a)]

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 Fig. 7. Fig. 7(a) shows a typical flow rate data series $\overline{f}_F[i]$ (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. Fig. 7(b) 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). Eq.

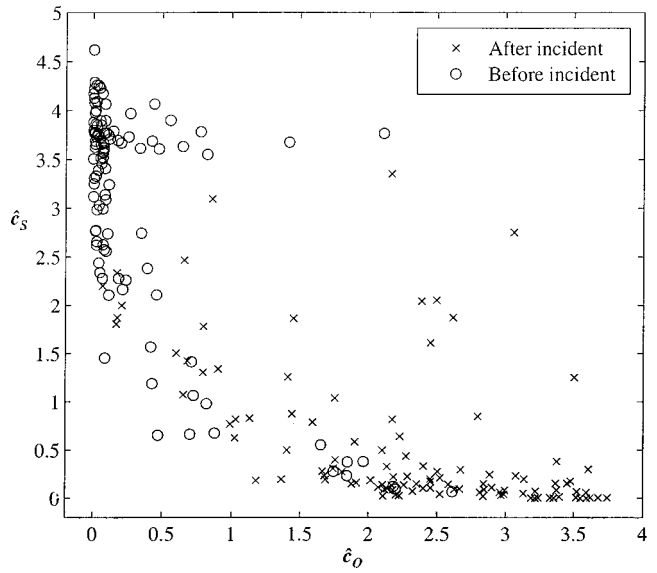

Fig. 8. Scatter plot of upstream lane occupancy and speed wavelet energy coefficients before and after incidents

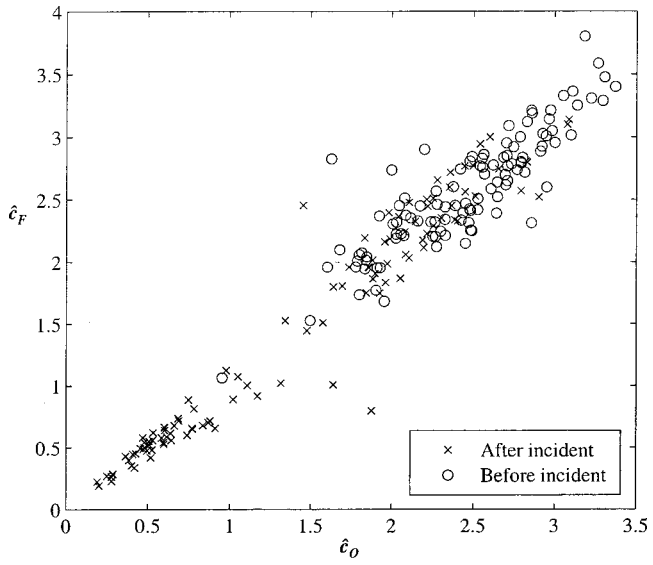

Fig. 9. Scatter plot of downstream lane occupancy and flow rate wavelet energy coefficients before and after incidents

(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 denoised shape or form. Also, through the two-stage low-pass filtering the 32-point time series is now reduced to an eight-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 four values of the eight 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 Figs. 1 and 2 that an incident condition pattern exhibits either a sudden decrease or a sudden increase in magnitude of data values that occur in the last few data points. This feature, which distinguishes an incident condition from a nonincident 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 denoised 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 nonincident 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 denoising and feature enhancement is demonstrated in Figs. 8 and 9. Fig. 8 is a scatter plot of $\hat{c}_O[i]$ and $\hat{c}_S[i]$ based on the same data used in Fig. 3. Fig. 9 is a scatter plot of $\hat{c}_O[i]$ and $\hat{c}_F[i]$ based on the same data used in Fig. 6. Comparisons of Fig. 3 with Fig. 8 and Fig. 6 with Fig. 9 indicate the improvement in pattern separation achieved by wavelet-based denoising 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 four 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 8D space to a 1D space. For this purpose, the radial basis function 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 Fig. 10. The input layer has eight 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 $\boldsymbol{\mu}_i$ ($1 \leq i \leq N_h$) 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:

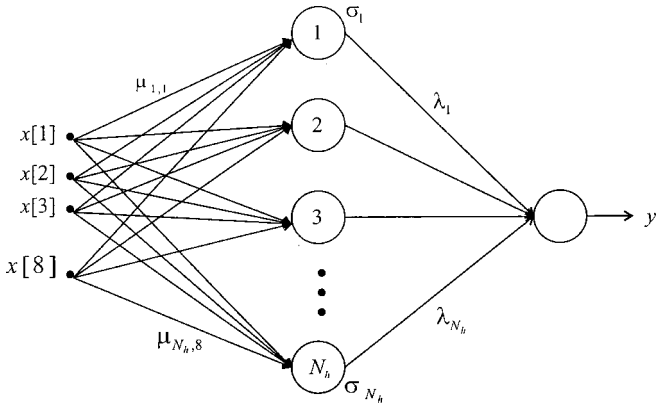

Fig. 10. Topology of radial basis function neural network for traffic pattern classification

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

where $\|\cdot\|$ represents the Euclidean norm of a vector. 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

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

where N is the total number of training instances. Eq. (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

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

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 values of λ_j we solve the following unconstrained optimization problem:

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

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 h 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 nonincident patterns extracted from the simulated data. This sample is not reused for testing. 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 recalibration or retraining 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. In the simulations, the network entry flow rate per lane is varied from 1,000 to 2,000 vph. The actual flow rate at detector stations, however, ranged from 480 to 2,640 vph per lane. The traffic incident consists of the blockage of one lane (the blockages are distributed evenly between the lanes) and a 50% 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 1,000 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 nonincident 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 s, with a range varying from 40 to 54 s. 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 min. 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 s. The upstream pattern is based on the formation of a queue that takes a rather long time to develop (on the order of 1 to 4 min).

In 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 nonincident. This was intended to eliminate 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 three-lane freeway for entry flow rates ranging from 1,250 vph to 2,000 vph per

Table 1. Performance of New Incident Detection Model on Two-Lane Freeway

Flow rate (vph per lane)	Location (m)[a]	Downstream station[b]			Upstream station[b]		
		Detections	False Alarms	Detection Time (s)	Detections	False Alarms	Detection Time (s)
1,000	244	10/10	1/40	50	10/10	0/40	192
1,000	122	10/10	0/40	40	10/10	0/40	142
1,100	244	10/10	0/40	40	10/10	0/40	228
1,100	122	10/10	0/40	40	10/10	0/40	126
1,250	244	10/10	0/40	48	10/10	0/40	172
1,250	122	10/10	0/40	46	10/10	0/40	110
1,500	244	10/10	0/40	48	10/10	0/40	130
1,500	122	10/10	0/40	48	10/10	0/40	82
1,750	244	10/10	0/40	44	10/10	0/40	114
1,750	122	10/10	0/40	48	10/10	0/40	70
2,000	244	10/10	0/40	54	10/10	0/40	88
2,000	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%)	

[a]Distance of the traffic incident from the upstream station. Distance between stations is 460 m.

[b]Numbers after indicate the total number of simulations.

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 four incidents during low demand conditions. The missed detections by the upstream detection station are understandable because the remaining capacity (about 4,000 vph) is still able to handle the demand (3,750 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 1 min 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

A compression wave in a traffic stream is characterized by a pattern of increased occupancy and flow rate that moves in the direction of flow and lasts for a few minutes at any given location (usually less than 5 min). Compression waves appear as a clustering of vehicles within the traffic stream. They are a major source of false alarms generated by automated incident detection algorithms. To test the model's performance under compression wave-like conditions, 100 min of data are generated for a two-lane freeway with a moderate flow rate of 1,500 vph per lane and with several periods of increased flow rate up to 500 vph. The periods of increased flow rate are limited to 5 min or less based on the assumption that compression waves are temporary conditions. A typical 25-min plot of lane occupancy is shown in Fig. 11. The higher flow rate period lasts from 600 to 900 s. In all, there are 374 patterns in this 100-min of data. The model correctly identified all of them as nonincident conditions.

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

The freeway service patrol 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 mul-

Table 2. Performance of New Incident Detection Model on Three-Lane Freeway

Flow rate (vph per lane)	Location (m)[a]	Downstream station[b]			Upstream station[b]		
		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%)	

[a]Distance of the traffic incident from the upstream station. Distance between stations is 460 m.

[b]Numbers after / indicate the total number of simulations.

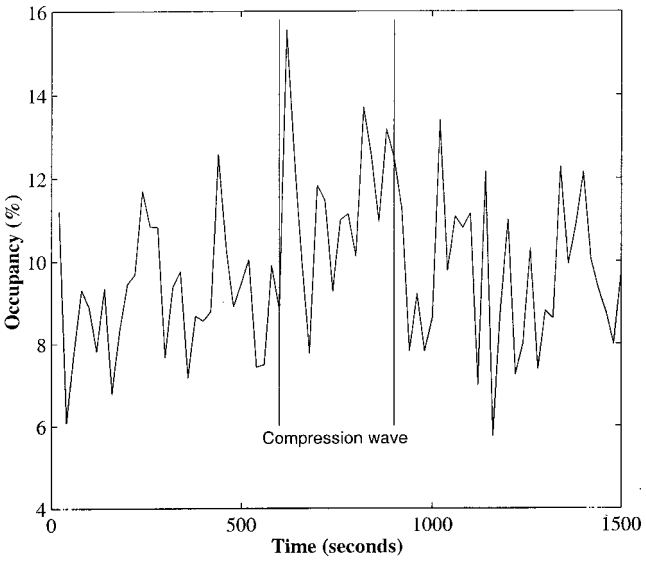

Fig. 11. Typical lane occupancy time-series plot for compression wave traffic condition

tiple 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 h 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% 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. The times recorded in the database are the times at which a patrol vehicle first encountered the incident, and in most cases, this time is after the effects of the incident became visible in data recorded at the nearest detector station. Thus, the detection times—which by definition requires the exact time at which an incident occurred—cannot be computed for the model.

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 speed data are available, the upstream detector station logic provides an additional level of reliability without any significant increase in computation. In general, the downstream station logic will signal an incident before it is signaled by the upstream station logic, thus making the upstream logic a backup in situations where a detector failure has occurred. 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 RB-FNN 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 because it does not require the computation of the inverse wavelet transform and the fuzzy *c*-mean at each time interval.

Table 3. Performance of New Incident Detection Model using Real Data from Freeway Service Patrol (FSP) Project's Database

Downstream station		Upstream station	
Detections	False Alarms	Detections	False Alarms
20/21 (95.2%)	0/480 (0%)	20/21 (95.2%)	0/480 (0%)

Note: Numbers after "/" indicate the total number of tests.

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 verified based on data obtained from the upstream detector station only. Wavelet domain processing is used to denoise, compress, and enhance the raw traffic data for classification. It is found that an energy representation of the data best characterizes incident and nonincident 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 nonincident 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 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 for nonincident conditions without triggering a false alarm;
- The model signals the presence of an incident within 1 min of its occurrence, to a great extent independent of the prevailing traffic and roadway conditions;
- The model does not require recalibration for its on-line implementation and thus is readily transferable; and
- The model is computationally highly efficient because DWT operations require a small number of multiplications and additions in every sensor reporting interval (say, 20 s) and we have reduced the dimensionality of the RBF neural network through wavelet-based energy representation of input.

These characteristics make this 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.

Acknowledgments

This work was based on a research project sponsored by the Ohio Dept. of Transportation and the Federal Highway Administration.

References

Adeli, H., and Hung, S. L. (1995). *Machine learning—neural networks, genetic algorithms, and fuzzy systems*, Wiley, New York.

Adeli, H., and Karim, A. (2000). ''A fuzzy-wavelet RBFNN model for freeway incident detection.'' *J. Transp. Eng.*, 126(6), 464–471.

Ahmed, S. A., and Cook, A. R. (1982). ''Application of time-series analysis techniques to freeway incident detection.'' *Transp. Res. Rec.*, 841, Transportation Research Board, Washington, D.C., 19–28.

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*, Plenum, New York.

Burrus, C. S., Gopinath, R. A., and Guo, H. (1998). *Introduction to wavelets and wavelet transforms—A primer*, Prentice-Hall, Upper Saddle River, N.J.

Cannon, R. L., Dave, J. V., and Bezdek, J. C. (1986). ''Efficient implementation of the fuzzy *c*-means clustering algorithms.'' *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-8(2), 248–255.

Chassiakos, A. P., and Stephanedes, Y. J. (1993). ''Smoothing algorithms for incident detection.'' *Transp. Res. Rec.*, 1394, Transportation Research Board, Washington, D.C., 8–16.

Cheu, R. L., and Ritchie, S. G. (1995). ''Automated detection of lane-blocking freeway incidents using artificial neural networks.'' *Transp. Res., Part C* 3(6), 371–388.

Cook, A. R., and Cleveland, D. E. (1974). ''Detection of freeway capacity-reducing incidents by traffic stream measurements.'' *Transp. Res. Rec.*, 495, Transportation Research Board, Washington, D.C., 1–11.

Daubechies, I. (1992). *Ten lectures on wavelets*, SIAM, Philadelphia.

Dia, H., and Rose, G. (1997). ''Development and evaluation of neural network freeway incident detection models using field data.'' *Transp. Res., Part C*, 5(5), 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.'' *J. Transp. Eng.*, 120(5), 753–772.

Ishak, S. S., and Al-Deek, H. M. (1998). ''Fuzzy ART neural network model for automated detection of freeway incidents.'' *Transp. Res. Rec.*, 1634, Transportation Research Board, Washington, D.C., 56–63.

Lin, C.-K., and Chang, G.-L. (1998). ''Development of a fuzzy-expert system for incident detection and classification.'' *Math. Comput. Modell.*, 27(9–11), 9–25.

Lin, W.-H., and Daganzo, C. F. (1997). ''A simple detection scheme for delay-inducing freeway incidents.'' *Transp. Res., Part A*, 31(2), 141–155.

Moody, J., and Darken, C. J. (1989). ''Fast learning in networks of locally-tuned processing units.'' *Neural Comput.*, 1, 281–294.

Payne, H. J., and Tignor, S. C. (1978). ''Freeway incident-detection algorithms based on decision trees with states.'' *Transp. Res. Rec.*, 682, Transportation Research Board, Washington, D.C., 30–37.

Persaud, B. N., and Hall, F. L. (1989). ''Catastrophe theory and patterns in 30-second freeway traffic data—implications for incident detection.'' *Transp. Res., Part A*, 23A(2), 103–113.

Samant, A., and Adeli, H. (2000). ''Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis.'' *Comput.-Aided Civ. Infrastruct. Eng.*, 15(4).

Weil, R., Wootton, J., and Garcia-Ortiz, A. (1998). ''Traffic incident detection: Sensors and algorithms.'' *Math. Comput. Modell.*, 27(9–11), 257–291.

Xu, H., Kwan, C. M., Haynes, L., and Pryor, J. D. (1998). ''Real-time adaptive on-line traffic incident detection.'' *Fuzzy Sets Syst.*, 93, 173–183.