Fuzzy-Wavelet RBFNN Model for Freeway Incident Detection

By Hojjat Adeli[1] and Asim Karim[2]

(Reviewed by the Urban Transportation Division)

Abstract: Traffic incidents are nonrecurrent and pseudorandom 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 the systemwide effect of incidents is most severe. Model-based solutions to the incident detection problem have not produced practical, useful results primarily because the complexity of the problem does not lend itself to accurate mathematical and knowledge-based representations. A new multiparadigm 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 denoising 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 (RBFNN) is developed to classify the denoised and clustered observed data. The new model produced excellent incident detection rates with no false alarms when tested using both real and simulated data.

INTRODUCTION

According to one estimate, about 60% 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.

In the United States 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 >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 3 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.

This article presents 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. This article presents an advanced denoising technique based on wavelet theory to overcome this problem and improve the efficiency and effectiveness of the algorithm.

[1]Prof., Dept. of Civ. and Envir. Engrg. and Geodetic Sci., Ohio State Univ., Columbus, OH 43210.

[2]Grad. Res. Assoc., Dept. of Civ. and Envir. Engrg. and Geodetic Sci., Ohio State Univ., Columbus, OH.

Note. Discussion open until May 1, 2001. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on September 3, 1999. This paper is part of the ***Journal of Transportation Engineering***, Vol. 126, No. 6, November/December, 2000. ©ASCE, ISSN 0733-947X/00/0006-0464–0471/$8.00 + $.50 per page. Paper No. 21803.

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 a 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 it 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 multidetector, 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 (Cook and Cleveland 1974; Dudek et al. 1974; Ahmed and Cook 1982; Stephanedes and Chassiakos 1993). Dudek et al. (1974) used the standard normal deviate of the data in their threshold-based algorithm, whereas Cook and Cleveland (1974) proposed the use of double exponential smoothing of traffic data in a similar algorithm logic. Ahmed and Cook (1982) presented 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 are then employed in the algorithm logic for incident detection.

More recently research has concentrated on model-free intelligent system 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) proposed 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 described the development of the fuzzy rules but presented no tested implementation of the algorithm. Weil et al. (1998) proposed 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 (ANNs) are powerful pattern recognizers and classifiers (Adeli and Hung 1995; Adeli and Park 1998). 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—multilayer perceptron, self-organizing feature map, and adaptive resonance theory Model 2—are investigated and compared with three common conventional algorithms using simulated data. Dia and Rose (1997) used field data to test a multilayer perceptron ANN as an incident detection classifier. Amin et al. (1998) proposed 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) presented a hybrid fuzzy logic–neural-network approach for the solution of the traffic incident detection problem. They used 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) used the fuzzy cerebral model arithmetic computer 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 artificial intelligence techniques and a multiparadigm 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% 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, which 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 the known behavior of traffic flow completely. This 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. The following paragraphs briefly introduce the mathematics of wavelet and multiresolution analysis.

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

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

where $\{w_{j,k}\}$ = set of coefficients corresponding to the expansion set $\{\psi_{j,k}\}$; and Z = space of integers. A 2D 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. Eq. (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:

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

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 (2) is done in powers of 2. For this dyadic formulation (2) can be rewritten as

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

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

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

Eq. (1) with the coefficients given by (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 when x is a continuous-time function. In the following discussion it is assumed that the signal is a discrete-time function and (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:

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

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

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

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

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

and there exists 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 (3). In (5)–(9), $V_0 \subset V_1$ means that V_0 is a subspace of V_1, $\cup$ represents the union of spaces, $\cap$ represents the intersection of spaces, the overbar denotes the closure of the space, $L^2(R)$ is the space of all square integrable functions of real variables, and $\Rightarrow$ and $\Leftrightarrow$ stand for one-way and two-way implications, respectively.

If (5)–(9) hold, then there exists a set of functions $\psi_{j,k}$ [(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 that, 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

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

where the first term is a coarse resolution at scale j_0 and the second term adds details of increasing resolutions. Eq. (12) also can 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 Fig. 1. Note that spaces spanned by different scales of wavelets are orthogonal to each other because they do not overlap (nonoverlapping functions are always orthogonal).

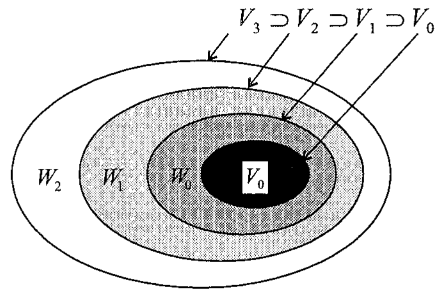


FIG. 1. Multiresolution Function Space Decomposition Using Wavelet Analysis

Computation of DWT

In practical wavelet analysis of discrete signals one usually does not have to deal with the functions themselves but instead works 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, (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. Eqs. (13) and (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 (12). This work uses Daubechies's 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, occupancy, and flow rate values at a given location every 20 to 30 s. 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 no-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 no-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 noncomparative approach is adopted in this research. This decision is based on the analysis of patterns on both the upstream and downstream sides of a 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 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 [Figs. 2(a and 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. Numerical experiments indicate $N = 16$ provides accurate results and is therefore used

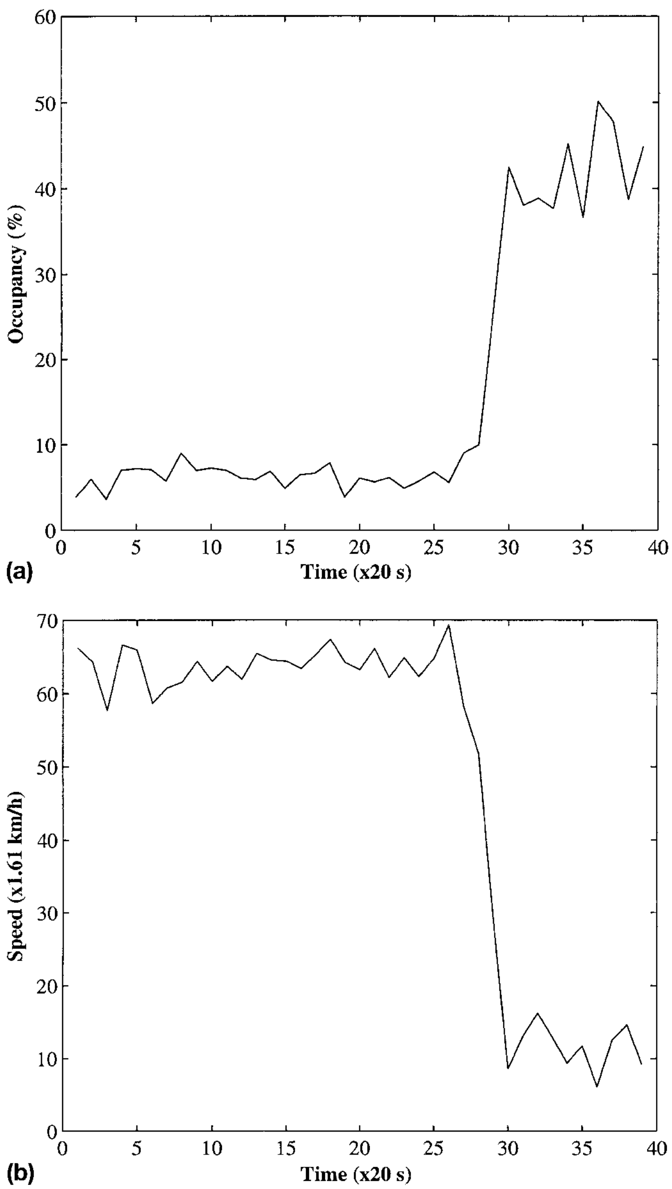


FIG. 2. Typical Time Histories Upstream of Incident (a) Occupancy Plot; (b) Speed Plot

in the model. The 16 data points constitute 5 min and 20 s of data, if data are obtained every 20 s. 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, whereas 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 $n = 1$ to 16.

WAVELET-BASED DENOISING

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 denoised 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 to produce the best results.

In recent years, formal wavelet-based denoising techniques have been presented in the literature (Donoho 1993, 1995; Polchlopek and Noonan 1997). These techniques perform a nonlinear filtering on the transformed signal, modifying the wavelet coefficients in such a way that the inverse transformation yields a denoised 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 denoised DWT of the original signal. To denoise 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 four 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 $(\cdot)^+$ is equal to $(\cdot)$ when $(\cdot)$ is positive and zero otherwise and the function sgn$(\cdot)$ 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 denoised 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-mean (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:

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

Subject to:

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

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

where J_β = objective function for a given value of β; A_{ij} = membership grade of vector i in class j; and $\|\cdot\|$ 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 nonconvex 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 one uses the following membership grade function based on the euclidean norm (Bezdek 1981):

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

where the superscript t denotes the iteration number.

To cluster the denoised data sequences $\bar{x}_O[n]$ and $\bar{x}_S[n]$ one defines 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

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

and uses the FCM algorithm in the following form:

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

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

3. Calculate the updated membership grade using (18).
4. If the maximum change in the membership grade is $<\varepsilon$, or

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

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 <10 iterations and the dimensionality of the data is small.

RADIAL BASIS FUNCTION NEURAL NETWORK (RBFNN) CLASSIFIER

The RBFNN learns an input-output mapping by covering the input space with basis functions that transform 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.

Fig. 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 this test example), and the dimension of each cluster (equal to 2, when occupancy and speed is used as in this 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 $\boldsymbol{\mu}_i$ ($1 \le i \le N_c$) are 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

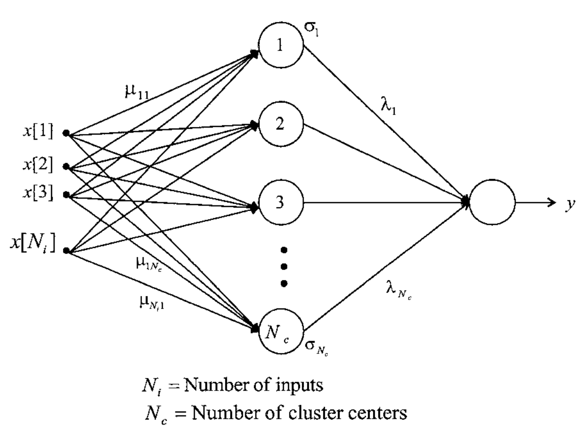


FIG. 3. RBFNN for Discriminating Incident and No-Incident Patterns

the cluster center $\boldsymbol{\mu}_j$. This work uses 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 = weight of the link from the hidden node j to the output node. The output value of 1 corresponds to an incident classification, whereas a value of -1 corresponds to a no-incident classification.

The variables λ_j and μ_{ji} 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 one solves the following unconstrained optimization problem:

$$\min E(\lambda) = \sum_{i=1}^{N_c} |y^i - y_d^i| \tag{24}$$

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

The spread parameters σ_j also can be treated as variables. However, one finds 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 preassign the value of σ_j:

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

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 are generated from the simulation package Traffic Software Integrated System (TSIS) (available via the Internet at ⟨http://www.fhwa-tsis.com⟩). The 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 are 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 are 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 (1,500–2,500 ft) apart. A total of 150 800-s simulations were performed with data obtained in 20-s intervals. Ninety of these simulations involve a traffic incident, whereas the remaining 60 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 2,000–2,500 vehicles/h. Low demand conditions are adopted for evaluation because these are the conditions under which currently available incident detection algorithms perform poorly.

Thirty incident and 30 no-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-s simulations such that the effects of the blockage are pronounced during the last few values of the sample. Fig. 4 shows the normalized occupancy plots for two simulation runs. Fig. 4(a) is for an incident 244-m downstream of the detector station, and Fig. 4(b) is for an incident 122-m downstream of the detector station. Fig. 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 <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

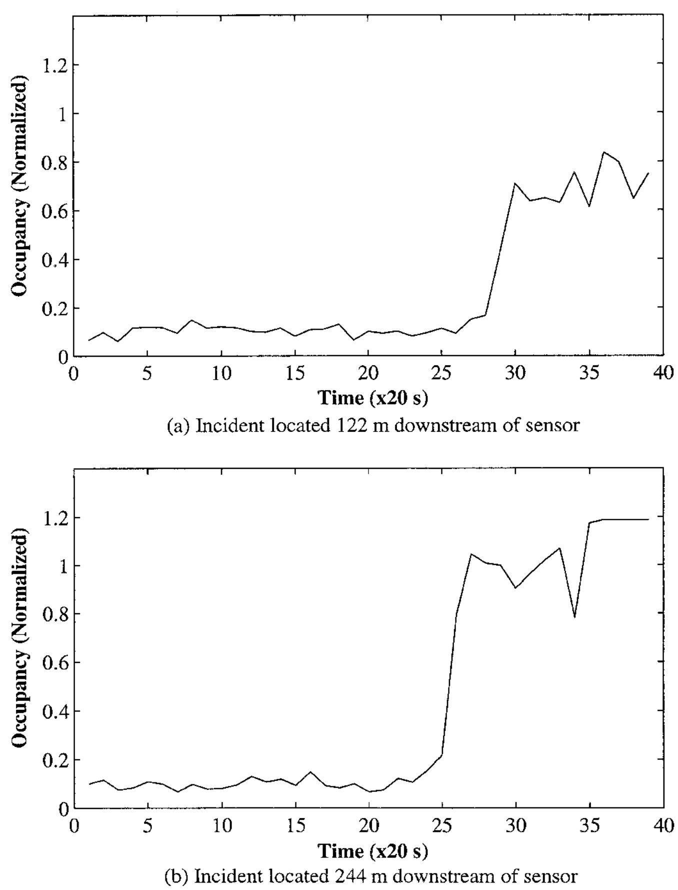
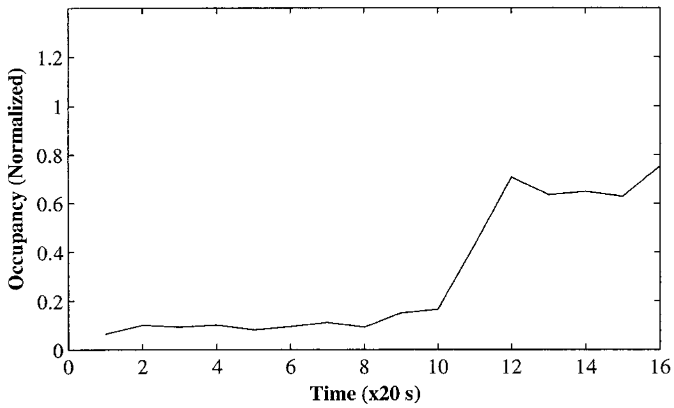
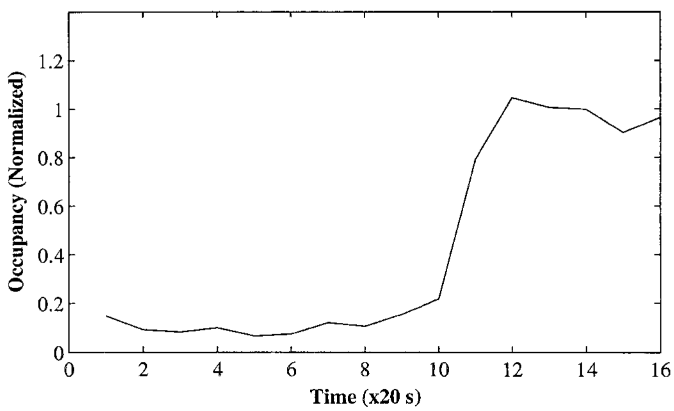


FIG. 4. Normalized Occupancy Plots Obtained from Simulating Traffic on Two-Lane Freeway (Incident Occurs at Time 400 s)



(a) Incident located 122 m downstream of sensor



(b) Incident located 244 m downstream of sensor

FIG. 5. Occupancy Incident Patterns Extracted from Simulations Presented in Fig. 4

to 0.3 is classified as an incident. Otherwise, it is classified as no incident. The model is tested using the simulated data by presenting each of the 90 800-s simulations as a continuous stream of data. An output is produced every 20 s after the first 320 s (16 data points). An incident is detected when the output becomes greater than the threshold for the first time. All 60 incidents were detected correctly during the testing of the model. Therefore, the detection rate is 100%. Also, none of the no-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. Fig. 6 shows the variation of the mean detection time of the algorithm with preincident 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-mi) long segment of the freeway from Oakland to San Jose, Calif. 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 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 are processed to extract 21 incidents that block one or more lanes. The loop detector data are averaged over a 30-s time interval. The incident detection model detected 20 of the 21 incidents, resulting in a detection rate of 95.2%. The traffic pattern corresponding to the missed incident did not exhibit the characteristics of an incident condition. This appears to be an error

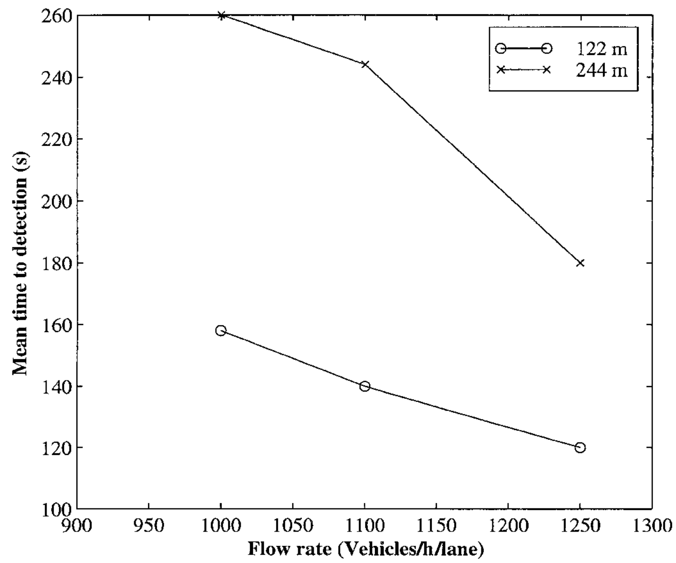


FIG. 6. Mean Detection Time of Incident as Function of Flow Rate and Distance from Upstream Sensor

in the incident data. The incident data, in general, are not accurate, as the location of incidents are reported approximately (like 1 mi from an 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 these 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. The new incident detection model correctly identified all 30 patterns as no-incident patterns. Thus, the false alarm rate is zero.

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

CONCLUSIONS

A new multiparadigm 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 denoising 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. An RBFNN is developed to classify the denoised 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 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% and the false alarm rate of zero. Real traffic data were 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 4 h of incident-free data.

The methodology presented provides a solid function for further research and development. The writers 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 the Federal Highway Administration, Washington, D.C.

APPENDIX. REFERENCES

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

Adeli, H., and Park, H. S. (1998). *Neurocomputing for design automation*, CRC, Boca Raton, Fla.

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.

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.'' *Comp.-Aided Civ. and Infrastructure Engrg.*, 13, 315–325.

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. on Pattern Anal. and Machine Intelligence*, 8(2), 248–255.

Chang, E. C., and Wang, S.-H. (1994). ''Improved freeway incident detection using fuzzy set theory.'' *Transp. Res. Rec. 1453*, Transportation Research Board, Washington, D.C., 75–82.

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

Cottrell, W. D. (1998). ''Estimating the probability of freeway congestion.'' *Transp. Res. Rec. 1634*, Transportation Research Board, Washington, D.C., 19–27.

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

Dave, R. N., and Krishnapuram, R. (1997). ''Robust clustering methods: A unified view.'' *IEEE Trans. on Fuzzy Sys.*, 5(2), 270–293.

Dia, H., and Rose, G. (1997). ''Development and evaluation of neural network freeway incident detection models using field data.'' *Transp. Res.*, 5C(5), 313–331.

Donoho, D. L. (1993). ''Nonlinear wavelet methods for recovery of signals, images, and densities from noisy and incomplete data.'' *Different perspectives on wavelets*, I. Daubechies, ed., American Mathematical Society, Providence, R.I.

Donoho, D. L. (1995). ''De-noising by soft thresholding.'' *IEEE Trans. on Information Theory*, 41, 613–627.

Dudek, C. L., Messer, C. J., and Nuckles, N. B. (1974). ''Incident detection on urban freeways.'' *Transp. Res. Rec. 495*, Transportation Research Board, Washington, D.C., 12–24.

Geng, J., and Lee, T. N. (1998). ''Freeway traffic incident detection using fuzzy CMAC neural networks.'' *Proc., IEEE Int. Conf. on Fuzzy Sys.*, Vol. 2, Piscataway, N.J., 1164–1169.

Hsiao, C.-H., Lin, C.-T., and Cassidy, M. (1994). ''Application of fuzzy logic and neural networks to automatically detect freeway traffic incidents.'' *J. Transp. Engrg.*, ASCE, 120(5), 753–772.

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

Lindley, J. A. (1987). ''Urban freeway congestion: Quantification of the problem and effectiveness of potential solutions.'' *ITE J.*, 57(1), 27–32.

Mallat, S. G. (1989). ''A theory for multiresolution signal decomposition: The wavelet representation.'' *IEEE Trans. on Pattern Anal. and Machine Intelligence*, 11(7), 674–693.

Moody, J., and Darken, C. J. (1989). ''Fast learning in networks of locally-tuned processing units.'' *Neural Computation*, 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.*, 23A(2), 103–113.

Persaud, B. N., Hall, F. L., and Hall, L. M. (1990). ''Congestion identification aspects of the McMaster incident detection algorithm.'' *Transp. Res. Rec. 1287*, Transportation Research Board, Washington, D.C., 167–175.

Poggio, T., and Girosi, F. (1990). ''Networks for approximation and learning.'' *Proc., IEEE*, 78, 1481–1497.

Polchlopek, H. M., and Noonan, J. P. (1997). ''Wavelets, detection, estimation, and sparsity.'' *Digital Signal Processing*, 7, 28–36.

Samant, A., and Adeli, H. (2000). ''Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis.'' *Comp.-Aided Civ. and Infrastructure Engrg.*, 15(4), 241–250.

Stephanedes, Y. J., and Chassiakos, A. P. (1993). ''Application of filtering techniques for incident detection.'' *J. Transp. Engrg.*, ASCE, 119(1), 13–26.

Al-Sultan, K. S., and Fediki, C. A. (1997). ''A Tabu search-based algorithm for the fuzzy clustering problem.'' *Pattern Recognition*, 30(12), 2023–2030.

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

Zadeh, L. A. (1978). ''Fuzzy set as a basis for a theory of possibility.'' *Fuzzy Sets and Sys.*, 1(1), 3–28.