Stochastic Optimal-Control Approach to Automatic Incident-Responsive Coordinated Ramp Control

Jiuh-Biing Sheu, *Member, IEEE*, and Mei-Shiang Chang

***Abstract*—This paper presents a stochastic optimal-control-based approach to real-time incident-responsive coordinated ramp control. The proposed coordinated ramp-control methodology includes two major functions, i.e., 1) identification of control-zone-based congestion patterns using the modified generalized-sequential-probability-ratio-testing technology and 2) group-based ramp control using stochastic optimal control coupled with extended Kalman filtering technologies. With the aid of the Paramics microscopic traffic simulator, numerical studies under various simulated incident-induced congestion conditions were conducted. Corresponding numerical results indicate the feasibility of the proposed ramp-control method in responding to diverse lane-blocking incident-induced traffic-congestion conditions.**

***Index Terms*—Coordinated ramp controls, incident management, stochastic optimal control.**

I. Introduction

Incident-induced traffic congestion remains a critical operational problem for freeway congestion management. As is widely recognized, lane-blocking incidents may rapidly cause temporal and spatial traffic phenomena, such as unexpected increases in delays and queue lengths. They can also contribute to more serious overcongestion problems, e.g., link queue overflows and network-wide gridlocks, particularly under high-volume lane-blocking-incident conditions. In addition, according to previous studies [1], [2], the main issue of real-time incident-induced congestion management is mainly rooted in the difficulty of characterizing in real time the variety of incident effects on time-varying traffic flows, either temporally or spatially. Consequently, efficiently alleviating freeway nonrecurrent congestion remains a challenge, despite the aid of intelligent-transportation-systems-related technologies.

Numerous ramp-metering technologies have emerged over the past decades, for either local ramp control [3]–[6] or coordinated ramp control [7]–[9]. An overview of ramp-metering algorithms can also be readily found in the study in [10], ranging from early fixed-time approaches to traffic-responsive regulators and to modern sophisticated nonlinear optimal-control schemes. Nevertheless, efficient solutions for freeway incident-induced congestion management warrant further in-depth investigations.

Accordingly, the purpose of this paper is to investigate the applicability of a stochastic optimal-control-based approach for real-time incident-responsive coordinated ramp control. In the proposed methodology, using the modified generalized-sequential-probability-ratio-testing (M-GSPRT) technology, traffic-congestion patterns on multiple mainline segments of a freeway are identified and grouped in real time. Then, a discrete-time nonlinear stochastic optimal-control-based model, coupled with extended Kalman filtering technologies, are employed to determine the group-based ramp-metering rates aiming to minimize the incident impacts on the corridor-wide traffic flows.

Manuscript received November 3, 2006; revised November 26, 2006 and February 18, 2007. This work was supported in part by the National Science Council, Taiwan, R.O.C., under Grant NSC 94-2416-H-009-009. The Associate Editor for this paper was S. Tang.

J.-B. Sheu is with the Institute of Traffic and Transportation, National Chiao Tung University, Taipei 10012, Taiwan, R.O.C. (e-mail: jbsheu@mail.nctu.edu.tw).

M.-S. Chang is with the Department of Civil Engineering, Chung Yuan Christian University, Chung Li 32023, Taiwan, R.O.C. (e-mail: mschang@cycu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2007.893327

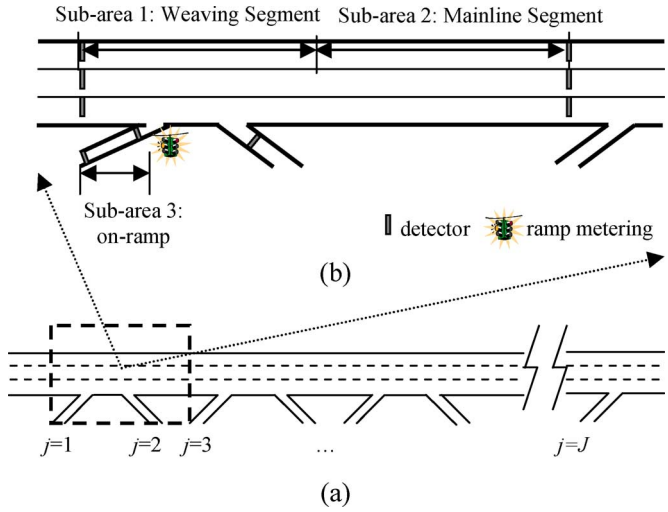

Fig. 1. Geographic specification of system scope. (a) Freeway corridor. (b) Local control zone.

II. System Specification

The system scope studied here considers any given freeway corridor, which is composed geometrically of multiple ramps and freeway mainline segments, as shown in Fig. 1. To collect real-time raw traffic data for the input of the proposed method, specific point detector layouts associated with local control zones of a given freeway corridor are proposed. Here, a local control zone refers to the area dominated by a given local ramp control, and the local control zone involves three subareas for the convenience of characterizing traffic dynamics in three corresponding subsections of the given control zone.

With the system scope and detector layouts specified above, a two-stage real-time coordinated ramp-control methodology is proposed. The conceptual framework of the proposed method is presented in Fig. 2. It is structured primarily with two layers: 1) zone-based congestion-pattern recognition and 2) stochastic optimization of coordinated ramp-metering control. The main functionality, which is provided in the first layer, is a grouping of control-zone-based congestion patterns for further coordinated ramp control. Herein, according to the measurements of specified decision variables, which are denoted later, the time-varying zone-based traffic-congestion patterns, including incident-induced and incident-free patterns, are identified and classified into several ramp-control groups via the proposed M-GSPRT algorithm executed in the first layer. Then, the second layer triggers a stochastic optimal-control-based algorithm to determine real-time group-based ramp meters, where those ramps assigned in the same zone group utilize the same ramp-metering rate estimated in the second layer to deal with the common group-based congestion problem.

Furthermore, three groups of system variables, including 1) decision variables; 2) state variables; and 3) control variables, are specified in order to facilitate model formulation in the following sections. The corresponding notations of these variables, as well as their utilization in the proposed method, are described below.

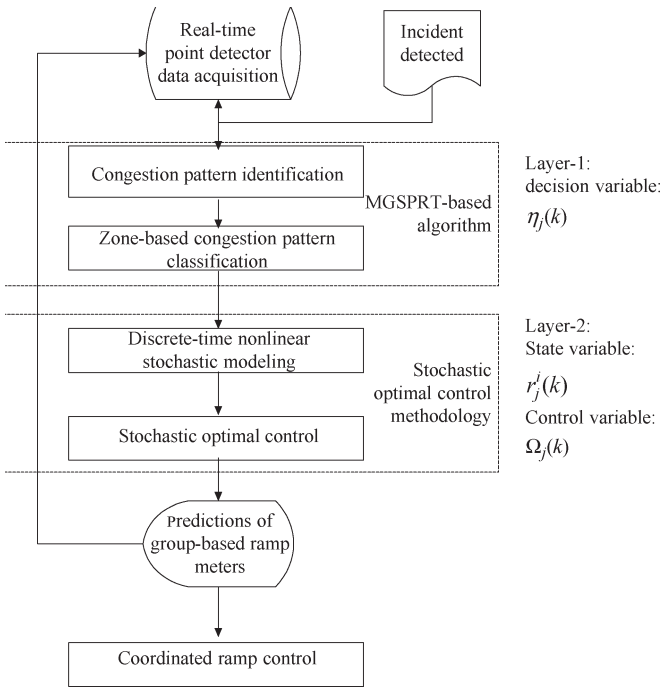

Fig. 2. Conceptual framework of the proposed real-time coordinated ramp-control system.

A. Decision Variables

Decision variables are used for the identification of congestion patterns performed by the proposed M-GSPRT-based algorithm executed in the first layer of the proposed method. Here, each given local control-zone j is associated with a time-varying decision variable $\eta_j(k)$ denoted by

$$\eta_j(k) = \sum_{i=1}^{3} w_j^i(k) \times \overline{q}_j^i(k) \tag{1}$$

where $\overline{q}_j^i(k)$ (in vehicles per lane per meter) represents the time-varying section-based density function, which indicates the congestion severity associated with the subarea i of a given control-zone j measured at the time-step k, and $w_j^i(k)$ is the time-varying weighting value associated with $\overline{q}_j^i(k)$. According to Fig. 1, there are three time-varying section-based density functions, i.e., $\overline{q}_j^1(k), \overline{q}_j^2(k)$, and $\overline{q}_j^3(k)$, associated with the respective subareas of any given control-zone j, and they are given, respectively, by

$$\overline{q}_j^1(k) = \frac{1}{n_j^1 \times L_j^1}\Big[a_j^1(k-1) + \overline{q}_j^1(k-1) \times n_j^1 \times L_j^1 + d_j^3(k-1) - \widetilde{d}_j^3(k-1)\Big] \tag{2}$$

$$\overline{q}_j^2(k) = \frac{1}{n_j^2 \times L_j^2}\Big[\left(\overline{q}_j^1(k-1) \times n_j^1 \times L_j^1\right) \times r_j^1(k-1) + \overline{q}_j^2(k-1) \times n_j^2 \times L_j^2\Big] \tag{3}$$

$$\overline{q}_j^3(k) = \frac{1}{n_j^3 \times L_j^3}\left[a_j^3(k-1) + \overline{q}_j^3(k-1) \times n_j^3 \times L_j^3\right] \tag{4}$$

where $a_j^1(k-1)$ and $a_j^3(k-1)$ represent the numbers of section-based traffic arrivals, i.e., the multilane traffic arrivals, measured by the upstream detector stations located in the subareas 1 and 3 of the given control-zone j at time-step $k-1$, respectively, $d_j^3(k-1)$ and $\widetilde{d}_j^3(k-1)$ represent the numbers of vehicles departing from the onramp (i.e., the subarea 3) and the offramp of the given control-zone j at time-step $k-1$, respectively, both of which can be measured via the detectors installed in the onramp and offramp areas and according to the specified detector layouts shown in Fig. 1, n_j^i and L_j^i represent the number of lanes and the geometric length associated with the subarea i of the given control-zone j, respectively, and $r_j^1(k-1)$ represents the state variable associated with the subarea 1 of a given control-zone j, estimated at the previous time-step $k-1$, which indicates the proportion of the vehicles present in the corresponding subarea that can pass the given subarea at time-step $k-1$. The definitions of state variables are detailed later.

It is noteworthy that $w_j^i(k)$ is herein proposed to dynamically determine the magnitude of the congestion effect associated with $\overline{q}_j^i(k)$ under various traffic-flow conditions in the proposed congestion-pattern-recognition procedure. In some cases, e.g., low-volume traffic-flow conditions on mainline segments of freeways, the relatively high congestion on ramps can be given with the tentatively greater weightiness, representing the greater significance of the onramp congestion effect in calculating the decision variable of congestion-pattern recognition, and vice versa. In addition, the summed value of $w_j^i(k)$ in a given control-zone j at any given time-step k should be unconditionally equal to one. Accordingly, $w_j^i(k)$ is given by

$$w_j^i(k) = \frac{\sum_{\varepsilon=1}^{\Xi} \overline{\theta}_j^{\,i}(k-\varepsilon)}{\sum_{i=1}^{3}\sum_{\varepsilon=1}^{\Xi} \overline{\theta}_j^{\,i}(k-\varepsilon)} \tag{5}$$

where $\overline{\theta}_j^{\,i}(k-\varepsilon)$ represents the historical section-based occupancy measured by averaging the lane-based occupancy data collected from the detector stations located in the given subarea i of the given control-zone j at time-step $k-\varepsilon$, Ξ is the maximum time-lag predetermined for the calculation of $w_j^i(k)$, herein set to be five consistent with the value used for real-time incident characterization in the previous research [11], and ε represents a time-lag index.

B. State Variables

State variables $r_j^i(k)$ are specified to characterize the ease of traffic flowing through a given subarea i of a given control-zone j at a given time-step k, and in the proposed method, they are used to formulate the proposed stochastic model. Accordingly, for each given control-zone j, we have three time-varying state variables, i.e., $r_j^1(k)$, $r_j^2(k)$, and $r_j^3(k)$, given, respectively, by

$$r_j^1(k) = \frac{D_j^1(k)}{a_j^1(k) + Q_j^1(k-1) + d_j^3(k) - \widetilde{d}_j^3(k)} \tag{6}$$

$$r_j^2(k) = \frac{d_j^2(k)}{D_j^1(k) + Q_j^2(k-1)} \tag{7}$$

$$r_j^3(k) = \frac{d_j^3(k)}{a_j^3(k) + Q_j^3(k-1)} \tag{8}$$

where $d_j^2(k)$ represents the number of vehicles departing from subarea 2 of a given control-zone j at time-step k, $D_j^1(k)$ represents the number of vehicles present in subarea 1 of a given control-zone j that pass through the corresponding subarea at time-step k, and $Q_j^1(k-1)$, $Q_j^2(k-1)$, and $Q_j^3(k-1)$ are the numbers of vehicles which are present in subareas 1–3 of a given control-zone j at time-step $k-1$.

TABLE I
CLASSIFICATION OF ZONE-BASED CONGESTION PATTERNS

Type of Congestion Patterns by Mainline Flow Condition	By On-ramp Flow Condition		
	High -volume	Medium -volume	Low -volume
high-volume/non-incident	type-1	type-2	type-3
high-volume/incident	type-10	type-11	type-12
medium-volume/non-incident	type-4	type-5	type-6
medium-volume/incident	type-13	type-14	type-15
low-volume/non-incident	type-7	type-8	type-9
low-volume/incident	type-16	type-17	type-18

C. Control Variables

The time-varying control variables $\Omega_j(k)$ are introduced to determine the length of the onramp green-time interval associated with each given control-zone j at time-step k in the process of real-time ramp optimal control and is given by

$$\Omega_j(k) = \frac{G_j(k)}{t} \tag{9}$$

where $G_j(k)$ represents the length of the corresponding green time (in seconds) associated with the given control-zone j embedded in time-step k, and t represents the length of a unit time step (in seconds). In the proposed method, the time-varying ramp-control variable is calculated in real time using the proposed stochastic optimal-control-based algorithm in response to the variety of incident-induced traffic congestion at each time step.

III. ZONE-BASED CONGESTION-PATTERN RECOGNITION

Given the assumption that all the aforementioned time-varying decision variables $\eta_j(k)$ follow Gaussian processes, the investigated zone-based congestion-pattern-recognition problem can then be treated as a specific multiclass pattern-recognition problem that can be solved readily by M-GSPRT techniques. Accordingly, the major function performed by the proposed M-GSPRT-based algorithm is to identify the respective traffic-congestion pattern associated with each control zone with decision variables estimated in real time. Then, the control zones that have the same congestion pattern are assigned to the same group by conducting the proposed specific rules for further coordinated ramp control. Note that, as presented in Table I, 18 types of congestion patterns, including incident-free and incident-induced congestion patterns, are prespecified according to the previous research [2]. These predetermined congestion patterns are conveniently employed in the proposed multizone congestion-pattern-recognition algorithm.

The following depicts the primary procedures of the proposed algorithm.

Step 0) Initialize the following procedures.

Step 0.1) Set the tolerable probability of incorrect recognition $e_{p,\tilde{p}}$ for each predetermined congestion pattern, where $e_{p,\tilde{p}}$ represents the acceptable probability of identifying the time-varying congestion pattern of a given control zone as the congestion-pattern p, given that its actual congestion pattern is $\widetilde{p}$.

Step 0.2) Initialize the iteration index $k = 0$.

Step 0.3) Specify the dynamic threshold associated with a given congestion-pattern $p(g_p(k))$ by

$$g_p(k) = \Psi_p \left(1 - \frac{k}{K}\right)^{r_p} \tag{10}$$

where the parameter r_p, determining the shape of the time-varying threshold, is subject to $0 < r_p \leq 1$, K represents the maximum number of time steps allowed for congestion-pattern recognition, and Ψ_p, which represents the initial value of the dynamic threshold associated with each congestion-pattern p, is given by

$$\Psi_p = \frac{1 - e_{p,p}}{\left[\prod_{\tilde{p}=1}^{\Delta}(1 - e_{p,\tilde{p}})\right]^{1/\Delta}}. \tag{11}$$

In (11), Δ represents the total number of predetermined congestion patterns and, herein, is equal to 18 in this paper.

Step 0.4) Set the hypothesis H_p to be associated with a given congestion-pattern p.

Step 1) Let $k = k + 1$, and target the first unidentified control zone as $j = 1$ for the onset of zone-based congestion-pattern recognition at the current time-step k.

Step 2) Given a control-zone j at time-step k, estimate the time-varying likelihood ratio function associated with each given congestion-pattern $p(\Lambda(\eta_j(k)|H_p))$ by

$$\Lambda\left(\eta_j(k)|H_p\right) = \frac{\prod_{\varepsilon=1}^{k} P\left(\eta_j(\varepsilon)|H_p\right)}{\left[\prod_{\tilde{p}=1}^{\Delta}\prod_{\varepsilon=1}^{k} P\left(\eta_j(\varepsilon)|H_{\tilde{p}}\right)\right]^{1/\Delta}} \tag{12}$$

where $p(\eta_j(k)|H_p)$ represents the conditional probability density function of $\eta_j(k)$ given hypothesis H_p and is calculated using the corresponding Gaussian process assumed previously. It is also noted that the time-varying value of $\eta_j(k)$ is estimated using the collected raw traffic data, according to the definition of the decision variable.

Step 3) Compare the estimated $\Lambda(\eta_j(k)|H_p)$ with the associated dynamic threshold $g_p(k)$ for all the presumed congestion-pattern hypotheses using the following rules.

IF

the congestion hypothesis H_p is the only one to be considered, then the congestion pattern of control-zone j is identified as the corresponding pattern p, and meanwhile, drop the control-zone j from the list of control zones targeted for congestion-pattern recognition in the following time steps.

ELSE IF

$$\Lambda\left(\eta_j fs(k)|H_p\right) < g_p(k) \tag{13}$$

the congestion hypothesis H_p is dropped from consideration for further computation. Correspondingly, the given control-zone j will not be assigned to the control-zone group associated with the congestion-pattern p in the following procedures.

ELSE

the corresponding status means that the congestion pattern of the given control-zone j cannot be identified at the current time step because the objective of the minimum decision cost has not been achieved at the current time-step k. Therefore, the next unidentified downstream control-zone $j + 1$ is taken into

account to continue the congestion-pattern recognition for the current time-step k.

THEN

let $j = j + 1$, and go to Step 2) to continue the classification for the next unidentified downstream control-zone $j + 1$ if there is any unidentified control zone that has not been processed at the current time-step k. Otherwise, go back to Step 1) to continue the classification for the next time-step $k + 1$ until the maximum number of time steps allowed for congestion-pattern recognition K is reached.

Step 4) Let $\hat{p}_j^K$ represent the identified congestion-pattern p associated with a given control-zone j during the time interval of K, scan the identified congestion patterns of all the control zones by starting from the first upstream control zone, i.e., given $j = 1$, and then conduct the following rule for grouping these control zones.

IF

$$\hat{p}_j^K = \hat{p}_{j+1}^K \tag{14}$$

the control-zone $j + 1$ is then assigned to the same group as that of the control-zone j. Otherwise, the control-zones j and $j + 1$ are regarded as being in two different control-zone groups.

THEN

let $j = j + 1$, and continue the above grouping rule until the last control zone is assigned.

According to the fundamentals of either M-GSPRT or MSPRT [12], the predetermined parameter K can be used to control the maximum number of time steps for final decision making. We know that the congestion pattern of any given control zone can be identified before K is reached in the proposed algorithm.

IV. Stochastic Optimization of Coordinated Ramp Control

This section presents a discrete-time nonlinear stochastic model to formulate the relationships among time-varying state variables, control variables, and measured raw traffic data under real-time coordinated ramp control. In addition, using extended Kalman filtering technologies, a stochastic optimal-control-based algorithm is proposed to estimate these time-varying state variables and control variables, as well as to determine real-time group-based ramp meters. Those ramps assigned in the same zone group utilize the same ramp-metering rate. To simplify the model presentation, the notations of all the variables defined in this section are summarized in Section IV-A. Then, Sections IV-B and C describe the main components of the proposed model and the corresponding algorithm, respectively.

A. Notations

$d_j^2(k)$	Number of vehicles departing from the subarea 2 of a given control-zone j at time-step k.
$D_j^1(k)$	Number of vehicles present in subarea 1 of a given control-zone j that pass through the corresponding subarea at time-step k.
$\mathbf{f}_{j_{g^K}}(k)$	Time-varying state vector associated with a given control-zone j_{g^K} of a given control-zone group g^K at time-step k.
$\mathbf{F}[\mathbf{r}(k), \mathbf{\Omega}(k), k]$	Time-varying state vector which involves the subvectors of time-varying state vector $\mathbf{r}(k)$ and corresponding control vector $\mathbf{\Omega}(k)$ at time-step k.
$\mathbf{h}_{j_{g^K}}(k)$	Time-varying measurement-component vector associated with a given control-zone j_{g^K} of a given control-zone group g^K at time-step k.
$\mathbf{H}[\mathbf{r}(k), k]$	Time-varying measurement-component vector in which each element is associated with a given element shown in $\mathbf{Z}(k)$, characterizing the components of the corresponding element of $\mathbf{Z}(k)$ with the state variables and collected upstream traffic counts at time-step k.
$\mathbf{L}[\mathbf{r}(k), \mathbf{\Omega}(k), k]$	Diagonal state-dependent noise matrix.
n_θ	Maximum number of sequential time steps which belong to a given cycle θ.
$Q_j^i(k-1)$	Numbers of vehicles present in the subareas i of a given control-zone j at time-step $k-1$.
$G_j(k)$	Green time associated with the given control-zone j at time-step k (in seconds).
$\mathbf{l}_{j_{g^K}}(k)$	Time-varying state-dependent noise vector associated with a given control-zone j_{g^K} of a given control-zone group g^K at time-step k.
$r_{j_{g^K}}^i(k+1)$	State variables associated with subareas i of the given control-zone j of a given control-zone group g^K at time-step $k+1$.
t	Length of a unit time step (in seconds).
$T_{\min}$	Minimum onramp green time in a given cycle θ (in seconds).
$\mathbf{V}(k)$	Gaussian noise vector, referring to the measurement errors of the collected traffic counts at time-step k.
$\mathbf{v}_{j_{g^K}}(k)$	Gaussian noise vector referring to the measurement errors associated with a given control-zone j_{g^K} of a given control-zone group g^K at time-step k.
$\mathbf{W}(k)$	State-independent Gaussian noise vector.
$\mathbf{w}_{j_{g^K}}(k)$	Time-varying state-independent noise vector associated with a given control-zone j_{g^K} of a given control-zone group g^K at time-step k.
$\mathbf{X}(k+1)$	Time-varying state vector at time-step $k+1$.
$\mathbf{x}_{j_{g^K}}(k+1)$	Time-varying state vector associated with a given control-zone j_{g^K} of a given control-zone group g^K at time-step $k+1$.
$\mathbf{Z}(k)$	Time-varying measurement vector which is composed of the downstream traffic counts collected in subareas 2 and 3 of each given control zone at time-step k.
$\mathbf{z}_{j_{g^K}}(k)$	Time-varying measurement vector associated with a given control-zone j_{g^K} of a given control-zone group g^K at time-step k.
$\omega_{j_{g^K}}^i(k)$	State-independent noises associated, respectively, with $r_{j_{g^K}}^i(k)$, $i = 1,\ 2,$ and 3.
$\Omega_j(k)$	Green-time split associated with the given control-zone j at time-step k.

B. Model Formulation

The proposed stochastic model is composed of three groups of dynamic equations, including state equations, measurement equations, and state boundaries.

1) State Equations:

$$\mathbf{X}(k+1) = \mathbf{F}\left[\mathbf{r}(k), \mathbf{\Omega}(k), k\right] + \mathbf{L}\left[\mathbf{r}(k), \mathbf{\Omega}(k), k\right] \mathbf{W}(k) \tag{15}$$

where

$$\mathbf{X}(k+1)=\mathbf{col}\Big(\mathbf{x}_{j_{g^K}}(k+1),\quad j_{g^K}=1,2,\ldots,m_{g^K}\ \ \forall g^K\Big) \tag{16}$$

$$\mathbf{x}_{j_{g^K}}(k+1)=\begin{bmatrix} r^1_{j_{g^K}}(k+1)\\ r^2_{j_{g^K}}(k+1)\\ r^3_{j_{g^K}}(k+1)\end{bmatrix} \tag{17}$$

$$\mathbf{F}\left[\mathbf{r}(k),\mathbf{\Omega}(k),k\right]=\mathbf{col}\Big(\mathbf{f}_{j_{g^K}}(k),\qquad j_{g^K}=1,2,\ldots,m_{g^K}\ \ \forall g^K\Big) \tag{18}$$

$$\mathbf{f}_{j_{g^K}}(k)=\begin{bmatrix} r^1_{j_{g^K}}(k)\\ r^2_{j_{g^K}}(k)\\ \Omega_{j_{g^K}}(k)\times r^3_{j_{g^K}}(k)\end{bmatrix} \tag{19}$$

$$\mathbf{L}\left[\mathbf{r}(k),\mathbf{\Omega}(k),k\right]=\mathbf{dia}\Big(\mathbf{l}_{j_{g^K}}(k),\qquad j_{g^K}=1,2,\ldots,m_{g^K}\ \ \forall g^K\Big) \tag{20}$$

$$\mathbf{l}_{j_{g^K}}(k)=\begin{bmatrix} \Omega_{j_{g^K}}(k)\times r^3_{j_{g^K}}(k)\times\Big[1-r^1_{j_{g^K}}(k)\Big]\\ r^1_{j_{g^K}}(k)\times\Big[1-r^2_{j_{g^K}}(k)\Big]\\ \Big[1-\Omega_{j_{g^K}}(k)\times r^3_{j_{g^K}}(k)\Big]\times\Big[1-r^1_{j_{g^K}}(k)\Big]\end{bmatrix} \tag{21}$$

$$\mathbf{W}(k)=\mathbf{col}\Big(\mathbf{w}_{j_{g^K}}(k),\qquad j_{g^K}=1,2,\ldots,m_{g^K}\ \ \forall g^K\Big) \tag{22}$$

$$\mathbf{w}_{j_{g^K}}(k)=\begin{bmatrix} \omega^1_{j_{g^K}}(k)\\ \omega^2_{j_{g^K}}(k)\\ \omega^3_{j_{g^K}}(k)\end{bmatrix}. \tag{23}$$

Equation (15) is the state equation to characterize the changing patterns of time-varying state variables by following the Gaussian–Markov processes. The rationalization of the assumption relies, to a certain extent, on our experience in utilizing the concepts of random walk models to deal with the complexity of the patterns of system states changing in the presence of a lane-blocking incident [1], [2]. Equations (16)–(23) are the relevant definitional equations to the variables in (15). It is noteworthy that the state-independent noises are mainly derived from the change patterns of traffic arrivals at the corresponding subareas in consideration of the phenomenon that the magnitude of the aforementioned state-dependent noises, shown in $\mathbf{l}_{j_{g^K}}(k)$, may also depend to a certain extent on the diversity of new traffic arrivals in the proposed model. Equation (19) denotes that the change pattern of the state variable $r^3_{j_{g^K}}(k)$ is influenced directly by the control variable $\Omega_{j_{g^K}}(k)$, and thus, the associated deterministic term is formulated as shown later in (35). As shown in (21), the equilibrium condition of $r^1_{j_{g^K}}(k)$ could be affected both by the time-varying proportion of onramp vehicles that are expected to enter into the mainline segment, i.e., subarea 1 of the given control-zone j_{g^K} at the next time-step $\Omega_{j_{g^K}}(k)\times r^3_{j_{g^K}}(k)$, and the time-varying proportion of vehicles remaining in subarea 1 at the current time-step $1-r_{j_{g^K}}(k)$. Accordingly, the magnitude of the corresponding disturbance effect is measured by multiplying the aforementioned two terms. Employing similar concepts, the other elements of $\mathbf{l}_{j_{g^K}}(k)$ are then formulated, as shown above.

2) Measurement Equations:

$$\mathbf{Z}(k)=\mathbf{H}\left[\mathbf{r}(k),k\right]+\mathbf{V}(k) \tag{24}$$

where

$$\mathbf{Z}(k)=\mathbf{col}\Big(\mathbf{z}_{j_{g^K}}(k),\qquad j_{g^K}=1,2,\ldots,m_{g^K}\qquad\forall g^K\Big) \tag{25}$$

$$\mathbf{z}_{j_{g^K}}(k)=\begin{bmatrix} z^2_{j_{g^K}}(k)\\ z^3_{j_{g^K}}(k)\end{bmatrix} \tag{26}$$

$$\mathbf{H}(k)=\mathbf{col}\Big(\mathbf{h}_{j_{g^K}}(k),\qquad j_{g^K}=1,2,\ldots,m_{g^K}\qquad\forall g^K\Big) \tag{27}$$

$$\mathbf{h}_{j_{g^K}}(k)=\begin{bmatrix} \Big(\Big(a^1_{j_{g^K}}(k)+Q^1_{j_{g^K}}(k-1)+d^3_{j_{g^K}}(k)-\widetilde{d}^3_{j_{g^K}}\Big)\\ \qquad\times r^1_{j_{g^K}}(k)+Q^2_{j_{g^K}}(k-1)\Big)\times r^2_{j_{g^K}}(k)\\ \Big[a^3_{j_{g^K}}(k)+Q^3_{j_{g^K}}(k-1)\Big]\times r^3_{j_{g^K}}(k)\end{bmatrix} \tag{28}$$

$$\mathbf{V}(k)=\mathbf{col}\Big(\mathbf{v}_{j_{g^K}}(k),\qquad j_{g^K}=1,2,\ldots,m_{g^K}\qquad\forall g^K\Big) \tag{29}$$

$$\mathbf{v}_{j_{g^K}}(k)=\begin{bmatrix} \nu^2_{j_{g^K}}(k)\\ \nu^3_{j_{g^K}}(k)\end{bmatrix}. \tag{30}$$

Equation (24) is the measurement equation which denotes the time-varying relationship between traffic measurements at time-step k. In (24), the elements of $\mathbf{Z}(k)$ referring to traffic measurements represent the time-varying section-wide traffic counts collected from the downstream detector stations of subareas 2 and 3. In the stochastic model, they can be further characterized by the elements of $\mathbf{H}[r(k),k]$, which decompose the measured downstream traffic counts into the state variables and the collected traffic arrivals, and $\mathbf{V}(k)$ represents the error terms associated with the measured traffic counts, which are assumed to follow Gaussian processes. Utilizing the specified relationships between the collected traffic counts and the basic lane-traffic states, the prior predictions of system states derived from state equations are updated in real time in response to the change patterns of zone-based traffic flows under the proposed ramp control. Equations (24)–(30) are the relevant definitional equations to the variables in (24).

3) State Boundaries:

$$0\le r_{j_{g^K}}(k+1)\le 1\qquad\forall r_{j_{g^K}}(k+1),k \tag{31}$$

$$0\le\Omega_{j_{g^K}}(k)\le 1\qquad\forall\Omega_{j_{g^K}}(k),k \tag{32}$$

$$\sum_{\varepsilon=0}^{n_\theta}\Big[\Omega_{j_{g^K}}(k+\varepsilon)\times t\Big]\ge T_{\min}\qquad\forall k,\theta. \tag{33}$$

Equations (31)–(33) are the state-boundary constraints associated, respectively, with state variables $r_{j_{g^K}}(k+1)$, control variables $\Omega_{j_{g^K}}(k)$, and the condition of the minimum ramp-metering rate. Herein, n_θ is the maximum number of sequential time steps, which belong to a given cycle θ, and $T_{\min}$ represents the minimum onramp green time in a given cycle θ (in seconds).

C. Stochastic Optimal-Control-Based Algorithm

A stochastic optimal-control-based algorithm is proposed to perform real-time coordinated ramp control. Herein, we attempt to minimize the differences between the estimated values of state variables and their corresponding ideal values during any given K-based time interval.

1) Aggregated Objective Function for Each Given K-Based Time Interval: The ideal values of state variables refer to the desired values

of the state variables that facilitate moving group-based traffic flows through the corresponding control-zone groups to the greatest extent under the conditions of identified congestion patterns. Accordingly, we have the aggregated objective function ξ_K as follows for each given K-based time interval:

$$\xi_K = \min \sum_{\forall g^K} \zeta_{g^K} \tag{34}$$

where g^K represents a given control-zone group g, which is identified in a given K-based time interval, ζ_{g^K} is the disaggregated objective function associated with g^K, which is given by

$$\zeta_{g^K} = E\left\{\sum_{k=1}^{K}\left[\mathbf{X}_{g^K}(k)-\mathbf{I}\right]^T \mathbf{\Phi}_{g^K}(k)\left[\mathbf{X}_{g^K}(k)-\mathbf{I}\right]\right\} \tag{35}$$

where $\mathbf{X}_{g^K}(k)$ is a $3m_{g^K}\times 1$ time-varying state vector containing the estimates of time-varying state variables associated with g^K, in which m_{g^K} represents the number of control zones involved in the given control-zone group g^K, $\mathbf{I}$ is a $3m_{g^K}\times 1$ unit vector, and $\mathbf{\Phi}_{g^K}(k)$ represents a $3m_{g^K}\times 3m_{g^K}$ time-varying diagonal positive-definite weighting matrix associated with a given control-zone group g^K. Herein, $\mathbf{\Phi}_{g^K}(k)$ can be further expressed as

$$\mathbf{\Phi}_{g^K}(k) = \mathbf{dia}\left(\boldsymbol{\varphi}_\iota(k), \qquad \iota = 1,2,\ldots,m_{g^K}\right)_{3m_{g^K}\times 3m_{g^K}} \tag{36}$$

where $\boldsymbol{\varphi}_\iota(k)$ represents the time-varying diagonal weighting matrix associated with a given control-zone ι in the given zone group g^K, which is denoted by

$$\boldsymbol{\varphi}_\iota(k) = \begin{bmatrix} \phi_\iota & 0 & 0 \\ 0 & \phi_\iota & 0 \\ 0 & 0 & \phi_\iota \end{bmatrix} \tag{37}$$

where each given diagonal element of $\boldsymbol{\varphi}_\iota(k)$ is identically determined by

$$\boldsymbol{\varphi}_\iota(k) = \frac{\eta_\iota(k)}{\sum_{\forall j\in g^K}\eta_j(k)}. \tag{38}$$

This means that state variables in a given control zone will share a weighting value in common, thus forming an identical effect of intrazone state variables on the disaggregated objective function ζ_{g^K}, relative to state variables in the other control zones of the given control group g^K.

2) Major Computational Steps of the Ramp-Control Logic: The primary computational scenarios involved in the proposed algorithm include 1) system initialization, 2) prior prediction of state variables, 3) stochastic optimal estimation of state variables, and 4) determination of the time-varying group-based ramp-control variables. In order to obtain the minimum mean-square estimates of the state variables through the aforementioned Scenarios 2 and 3, the fundamentals of an extended Kalman filter are applied. Except for the scenario of system initialization, the other three computational scenarios are executed in sequence at each time step until the identified incident-induced congestion patterns no longer exist. The following summarizes the major computational steps of the proposed ramp-control logic.

Scenario 1: System initialization

Step 0) Initialize system states, and input collected raw traffic data. Given $k=0$, system states are initialized, including the following.

1) the state vector $\mathbf{X}(0|0)$, referring to the vector of state variables estimated at the beginning of the initial time step, i.e., $k=0$;
2) the covariance matrix of the state-estimation error $\mathbf{Cov}_{\mathbf{x}}(0|0)$;
3) the weighting matrix $\mathbf{\Phi}_{g^K}(k)$ associated with each disaggregated objective function (In addition, let the time-varying ramp-control variable $\Omega_{j_{g^K}}(0)=0$).

It is also noteworthy that the ramp-control groups g^K and the corresponding control-zones j_{g^K} are given at this stage, according to the output of the previous zone-based congestion-pattern-recognition mechanism.

Scenario 2: Prior prediction of system states

Step 1) Compute the prior prediction of the time-varying state vector $\mathbf{X}(k+1|k)$ and the covariance matrix of the state-estimation error $\mathbf{Cov}_{\mathbf{x}}(k+1|k)$, respectively, by

$$\mathbf{X}(k+1|k) = \mathbf{F}\left[\mathbf{r}(k), \mathbf{\Omega}(k), k\right] \tag{39}$$

$$\begin{aligned}\mathbf{Cov}_{\mathbf{x}}(k+1|k) = {}& \dot{\mathbf{F}}(k)\mathbf{Cov}_{\mathbf{x}}(k|k)\dot{\mathbf{F}}^T(k) \\ & + \mathbf{L}\left[\mathbf{r}(k), \mathbf{\Omega}(k), k\right]\mathbf{\Phi}(k)\mathbf{L}\left[\mathbf{r}(k), \mathbf{\Omega}(k), k\right]^T\end{aligned} \tag{40}$$

where $\mathbf{\Phi}(k)$ represents a $(3\sum_{\forall g^K} m_{g^K})\times(3\sum_{\forall g^K} m_{g^K})$ time-varying diagonal positive-definite weighting matrix incorporated in the objective function, as described previously, and $\dot{\mathbf{F}}^T(k)$ is the transpose matrix of $\dot{\mathbf{F}}(k)$, in which $\dot{\mathbf{F}}(k)$ is given by

$$\dot{\mathbf{F}}(k) = \left.\frac{\partial \mathbf{F}\left[\mathbf{r}(k), \mathbf{\Omega}(k), k\right]}{\partial \mathbf{r}(k)}\right|_{\mathbf{r}(k)=\mathbf{r}(k|k)}. \tag{41}$$

Step 2) Calculate the Kalman gain $\boldsymbol{\delta}(k+1)$ in real-time by

$$\begin{aligned}\boldsymbol{\delta}(k+1) = {}& \mathbf{Cov}_{\mathbf{x}}(k+1|k)\dot{\mathbf{H}}^T(k+1)\left[\dot{\mathbf{H}}(k+1)\right. \\ & \left.\times\, \mathbf{Cov}_{\mathbf{x}}(k+1|k)\dot{\mathbf{H}}^T(k+1) + \mathbf{Cov}_{\mathbf{v}}(k+1)\right]^{-1}\end{aligned} \tag{42}$$

where $\mathbf{Cov}_{\mathbf{v}}(k+1)$ is the covariance matrix of $\mathbf{V}(k+1)$, and $\dot{\mathbf{H}}(k+1)$ is denoted by

$$\dot{\mathbf{H}}(k+1) = \left.\frac{\partial \mathbf{H}\left[\mathbf{r}(k+1), k+1\right]}{\partial \mathbf{r}(k+1)}\right|_{\mathbf{r}(k+1)=\mathbf{r}(k+1|k)}. \tag{43}$$

Scenario 3: Stochastic optimal estimation of system states

Step 3) Update the prior prediction of the time-varying state vector $(\mathbf{X}(k+1|k+1))$ by

$$\mathbf{X}(k+1|k+1) = \mathbf{X}(k+1|k) + \boldsymbol{\delta}(k+1)\Delta\mathbf{Z}(k+1|k) \tag{44}$$

where $\Delta\mathbf{Z}(k+1|k)$ is given by

$$\Delta\mathbf{Z}(k+1|k) = \mathbf{Z}(k+1) - \mathbf{H}\left[\mathbf{r}(k+1|k), k+1\right]. \tag{45}$$

In (45), $\mathbf{H}[\mathbf{r}(k+1|k), k+1]$ is the prior measurement-component vector, in which each element characterizes the components of the corresponding element of $\mathbf{Z}(k+1)$ using the prior predictions of state variables at the beginning of the given time-step $k+1$ and the upstream traffic counts collected in the given time-step $k+1$.

Step 4) Truncate the updated state vector $(\mathbf{X}(k+1|k+1))$ with the aforementioned boundary constraint, such that each updated state variable is feasible, subject to the corresponding upper and lower bounds.

Step 5) Update the covariance matrix of the state-estimation error as follows:

$$\mathbf{Cov}_{\mathbf{x}}(k+1|k+1) = \left[\mathbf{I} - \boldsymbol{\delta}(k+1)\dot{\mathbf{H}}(k+1)\right]\mathbf{Cov}_{\mathbf{x}}(k+1|k). \tag{46}$$

Step 6) Update the number of vehicles presented in each given subarea associated with each given control-zone j_{g_K} at the end of time-step $k+1$, i.e., $Q^1_{j_{g_K}}(k+1)$, $Q^2_{j_{g_K}}(k+1)$, and $Q^3_{j_{g_K}}(k+1)$ for subareas 1–3, respectively. The formulas for updating $Q^1_{j_{g_K}}(k+1)$, $Q^2_{j_{g_K}}(k+1)$, and $Q^3_{j_{g_K}}(k+1)$ are expressed as

$$Q^1_{j_{g_K}}(k+1) = \left[a^1_{j_{g_K}}(k+1) + Q^1_{j_{g_K}}(k) + d^3_{j_{g_K}}(k+1) - \widetilde{d}^3_{j_{g_K}}(k+1)\right] \times \left[1 - r^1_{j_{g_K}}(k+1)\right] \tag{47}$$

$$Q^2_{j_{g_K}}(k+1) = \left\{\left[a^1_{j_{g_K}}(k+1) + Q^1_{j_{g_K}}(k) + d^3_{j_{g_K}}(k+1) - \widetilde{d}^3_{j_{g_K}}(k+1)\right] \times r^1_{j_{g_K}}(k+1) + Q^2_{j_{g_K}}(k)\right\} \times \left[1 - r^2_{j_{g_K}}(k+1)\right] \tag{48}$$

$$Q^3_{j_{g_K}}(k+1) = \left[a^3_{j_{g_K}}(k+1) + Q^3_{j_{g_K}}(k)\right] \times \left[1 - r^3_{j_{g_K}}(k+1)\right]. \tag{49}$$

Scenario 4: Determination of the time-varying ramp-control variable

Step 7) Calculate the control-variable vector $\boldsymbol{\Omega}(k+1)$ by

$$\boldsymbol{\Omega}(k+1) = -\mathbf{E}(k+1)\mathbf{X}(k+1|k+1) + \mathbf{C}(k+1) \tag{50}$$

where $\mathbf{E}(k+1)$, which is referred to as a time-varying control gain vector, and $\mathbf{C}(k+1)$ are given, respectively, by

$$\mathbf{E}(k+1) = \left[\mathbf{B}^T(k+1)\mathbf{S}(k+2)\mathbf{B}(k+1) + \mathbf{Cov}_{\mathbf{v}}(k+1)\right]^{-1} \times \mathbf{B}^T(k+1)\mathbf{S}(k+2)\dot{\mathbf{F}}(k+1) \tag{51}$$

$$\mathbf{C}(k+1) = \left[\mathbf{B}^T(k+1)\mathbf{S}(k+2)\mathbf{B}(k+1) + \mathbf{Cov}_{\mathbf{v}}(k+1)\right]^{-1} \times \left[\mathbf{B}(k+1)\boldsymbol{\Phi}(k+1)\mathbf{X}(k+1|k+1) + \mathbf{Cov}_{\mathbf{v}}(k+1)\boldsymbol{\Omega}(k+1)\right] \tag{52}$$

where the matrix $\mathbf{S}(k+2)$, which should satisfy the Riccati equation, is predetermined by

$$\mathbf{S}(k+1) = \boldsymbol{\Phi}(k+1) + \dot{\mathbf{F}}^T(k+1)\mathbf{S}(k+2)\dot{\mathbf{F}}(k+1) - \dot{\mathbf{F}}^T(k+1)\mathbf{S}(k+2)\mathbf{B}(k+1)\mathbf{E}(k+1) \tag{53}$$

and the matrix $\mathbf{B}(k+1)$ is determined by

$$\mathbf{B}(k+1) = \frac{\partial \mathbf{F}\left[\mathbf{r}(k), \boldsymbol{\Omega}(k), k\right]}{\partial \boldsymbol{\Omega}(k)}. \tag{54}$$

Step 8) Check the estimates of the time-varying control variables presented in $\boldsymbol{\Omega}(k+1)$ with (32) and (33) to satisfy the conditions in terms of control variables.

Step 9) Input the next-time-step raw traffic data; let the time step index $k = k+1$, and then, go back to Step 1) for the next computational iteration of the real-time coordinated ramp-control algorithm.

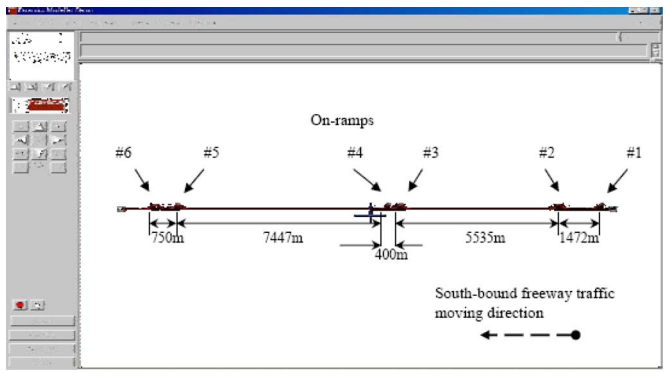

Fig. 3. Illustration of the simulated site.

V. Numerical Study

This numerical study demonstrates the efficiency and suitability of the proposed approach to real-time coordinated ramp control under diverse conditions of freeway lane-blocking incidents. Conveniently, a simplified four-lane freeway corridor, which mimics the mainline segments of the southbound Sun Yat-Sen freeway of Taiwan, R.O.C., located in the metropolitan area of Taipei, was selected as the study site. To illustrate the relative advantages of the proposed method, the numerical results were compared with results obtained using two other ramp-control alternatives, which are presently employed in the study site. Preliminary tests and numerical results with respect to the validity of the estimated state variables were conducted previously, although they are not presented in this paper.

In view of the difficulty in gathering enough real incident-related traffic data for diverse incident cases, simulation data generated from Paramics, Version 3.0, which is a microscopic traffic simulator, was used in the test scenario. The Paramics simulator was also calibrated and tested prior to this study. Details about the calibration and validation tasks required for Paramics can also be found in previous research [13], [14].

To simulate diverse lane-blocking incidents in the aforementioned site, a simulation network mimicking the site was constructed using Paramics, as illustrated in Fig. 3. The site contains six onramps and four offramps. Employing the Paramics simulator, different lane-blocking incidents were generated randomly on the mainline segments of the site, and then, the 10-s simulation data were collected in each given 1-h simulation duration. In this test scenario, 27 types of lane-blocking incidents associated with diverse incident attributes, including incident duration, incident location in a given lane, the lane blocked, and traffic-flow conditions, were simulated. The corresponding simulation prerequisites and key parameters set in Paramics are summarized in Table II.

In the test scenario, three evaluation measures are utilized: 1) zone-based average travel time $\overline{\mathrm{AI}}$; 2) average density-based incident-impact index $\overline{\mathrm{DI}}$; and 3) zone-based average hourly throughput index $\overline{\mathrm{TI}}$. The definitions of these three evaluation measures are denoted, respectively, as follows:

$$\overline{\mathrm{AI}} = \frac{\sum_{k=1}^{K_I}\sum_{j^I=1}^{N_{j^I}}\left[2 \times \left(L^1_{j^I} + L^2_{j^I}\right) \Big/ \left(\overline{\mu}^1_{j^I}(k) + \overline{\mu}^2_{j^I}(k)\right)\right]}{K_I \times N_{j^I}} \tag{55}$$

$$\overline{\mathrm{DI}} = \frac{\sum_{k=1}^{K_I}\sum_{j^I=1}^{N_{j^I}}\left[\overline{q}^1_{j^I}(k) + \overline{q}^2_{j^I}(k)\right]}{2 \times K_I \times N_{j^I}} \tag{56}$$

$$\overline{\mathrm{TI}} = \frac{\sum_{k=1}^{K_I}\sum_{j=1}^{N_j} d^2_j(k)}{K_I \times N_j} \times \frac{3600}{t} \tag{57}$$

TABLE II
SIMULATION PREREQUISITES

Geometric Characteristics			
number of lanes in mainline segments	4	number of lanes in the ramp segments	1
number of on-ramps	6	on-ramp length	250 m
number of off-ramps	4		
Traffic Characteristics			
high-volume	6000 vph	speed limit	100 kph
medium-volume	3000 vph	mean headway	0.86 sec.
low-volume	1500 vph	mean reaction time	0.63 sec.
Ramp Metering			
the proposed coordinated method		dynamic and determined at each given time step	
alternative-1: isolated multi-ramp control		dynamic and determined at each given time step	
alternative-2: fixed-time		preset according to the ramp control manual provided by the National Freeway Traffic Management Agent of Taiwan	
alternative-3: non ramp control		all-green	
Simulation Operational Factors			
simulation time for each case		60 minutes	
incident duration		varying with cases (5 min.-50 min.)	
onset of incident duration		10 minutes from the beginning	
length of a unit time step		60 sec.	

where N_{j^I} represents the total number of control zones upstream from and involving the incident site, and the corresponding control zone is indicated herein by j^I, K_I is the incident duration indicated by the total number of time steps for a given incident case I, $L^1_{j^I}$ and $L^2_{j^I}$, as defined previously, represent the geometric lengths associated with subareas 1 and 2 of a given control-zone j^I, $\overline{\mu}^1_{j^I}(k)$ and $\overline{\mu}^2_{j^I}(k)$ are the section-wide averaged speeds measured from the upstream and downstream detector stations of a given control-zone j^I at a given time-step k, $\overline{q}^i_{j^I}(k)$ represents the time-varying section-based density function which indicates the congestion severity associated with the subarea i of a given control-zone j^I measured at a given time-step k, $d^2_j(k)$, as defined previously, represents the number of vehicles departing from the subarea 2 of a given control-zone j at a given time step, and herein, $d^2_j(k)$ can be collected from the downstream detector station installed in the mainline segment associated with the given control-zone j.

Utilizing the aforementioned criteria, the proposed incident-responsive coordinated ramp-control method was tested with simulated incident data and compared with three ramp-control alternatives, including incident-responsive isolated multiramp control and two existing control strategies implemented in the site, i.e., pretimed ramp control and no control. Here, the incident-responsive isolated multiramp control can be regarded as a special control case of the proposed method, where using the same stochastic optimal-control technique, the local onramp traffic flows are regulated independently without considering the mechanism of zone-based congestion-pattern recognition. To reveal the potential advantages of the proposed coordinated ramp-control method relative to isolated ramp-control strategies, such comparisons are involved in the test scenario.

In addition, it is worth mentioning that, as pointed out in our literature review, the published ramp-control models may aim at incident-free traffic-flow cases rather than incident cases, and specific incident-responsive control functions may not be embedded in these models. In view of this, direct comparisons with these published ramp-control models are not taken into account in this, since such comparisons may not be suitable to illustrate the potential advantages of the proposed model for incident cases.

TABLE III
COMPARISON RESULTS OF RAMP-CONTROL PERFORMANCE

Incident Case	Evaluation Measure	Control Alternatives			
		Proposed Coordinated Control	Isolated Multi-ramp Control	Pretimed Control	No Control
high -volume	$\overline{AI}$ (sec.)	340.5	358.3	411.5	420.8
	$\overline{DI}$ (veh/km)	40.6	43.1	47.5	51.8
	$\overline{TI}$ (veh/hr)	1908.2	1820.8	1567.5	1499.3
medium -volume	$\overline{AI}$ (sec.)	222.9	251.6	267	340.8
	$\overline{DI}$ (veh/km)	31.6	35.4	38.4	42.9
	$\overline{TI}$ (veh/hr)	2730.3	2618.5	2534.2	1630.2
low -volume	$\overline{AI}$ (sec.)	180.1	204.3	238.5	312.3
	$\overline{DI}$ (veh/km)	14.8	18.6	20.7	25.2
	$\overline{TI}$ (veh/hr)	2723.2	2653.1	2608.2	2469.6

The comparison results classified by three levels of traffic-volume conditions are summarized in Table III. Overall, the numerical results shown in Table III may reveal the potential advantages of the proposed coordinated ramp-control method in response to diverse incident-induced freeway congestion patterns in comparison with the other three control strategies. Particularly, the relative improvement in terms of the zone-based average travel time ($\overline{\text{AI}}$) reaches up to 73.4% when compared with the no-control strategy in the low-volume incident cases.

In addition, several significant findings are summarized as follows for discussion: 1) By comparing the measurements of $\overline{\text{TI}}$ in different traffic-flow conditions, high-volume incident-induced traffic congestion is still a critical issue in the corridor-wide ramp-control problem. It is observed that the corresponding $\overline{\text{TI}}$ values measured in high-volume cases are less than those in any other cases, despite the fact that a certain level of the corresponding relative improvements have been achieved using the proposed method. In reality, it should be noted that the main effect of ramp control is on the onramp traffic flows to regulate traffic merging in the weaving area. Once the proportion of the number of onramp vehicles to the number of vehicle in the mainline segment is too small, as measured in such cases as high-volume incident conditions, the efficiency of ramp control may turn out to be rather insignificant. This is agreeable since the averaged throughput measurements observed in high-volume incident cases are less than the others. 2) Under low-volume incident cases, the zone-based averaged travel time can be improved to the greatest extent by using the proposed method, compared with other traffic-flow conditions. It is implied that through appropriate coordinated ramp control, the stability of the zone-based traffic flows, in a given control group under low-volume conditions, can be easily achieved in spite of incident occurrence. A similar generalization can also be found in the medium-volume incident cases. Importantly, the accurate identification of incident-induced congestion patterns determines the efficiency of multiramp control under incident conditions. According to our observations from the numerical, distinguishing incident-induced congestion patterns from other patterns under both the low- and medium-volume traffic-flow conditions appears much easier than that under high-volume conditions. In reality, such a congestion-pattern-recognition effect can also be found while comparing the performance of the proposed coordinated control method and the isolated control strategy shown in Table III, where the average density-based incident-impact index $\overline{\text{DI}}$ can be improved up to 25.7% in low-volume cases. By contrast, under high-volume conditions, the identification of incident-impacted control zones upstream from the incident site is quite difficult and, thus,

may contribute to the inefficiency in determining the corresponding ramp meters for these zone groups. Accordingly, the identification of incident impact in either the temporal domain or the spatial domain appears to be important to address the incident-responsive coordinated ramp-control issues.

VI. Concluding Remarks

This paper has presented a real-time coordinated ramp-control approach to address the incident-induced traffic-congestion problems on multiple mainline segments of freeways. The proposed method embeds two sequential functions in the procedure of incident-responsive coordinated ramp control: 1) identification of control-zone-based congestion patterns and 2) stochastic optimization of group-based ramp-metering control to alleviate corridor-wide incident-induced traffic congestion as much as possible. To execute the aforementioned two major functions, M-GSPRT and stochastic optimal-control approaches are used in the development of the proposed ramp-control algorithm.

Our numerical results show the applicability of the proposed coordinated ramp-control method in responding to diverse incident-induced traffic-congestion conditions through appropriate identification of congestion patterns and dynamic estimation of group-based ramp meters. Results presented in this paper also suggested the relative advantages of the proposed control method compared with three specified ramp-control strategies presently used in the site. Importantly, the proposed approach finds a new and feasible solution that has never been explored in previous literature to deal with large-scale incident-induced traffic-congestion problems. In addition, the findings observed in the numerical may stimulate more valuable research, although some limitations still remain in the current version of the proposed approach.

References

[1] J.-B. Sheu, Y.-H. Chou, and L.-J. Shen, "A stochastic estimation approach to real-time prediction of incident effects on freeway traffic congestion," *Transp. Res. Part B*, vol. 35B, no. 6, pp. 575–592, 2001a.

[2] J.-B. Sheu and S. G. Ritchie, "Stochastic modeling and real-time prediction of vehicular lane-changing behavior," *Transp. Res. Part B*, vol. 35B, no. 7, pp. 695–716, 2001b.

[3] J. A. Wattleworth and D. S. Berry, "Peak-period control of a freeway system-some theoretical investigations," *Highw. Res. Rec.*, no. 89, pp. 1–25, 1965.

[4] L. Shaw and W. R. McShane, "Optimal ramp control for incident response," *Transp. Res.*, vol. 7, pp. 393–411, 1973.

[5] M. Papageorgiou, H. S. Habib, and J. M. Blosseville, "ALINEA: A local feedback control law for on-ramp metering," *Transp. Res. Rec.*, no. 1320, pp. 58–64, 1991.

[6] H. M. Zhang and S. G. Ritchie, "Freeway ramp metering using arterial neural networks," *Transp. Res. Part C*, vol. 5C, no. 5, pp. 273–286, 1997.

[7] G. L. Chang, P. K. Ho, and C. H. Wei, "A dynamic system-optimum control model for commuting traffic corridors," *Transp. Res. Part C*, vol. 1C, no. 1, pp. 3–32, 1993.

[8] Y. Stephanedes and K. K. Chang, "Optimal control of freeway corridors," *ASCE J. Transp. Eng.*, vol. 119, no. 4, pp. 504–514, 1993.

[9] H. Zhang and W. W. Recker, "On optimal freeway ramp control policies for congested traffic corridors," *Transp. Res. Part B*, vol. 33, no. 6, pp. 417–436, 1999.

[10] M. Papageorgiou, "Freeway ramp metering: An overview," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 4, pp. 271–281, Dec. 2002.

[11] J.-B. Sheu, "A fuzzy clustering based approach to automatic freeway incident detection and characterization," *Fuzzy Sets Syst.*, vol. 128, no. 3, pp. 377–388, 2002.

[12] R. F. Stengel, *Stochastic Optimal Control Theory and Application*. New York: Wiley, 1986.

[13] B. Abdulhai, J.-B. Sheu, and W. W. Recker, "Development and performance evaluation of an ITS-ready microscopic traffic model for Irvine, California," *Intell. Transp. Syst. J.*, vol. 7, no. 1, pp. 79–102, 2002.

[14] B. Abdulhai, J.-B. Sheu, and W. W. Recker, "Simulation of ITS on the Irvine FOT area using the 'Paramics 1.5' scalable microscopic traffic simulator—Phase I: Model calibration and validation," Uni. Calif., Berkeley, CA, Calif. PATH Res. Rep. UCB-ITS-PRR-99-12 1055-1425, 1999.

The Selective Random Subspace Predictor for Traffic Flow Forecasting

Shiliang Sun and Changshui Zhang, *Member, IEEE*

***Abstract*—Traffic flow forecasting is an important issue for the application of Intelligent Transportation Systems. Due to practical limitations, traffic flow data may be incomplete (partially missing or substantially contaminated by noises), which will aggravate the difficulties for traffic flow forecasting. In this paper, a new approach, termed the selective random subspace predictor (SRSP), is developed, which is capable of implementing traffic flow forecasting effectively whether incomplete data exist or not. It integrates the entire spatial and temporal traffic flow information in a transportation network to carry out traffic flow forecasting. To forecast the traffic flow at an object road link, the Pearson correlation coefficient is adopted to select some candidate input variables that compose the selective input space. Then, a number of subsets of the input variables in the selective input space are randomly selected to, respectively, serve as specific inputs for prediction. The multiple outputs are combined through a fusion methodology to make final decisions. Both theoretical analysis and experimental results demonstrate the effectiveness and robustness of the SRSP for traffic flow forecasting, whether for complete data or for incomplete data.**

***Index Terms*—Correlation coefficient, random subspace, subset selection, traffic flow forecasting.**

I. Introduction

Reliable analysis of historical trends of traffic flows is an important input to many of the traffic management and control systems in operation and under development. In recent years, utilizing signal processing and machine learning techniques for traffic flow forecasting has drawn more and more attention [1]–[5]. These innovative methods activate and enlighten Intelligent Transportation Systems to a large extent. In this paper, we concentrate on using the ideology of random subspace to deal with the issue of short-term traffic flow forecasting and propose a new predictor.

The problem of short-term traffic flow forecasting is to determine the traffic flow data in the next time interval, usually in the range of 5 min to half an hour. Up to the present, some approaches ranging from simple to complex are proposed for traffic flow forecasting. Simple ones, such as random walk, historical average, and informed historical average can only work well in specific situations [1]. One class of the complex approaches (Urban Traffic Control System prediction methods) are based on forecasting philosophies similar to the above methods. There are also other elaborate methods proposed including

Manuscript received October 31, 2005; revised July 23, 2006, October 11, 2006, October 23, 2006, November 3, 2006, and November 4, 2006. This work was supported by the National Basic Research Program of China under Project 2006CB705500. The Associate Editor for this paper was M. Papageorgiou.

The authors are with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: shiliangsun@gmail.com; sunsl02@mails.tsinghua.edu.cn; zcs@mail.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TITS.2006.888603