Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the past, traffic data are mainly collected by infrastructure-based approaches, e.g., loop detectors and CCTV. However, the high deployment and maintenance cost cause difficulty in pervasively data collection. As compared with traditional infrastructure-based approaches, *crowdsourcing* approaches have lower cost both in deployment and maintenance and thus become a new trend for pervasively discovering traffic information. To obtain accurate traffic information from crowdsourced data, effective mining techniques are required. In addition, the percentage of participators in the mining process affects directly to the accuracy of mining results. In short, *mining techniques* and the *penetration rate* are two major issues in the crowdsourcing approaches. In this paper, we adopt shockwave models to discovery traffic information from crowdsourced data.

Shockwaves are the propagation phenomenon of vehicle accumulation or relief on roads between two traffic flows with different speeds [1][2]. As the speed of traffic flow decreases,

vehicles start to accumulate in a line and form a congestion wave. Later on, as the speed of traffic flow increases, vehicles start to relieve and form a relief wave. For example, consider the shockwave propagation in signalized traffic, on encountering a red sign, vehicles gradually stop, forming a congestion wave. On the sign turning to green, these vehicles start to move, forming a relief wave. The congestion/relief waves of shockwave phenomenon in signalized traffic are called "*stop/go shockwaves*". From the stop/go shockwaves, important parameters of traffic information can be discovered.

Both the infrastructure-based approach and the crowdsourcing-based approach can be used to detect shockwaves. For the infrastructured-based approach, a pair of loop detectors is used to measure the speed and number of incoming and outgoing vehicles [3][4]. However, this approach can not deployed pervasively due to the need of infrastructure. Besides, the shockwave is identified by estimated the position where the flow speed changes, which is inaccurate. For the crowdsourcing-based approach, GPS tracking data are reported to measure the flow variation [5][6]. However, to obtain accurate mining results, as many as 3% penetration rate of GPS-tracking data is needed [5].

In this paper, we propose to use crowdsourced data for the identification of the stop/go shockwaves. The movement data of vehicles in front of a intersection are collected for the shockwave identification. To conquer the problem when the penetration rate of the movement data is low, we propose a folding heuristic that adopts traffic light cycle information to virtually increase the number of movement data used in the shockwave identification. Extensive simulations are performed to validate the proposed concepts, especially on the

2

penetration rate issue. Our results show that the shockwave models with the proposed folding technique are able to compute traffic information, including red/green light transition information and vehicle arrival/relief rate with *root mean square errors* of 6.0/0.6 seconds and 3.2/4.0 vehicles per minute, respectively under a low penetration rate of 1.6%.

The rest of this thesis is organized as follows. Chapter 2 covers some related works. In Chaprer 3, we introduce the shockwave models. Chaprer 4 presents the mining techniques. We verificate our shockwave models in Chaprer 5. Conclusions are drawn in Chapter 6.

Chapter 2

Related Works

In this Chapter, we review some related works about method of traffic information detection. This resherchs can be classfied to fix located sensor based and probe car sensor based.

2.1 Fix Located Sensing

The fix located sensing is means that detect the traffic information by the road side unit. The most popular device of road side unit is the loop detector, so we study some researches of loop detector-based method. Researches using data collected by the loop detector-based method devoted in the development of analytical models for traffic estimation and the improvement of data collection methods by loop detectors. Skabardonis and Geroliminis [7] developed an analytical model for traveling time estimation in signalized arterials. Liu *et al.* [8] discovered a *Queue-Over-Detector* (QOD) problem in the traditional input-output approach for queue length estimation in signalized road segments, and used high resolu-

4

tion traffic signal data with data collected by loop detectors to estimate time-dependant queue length. Followed by Wu *et al.* [3], *Oversaturated Severity Index* (OSI) is defined for quantifying the effects of spillovers, and further separated to temporal OSI and spatial OSI where the temporal OSI describes the detrimental effects created by a residual queue, *i.e.*, the detrimental effects in temporal dimension, and the spatial OSI describes the detrimental effects created by spillovers, *i.e.*, the detrimental effects in spatial dimension. After that, the QOD problem in signalized arterials was further discussed in [9], Wu *et al.* found that the QOD can significantly affects the accuracy of *Arterial Fundamental Diagram* (AFD) and concluded that after removing the QOD effects, one can use AFD to interpret the traffic flow in signalized arterials. In [4], Geroliminis and Skabardonis proposed a method to detect spillovers in signalized intersections.

2.2   Probe Car Sensing

Researches using data collected by the GPS-based method tried to utilize the new mobile sensing technologies to find other means for traffic data collection. In [10], Herrera *et al.* proposed to incorporate GPS tracking logs with data collected from loop detector for traffic reconstruction. However, they did not discuss the penetration rate of GPS in vehicles, which directly affects the accuracy of the estimated traffic flow. In [11], Izadpanah *et al.* proposed a clustering algorithm to automatically identify the trajectories of shockwaves. However, the penetration rate would still be a performance issue in their method. After that, the penetration issue was brought back by Herrera *et al.* [5], they performed a field trial to show

that a $2 - 3\%$ penetration of GPS in vehicles is enough to provide accurate measurement data. In [6], Ban *et al.* proposed to use traveling time between intersections from GPS tracking logs, and developed an analytical model from the concept *Queue Rear No-delay Arrival Time* (QRNAT) to estimate queue length.

2.3 Shockwave Models

The shockwave theory was first proposed by Lighthill, Whitham [1][2] and Richards [12] for modeling highway traffic and later expanded by Stephanopolos [13] and Michalopoulos [14] for modeling signalized traffic. The shockwave theory describes the dynamics of flow, speed and density in space and time domain. The shockwave occurs when two flows intersect to each other. For example, it often cause by a traffic signalm, an accident, change in capacity on roadways and merge free way.
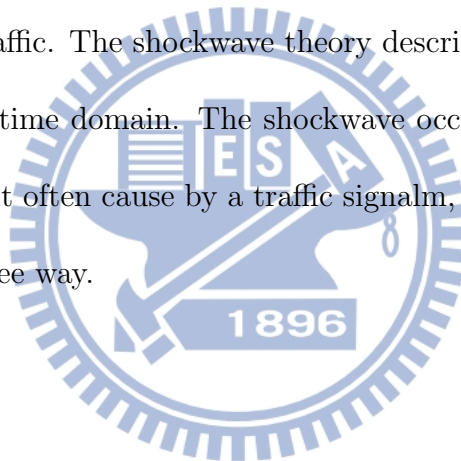
Chapter 3

Stop-Go Shockwave Models

Vehicles moving between signalized intersections that are the most frequently road type in urban areas are forced to stop and go due to the traffic light signal transition. They will gradually stop when a red signal is encountered, and start to move after the traffic light turns to green. The event that a vehicle transits from move to halt is called a stop event, and it can be described by the time and position pair that a vehicle stops. Similarly, the event that a vehicle transits from halt to move is called a go event, and it can also be described by the time and position pair that a vehicle starts to go.

Considering that $\{s_1, s_2, ..., s_n\}$ is a collection of stop events and $\{g_1, g_2, ..., g_n\}$ is a collection of go events. Fig. 3.1 that is a time-position graph illustrates the movement of vehicles in front of a traffic light. Each dash line represents a vehicle's moving track in the time-position plane, and the arrows indicate the moving direction of vehicles. Triangles represent stop events and circles represent go events. If these stop events $s_i = (t_i^s, p_i^s)$ are put together,

a linear trend called the stop shockwave may exist. Similarly, a linear trend called the go shockwaves may exist among these go events $g_i = (t_i^g, p_i^g)$. The propagation of the stop and the go shockwaves in front of a traffic light are depicted by a solid arrow and a chain arrow, respectively, in Fig. 3.1. These stop and go shockwaves can thus be expressed by equations

$$L^s : p^s \doteq \alpha^s t^s + \beta^s \tag{3.1}$$

$$L^g : p^g \doteq \alpha^g t^g + \beta^g, \tag{3.2}$$

respectively. From the equations, many traffic parameters can be found including the arrival and relief rates of vehicles and the traffic light transition times. The mining techniques used to extract these traffic information and a folding heuristic are presented in the next section.
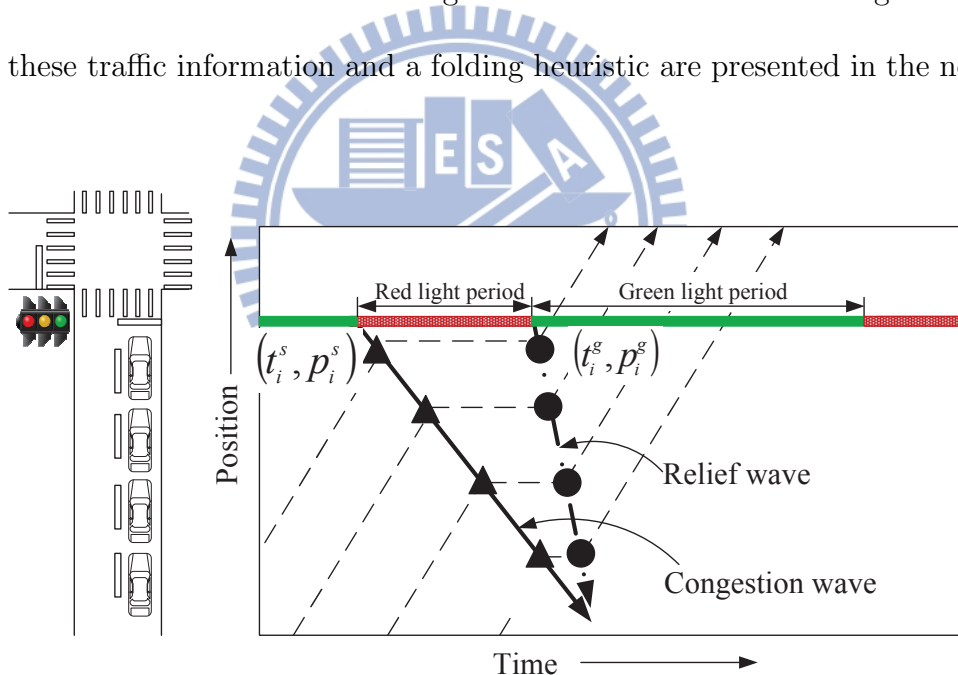
Figure 3.1: Propagation of two shockwaves in a signalized road segment.

Once the equations of the stop-go shockwaves are determined, many useful traffic infor-

mation can be inferred from the parameters of the equations. If we regard the position of the traffic light as the origin, the moment of $p^s = 0$ indicates the start time of a red light period. So, from $0 = \alpha^s t^s + \beta^s$, $t^s = -\beta^s/\alpha^s$ is the start time of a red light period. Similarly, $t^g = -\beta^g/\alpha^g$ is the start time of a green light period. In addition, the arrival and relief rate of vehicles can also be discovered from α^s and α^g in the shockwave equations, respectively. Suppose H is the average space headway between vehicles and L is the number of lanes in a direction. Since the arrival rate of vehicles is the number of incoming vehicles per unit time, and $|\alpha^s|$ is the vehicle accumulating distance per unit time per lane, $|\alpha^s| \times L/H$ is the average arrival rate of vehicles. Similarly, since the relief rate of vehicles is the number of outgoing vehicles per unit time, and $|\alpha^g|$ is the vehicle discharge distance per unit time per lane, $|\alpha^g| \times L/H$ is the average queue discharge rate of vehicles.
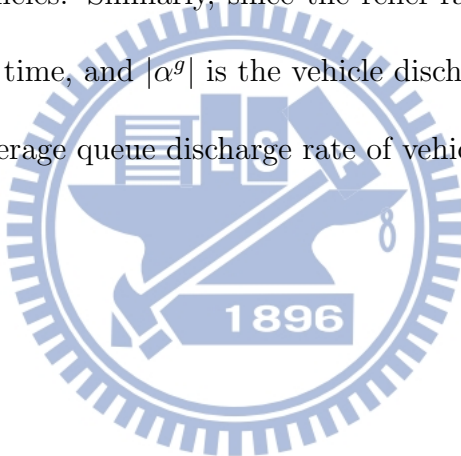
Chapter 4

Mining Techniques

Below, we present the mining techniques to identify the shockwave models for extraction of traffic information, followed by a folding heuristic that is useful to reduce the requirement on the penetration rate.

4.1 Detect Stop-Go Events

The trip of a vehicle is composed of a sequence of alternative stop periods and moving periods separated by stop events and go events. A stop event indicates the transition from a go period to a stop period; and a go event indicates the transition from a stop period to a go period. Let $\{s_1, s_2, ...\}$ denote a collection of the stop events of a trip and $\{g_1, g_2, ...\}$ denote a collection of the go events of the trip. In details, let $s_i = (t_i^s, p_i^s)$ denote a stop event occurring at time t_i^s and location p_i^s, and $g_i = (t_i^g, p_i^g)$ denote a go event occurring at time t_i^g and p_i^g. In addition, $t_i^s < t_i^g < t_{i+1}^s$ for all $i = 1, 2, \cdots$. Ideally, we assume $p_i^s = p_i^g$ for

10

all $i = 1, 2, \cdots$. However, due to the drifting problem of GPS and the inaccuracy of go and stop event detection algorithms, $p_i^s \neq p_i^g$ could happen. We call $s_1, g_1, s_2, g_2, \ldots$ the stop-go (denoted as SG in short) sequence of the trip. The duration from t_i^s to t_i^g is corresponding to the stop period at p_i^s, and the duration from t_i^g to t_{i+1}^s is corresponding to the moving period from p_i^g to p_{i+1}^s. A SG event pair is composed of a stop event and the subsequent go event, $i.e.$, (s_i, g_i) for $i = 1, 2, \cdots$ in the notation.

Practically, it is not so definitely that the stop and go events can be detected. In this work, a *Finite State Machine* (FSM) is used to model the status of a vehicle and also to sketch an algorithm detecting the stop and go events. See Fig. 4.1. The GO state means
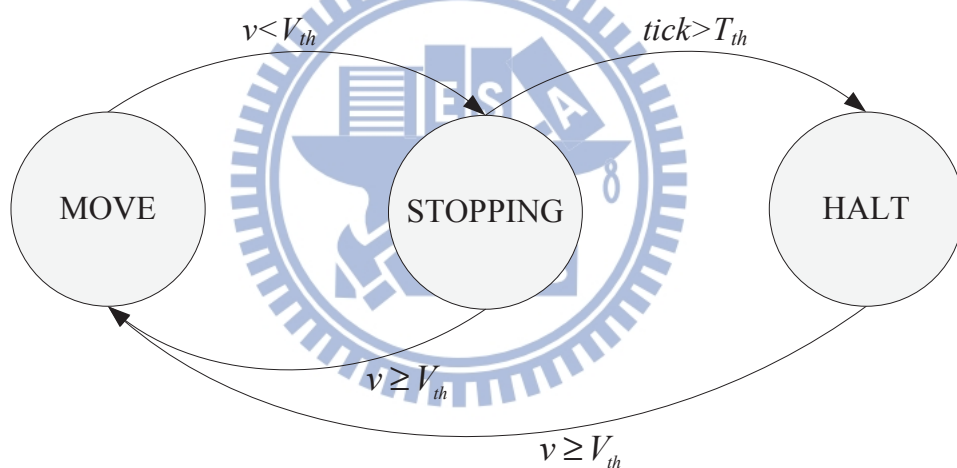

Figure 4.1: The finite state machine that is used to recognize SG events from GPS tracking logs.

the vehicle is moving, and the STOP state means the vehicle is stopped. As mentioned before, the moving speed reported by GPS is more or less inaccurate and somewhat with

delay. Even more, vehicles may move slowly before stopped. So, there exists a grey area between the GO and STOP states. To reflect these facts, we add one more state called STOPPING to quarantine the entrance to the STOP state as the speed is low. To sum up, in the FSM, there are three states, GO, STOP and STOPPING, and two parameters, $toStop$ and $QuarDur$. Let v denote the current speed of the vehicle and $tick$ denote the duration after entering the STOPPING state. The vehicle is in the GO state whenever $v \geq toStop$. In the GO state, if v goes below the threshold $toStop$, $i.e.$, $v < toStop$, the FSM transits to the STOPPING state, and at the same time, $tick$ is reset. In the STOPPING state, as v stays below $toStop$ for $QuarDur$ unit time, $i.e.$, $tick = QuarDur$, the FSM transits to the STOP state. Remind that no matter when, as soon as $v \geq toStop$, the FSM goes to the GO state. As the FSM transits from the STOP state to the GO state, a go event with the location and the time of the transition is reported. As the FSM transits from the STOPPING state to the STOP state, a stop event with the location and the time of the previous transition from the GO state to the STOPPING state is reported. The details of the GPS-based event detection algorithm are depicted in PROCEDURE 1.

The inaccuracy, drifting and non-instant nature and low sampling rate (about 1Hz) of GPS all make the GPS-based algorithm not precise. We use an IMU to measure the vibration of the vehicle to refine the SG events reported by the GPS-based algorithm. We observe that the amplitude of the vibration of g-value, $i.e.$, the reading of the accelerometer, is small as the vehicle is moving; and on the other hand, the amplitude of the vibration of the g-value is relatively large as the vehicle is stopped. The flow chart of the proposed

12

PROCEDURE 1 EventSearch

BEGIN

1: *state* =STOP; *tick* = 0; *i* = 0;

2: **while** there is a datum in GPS buffer **do**

3:　　get the *i*th GPS datum from GPS buffer

4:　　v_i: the speed of the *i*th GPS datum

5:　　**if** $v_i \geq toStop$ **then**

6:　　　**if** *state* =STOP **then**

7:　　　　Report a go event.

8:　　　**end if**

9:　　　*state* =GO

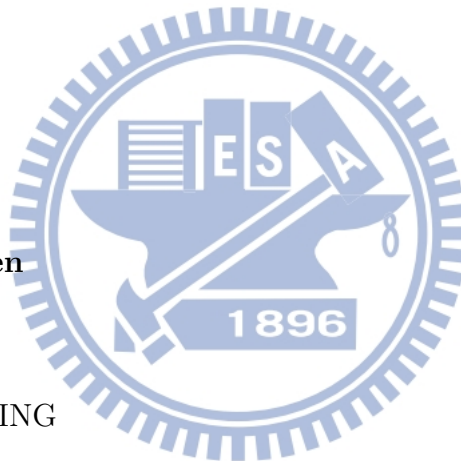10:　　**else**

11:　　　**if** *state* =GO **then**

12:　　　　*tick* = 0

13:　　　　*state* =STOPPING

14:　　　　Keep the location and time

15:　　　**else if** *state* =STOPPING **then**

16:　　　　*tick* = *tick* + 1

17:　　　　**if** *tick* = *QuarDur* **then**

18:　　　　　Report a stop event.

19:　　　　　*state* =STOP

20:　　　　**end if**

21:　　　**end if**

13

22:　　**end if**

23:　　*i* = *i* + 1

24: **end while**

IMU-refine algorithm is depicted in Fig. 4.2. Let g_i be the norm of the g-value at time t_i.
We say g_i is a peak if g_i is a local maximum or local minimum, $i.e.$, $g_i > \max(g_{i-1}, g_{i+1})$
or $g_i < \min(g_{i-1}, g_{i+1})$. Let Δa_j^l and Δa_j^r respectively denote the left amplitude and right
amplitude of the peak at time t_j. In other words, if g_i, g_j, g_k are three consecutive peaks, then
$\Delta a_j^l = |g_i - g_j|$ and $\Delta a_j^r = |g_k - g_j|$. Now, if a stop event (t^s, p^s) is reported by the GPS-
based algorithm, the IMU-refine algorithm is triggered to scan the g-value backward from
time t^s to time $t^s - QuarDur$ to find a peak whose left amplitude is larger then a threshold
a_{stop}. Then, the time and position related to the first found peak are reported to replace the
original stop event. Similarly, if a go event (t^g, p^g) is reported by the GPS-based algorithm,
the IMU-refine algorithm is triggered to scan the g-value forward from time $t^g - 2QuarDur$
to time t^g to find a peak whose right amplitude is larger then a threshold a_{go}. Then, the time
and position related to the first found peak are reported to replace the original go event.
However, no matter for the stop or go event, if there are no such peaks found, the event is
not updated. The pseudocode of the proposed algorithm is given in PROCEDURE 2.

4.2 Identify Stop-Go Shockwaves by LSM

We provide the idea to find out the parameters α^s, β^s, α^g and β^g in Eq. (3.1) and (3.2).
Assume $\{s_1, s_2, \cdots, s_n\}$ are stop events due to a red light period. For any $1 \leq i \leq n$,
$s_i = (t_i^s, p_i^s)$ fits to a stop shockwave

$$L^s : p_i^s \doteqdot \alpha^s t_i^s + \beta^s. \tag{4.1}$$

14
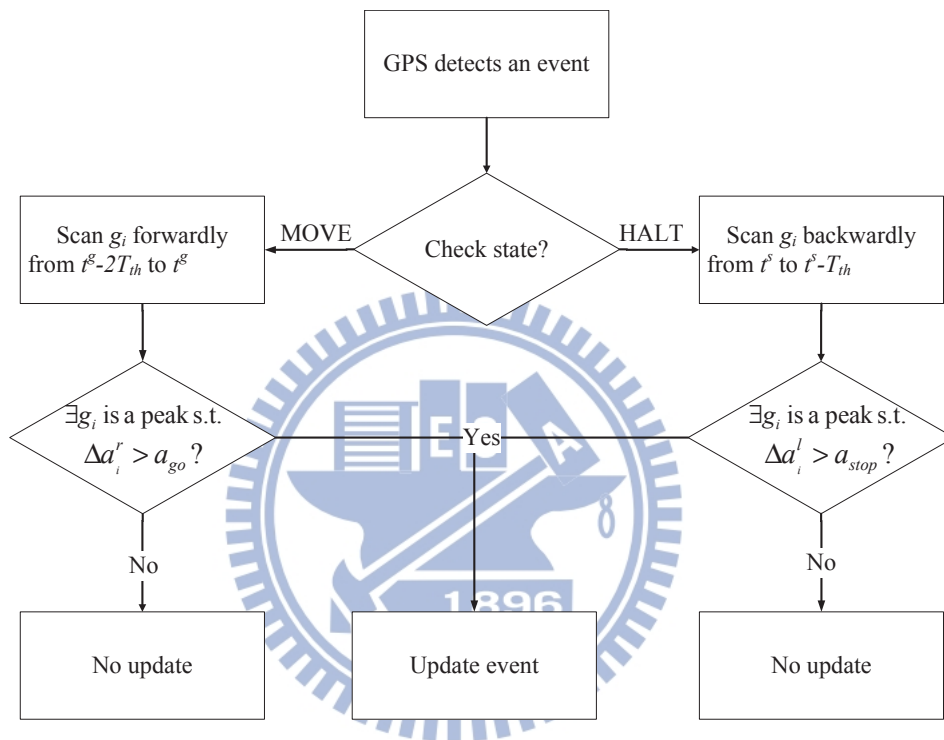
Figure 4.2: The flow chart of the IMU-refine algorithm.

PROCEDURE 2 IMU-refine

BEGIN

1: $state$: the vehicle state

2: **if** $state =$GO **then**

3:     **for** $i = t^s - 2QuarDur$ to t^s **do**

4:         **if** g_i is a peak and $\Delta a_i^r > a_{go}$ **then**

5:             **return** the position and time related to g_i

6:         **end if**

7:     **end for**

8:     No update

9: **else**

10:     **for** $i = t^s$ to $t^s - QuarDur$ **do**

11:         **if** g_i is a peak and $\Delta a_i^l > a_{stop}$ **then**

12:             **return** the position and time related to g_i

13:         **end if**

14:     **end for**

15:     No update

16: **end if**
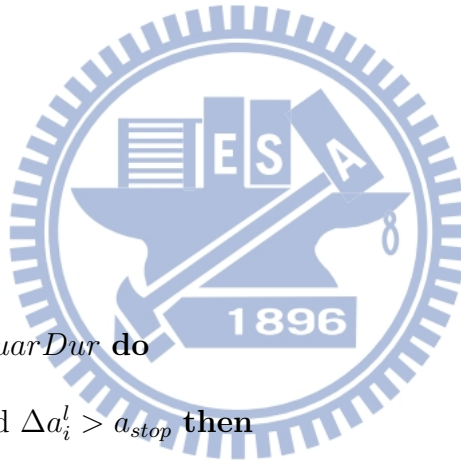
Let

$$
\mathbb{T}^s = \begin{bmatrix} t_1^s & 1 \\ t_2^s & 1 \\ \vdots & \vdots \\ t_n^s & 1 \end{bmatrix}, \ \mathbf{p}^s = \begin{bmatrix} p_1^s \\ p_2^s \\ \vdots \\ p_n^s \end{bmatrix}, \ \text{and} \ \mathbf{x}^s = \begin{bmatrix} \alpha^s \\ \beta^s \end{bmatrix}. \tag{4.2}
$$

The linear system can be written as $\mathbb{T}^s \mathbf{x}^s \doteq \mathbf{p}^s$. Since this is an approximately system, we

apply the *Least Square Method* (LSM) to find the linear regression equation for L^s. It follows

that

$$
(\mathbb{T}^s)^T \mathbb{T}^s \mathbf{x}^s = (\mathbb{T}^s)^T \mathbf{p}^s, \tag{4.3}
$$

and we have

$$
\mathbf{x}^s = \left((\mathbb{T}^s)^T \mathbb{T}^s \right)^{-1} \left((\mathbb{T}^s)^T \mathbf{p}^s \right). \tag{4.4}
$$

Similarly, if $\{g_1, g_2, \cdots, g_n\}$ are go events due to a green light period, for any $1 \le i \le n$,

$g_i = (t_i^g, p_i^g)$ fits to a go shockwave

$$
L^g : p_i^g \doteq \alpha^g t_i^g + \beta^g. \tag{4.5}
$$

Let

$$
\mathbb{T}^g = \begin{bmatrix} t_1^g & 1 \\ t_2^g & 1 \\ \vdots & \vdots \\ t_n^g & 1 \end{bmatrix}, \ \mathbf{p}^g = \begin{bmatrix} p_1^g \\ p_2^g \\ \vdots \\ p_n^g \end{bmatrix}, \ \text{and} \ \mathbf{x}^g = \begin{bmatrix} \alpha^g \\ \beta^g \end{bmatrix}, \tag{4.6}
$$
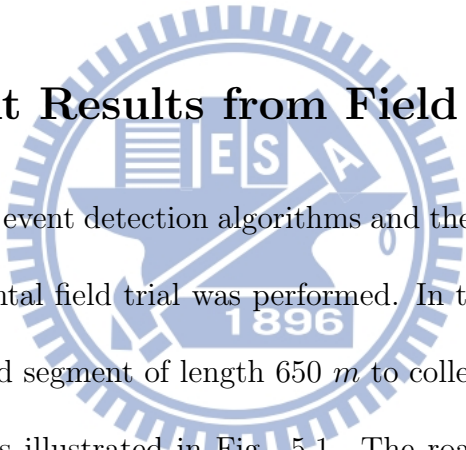
and we have

$$
\mathbf{x}^g = \left((\mathbb{T}^g)^T \mathbb{T}^g \right)^{-1} \left((\mathbb{T}^g)^T \mathbf{p}^g \right). \tag{4.7}
$$

Chapter 5

Model Verification

5.1 Experiment Results from Field Trials

To verify the proposed SG event detection algorithms and the shockwave equations obtained by the LSM, an experimental field trial was performed. In the experiment, two vehicles go around on a two-lanes road segment of length 650 m to collect SG events happening before a signalized intersection as illustrated in Fig. 5.1. The road segment is part of BaoShan Road near NCTU in Hsinchu City from A to B. The experiment is performed in a rush hour from 7:37 am to 8:20 am. The traffic light is located at B and two vehicles go along the black line from A to B to collect the SG events. The vehicles then take U turns at B and go back to A to start another round of data collection. Totally, ten rounds were performed in the experiment. A smartphone is installed in each vehicle to log GPS and IMU data. The sampling rates of GPS and IMU are 1Hz and 20Hz, respectively. Besides the GPS and IMU

Figure 5.1: The road segment for field trial experiments.

data, the program also provides a function to let users manually log stop and go events. Meanwhile, the traffic light cycle is 48 seconds for a red light period, including a 3-seconds-long yellow light period, and 102 seconds for a green light period. The total average vehicle arrival rate is 24 vehicles per minute in which the average vehicle arrival rate of the left lane is 18.6 vehicles per minute. The average vehicle arrival rate is counted from recorded video. Note that, the experiment is performed in the left lane since the traffic flow in Taiwan is mixed with heavy vehicles, passenger vehicles and motorcycles, and motorcycles usually drive in the right lane. In addition, in the ten-round experiment, there are total ten samples among the 43-minute-length experiment and there are about $42.5 \times 18.6 = 790.5$ vehicles during the experiment. Therefore, the penetration rate is $10/790.5 \approx 1.2\%$. In the rest of this sections, we will first give the analysis on the SG event detection algorithm and then on the event-based shockwave model discovery.

5.1.1 Analysis on SG Event Detection

Fig. 5.2 depicts the time domain waveforms of GPS and IMU tracking logs, the SG events recorded manually, and the SG events reported by the proposed algorithms. In the rest of this section, the GPS-based event detection algorithm and GPS+IMU-based event detection algorithm are called the GPS-based algorithm and GPS+IMU-based algorithm for short, respectively. In the figure, the (blue) line marked with diamonds is the speed reported by GPS and the y-axis on the left is the scale of speed in Km/hr; and the (red) line marked with squares is the g-norm measured by IMU and the y-axis on the right is the scale of g-norm
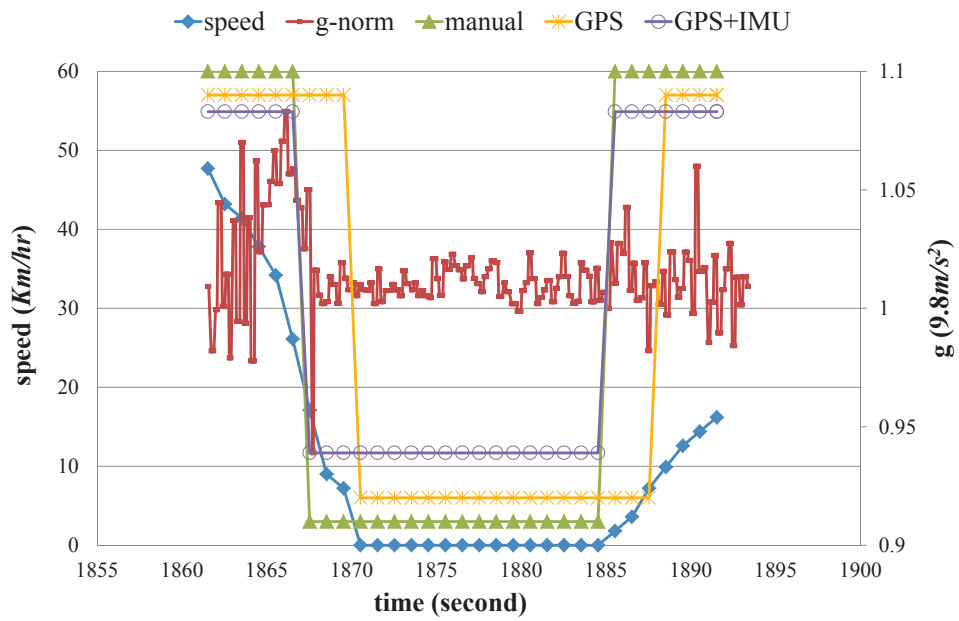
Figure 5.2: A snapshot of data waveforms from GPS tracking logs for illustration of GPS delay phenomenon on speed.

in g, *i.e.*, 9.8 m/s^2. The (green) line marked with triangles, the (yellow) line marked with asterisks and the (purple) line marked with circles respectively are the STOP and GO states manually logged, determined by GPS-based algorithm, and determined by the GPS+IMU-based algorithm. In the three SG state curves, the high level denotes the GO state and the low level denotes the STOP state. The thresholds used in the SG event detection algorithms, including *toStop*, *QuarDur*, a_{stop} and a_{go}, are set to 3 Km/hr, 3 seconds, 0.035g and 0.036g, respectively. The stop and go events are respectively marked at 1867.5 second and 1885.5 second manually, at 1870.5 second and 1888.5 second by the GPS-based algorithm, and at 1867.5 second and 1885.5 second by the GPS+IMU-based algorithm. We can see that GPS has a delayed response and therefore the SG events reported by the GPS-based algorithm is usually behind the manually logged events. However, from the example provided in Fig. 5.2, we can see that the delay problem of GPS can be significantly corrected by utilizing IMU to fine tune the reported stop and go events.

Next, we would like to point out the drifting nature of GPS. In Fig. 5.3, the (blue) line marked with diamonds is the distance of the vehicle to the traffic light, which is calculated based on the position reported by GPS, and the (green) line marked with triangles indicates the manually logged SG events. We can see that the distance drifts from -136.2 m to -139.3 m during the stop period. In other words, even the vehicle does not move, the position reported by GPS varies in different time. However, the distance should be a constant when a vehicle stops.

The SG event detection algorithms will be evaluated in terms of temporal and spatial
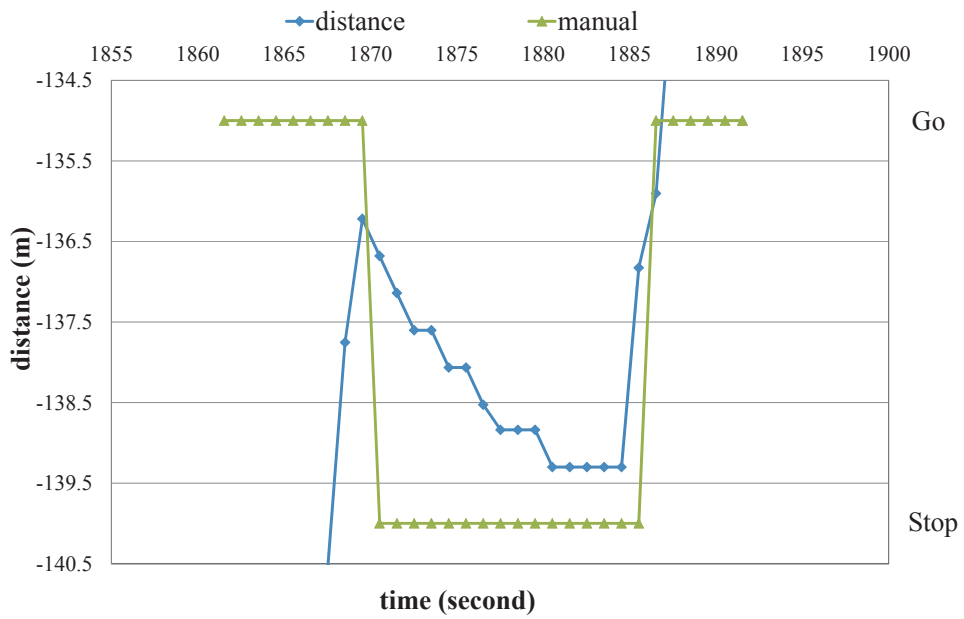
Figure 5.3: A snapshot of data wavefrom from GPS tracking logs for illustration of GPS drift phenomenon on position

errors, hitting rates, and false detection rates based on the manually logged SG events.
Fig. 5.4, in which the x-axis is the time line and the y-axis is the distance to the traffic
light, depicts the stop events collected in the experiment. The squares mark the manually
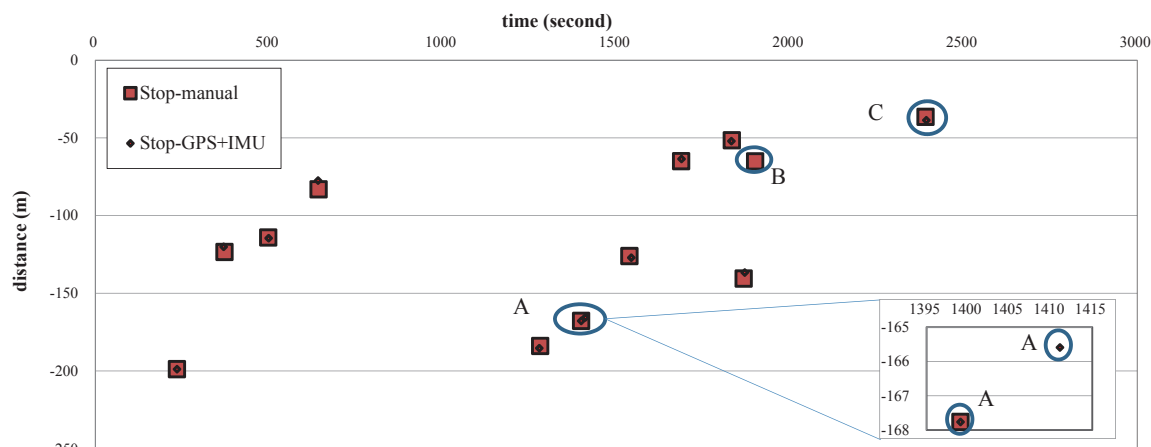
Figure 5.4: Stop events from the manual log and the GPS+IMU-based algorithm.

logged stop events and the diamonds mark the stop events reported by the GPS+IMU-based
algorithm. The major reason that causes the inconsistent between the two groups of stop
events is the drifting problem of GPS. In details, the inconsistent can be further categorized
into two levels of errors. One is the false detection of the algorithms. The false detection
can be classified to positive false detection and negative false detection. The positive false
detection, *e.g.*, the event marked by A in Fig. 5.4, means that an event is reported by the
algorithms but does not actually happen and thus is not manually logged. On the contrary,
the negative false detection, *e.g.*, the event marked by B in Fig. 5.4, means that an event
logged by users but is not reported by the algorithms. The other level of inconsistence is the

inaccuracy of the reported events, including the reported time and position. For example, in Fig. 5.4, we can see that the square marks and diamond marks do not precisely coincide. This is due to the inaccuracy of the time of reported events plus the drifting problem of GPS.

The temporal error is defined as the time differences between the manually logged events and the corresponding events reported by the algorithms. The GPS-based and GPS+IMU-based algorithms respectively have 2.74 seconds and 1.81 seconds temporal errors in average. Similarly, the spatial error is defined as the distance differences between the manually logged events and the corresponding events reported by the algorithms. The GPS-based and GPS+IMU-based algorithms respectively have 3.89 m and 2.61 m spatial errors in average. Both results indicate that IMU can improve the accuracy of the SG event detection.

The hitting rate is defined as the ratio of the number of the SG events correctly reported by the algorithms to the number of the manually logged SG events. The false detection rate, including the positive false detection rate and the negative false detection rate, is defined as the ratio of the number of the false SG events to the number of manually logged SG events. We can see from Fig. 5.4 that the hitting rate is 11/12 and the positive false detection rate is 1/12, and the negative false detection rate is 1/12. The results show that the proposed SG event detection algorithms can detect most SG events with a small false detection rate.

In summary, the experiment results suggest that it is sufficient to detect SG events by the GPS-based algorithm. However, to improve the accuracy of the position and time of the reported SG events, assistant devices such as IMU or OBD (On Board Diagnostic) may be

needed in the SG event detection.

5.1.2 Analysis on Shockwave Model Discovery by LSM

In this section, we will look closer to shockwave equations obtained by applying LSM to the SG event data sets. In the experiment, the length of traffic light cycle is 150 seconds. Due to the low penetration rate of vehicles equipped with our experiment devices, a 150-seconds cycle time is too short to log enough SG event data to discover the shockwave equations. Even in the 43-minutes experiment, there are only 12 pairs of SG events logged. Without further processing, no information about the shockwave equations can be easily extracted from the data illustrated in Fig. 5.4. Based on the assumptions that the arrival of vehicles is stable and all traffic light cycles during the experiment are with the same length, we can shift the SG events into one traffic light cycle by the modulo operation mentioned in Section **??**. Fig. 5.5 illustrates the distribution of the SG events reported by the GPS+IMU-based algorithm after shifting the SG events into the traffic light cycle starting at time 0. In the figure, the stop events are marked by (red) triangles, and the go events are marked by (green) circles.

About Fig. 5.5, first of all, we notice that the group of stop events and the group of go events both clearly reveal some trends but there are three outliers, or called noises, marked by A, B and C. The events reported by the algorithms marked by A and B respectively are the positive and negative false events discussed in the previous section. The outlier marked by C is the SG event caused by spillover. The spillover effects are treated as noises here, and
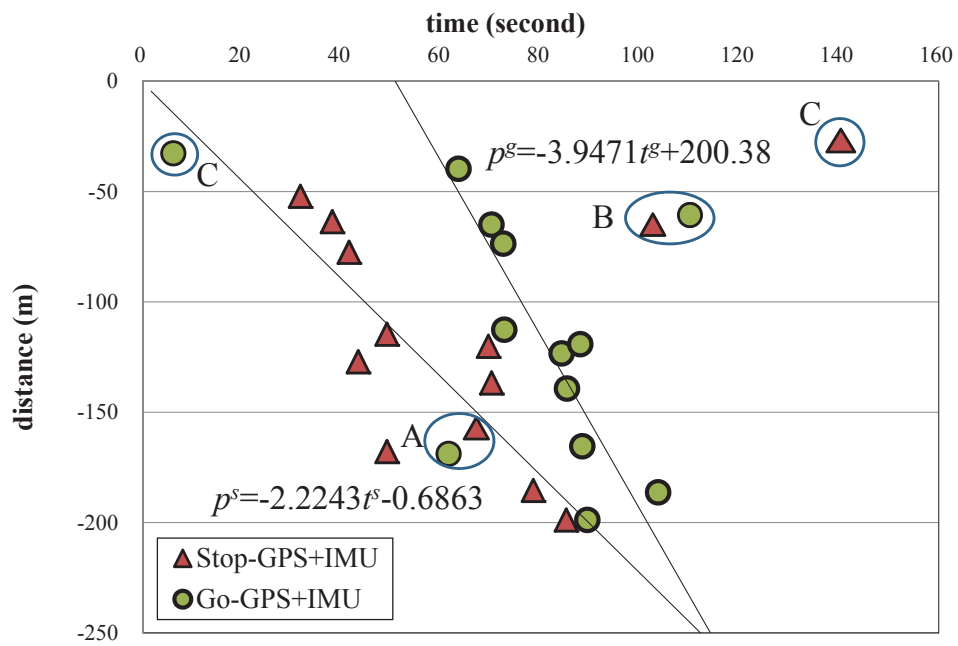
Figure 5.5: The SG events reported by the GPS+IMU-based algorithm after shifted into the same traffic light cycle.

Table 5.1: Coefficients and RMSE of the congestion wave and relief shockwave equations in different cycles.

Coefficients/RMSE	$L^s : (\alpha^s, \beta^s)$		$L^g : (\alpha^g, \beta^g)$	
T_1	$(-2.07, -10.64)$	12	$(-4.08, 212.76)$	11
T_2	$(-1.95, -12.18)$	11	$(-4.23, 244.80)$	12
T_3	$(-2.22, -0.69)$	12	$(-3.95, 200.38)$	12

Table 5.2: Coefficients of the congestion wave and relief shockwave equations

Coefficients	$L^s : (\alpha^s, \beta^s)$		$L^g : (\alpha^g, \beta^g)$	
Manual	$(-2.07, -10.64)$	9	$(-4.08, 212.76)$	8
GPS	$(-1.95, -12.18)$	10	$(-4.23, 244.80)$	11
GPS+IMU	$(-2.22, -0.69)$	12	$(-3.95, 200.38)$	12
GPS+OBD	$(-2.07, -10.64)$	9	$(-4.08, 212.76)$	9

further discussions are left to our future study. After removing the outliers and applying the LSM, the congestion wave equation $L^s : p^s = -2.22t^s - 0.69$ and the congestion relief equation $L^g : p^g = -3.95t^g + 200.38$ are obtained and depicted in Fig. 5.5.

To compare the shockwave equations obtained from three SG data sets that are collected manually, by the GPS-based algorithm, and by the GPS+IMU-based algorithm, in Table 5.2, the coefficients of the congestion wave and relief equations obtained by the LSM are given. The coefficients (α^s, β^s) of the congestion wave equation based on the stop event data sets

Table 5.3: Trffic information estimation for shockwave equations

Parameters	arrival rate (vel/min)		Red light period
	$h = 6.10$	$h = 9.14$	(sec)
Measured	18.63		48.1
Manual	20.67	13.59	57.3
GPS	19.51	12.83	64.2
GPS+IMU	22.21	14.60	51.1

logged manually, reported by the GPS-based algorithm, and reported by the GPS+IMU-based algorithm respectively are $(-2.07, -10.64)$, $(-1.95, -12.18)$ and $(-2.22, -0.69)$. The coefficients (α^g, β^g) of the congestion relief equation based on the go event data sets logged manually, reported by the GPS-based algorithm, and reported by the GPS+IMU-based algorithm respectively are $(-4.08, 212.76)$, $(-4.23, 244.8)$ and $(-3.95, 200.38)$. As mentioned before, if the time tick starts from the beginning of the red light period, $-\beta^s/\alpha^s$ that is the beginning of the red light period should be equal to 0, and $-\beta^g/\alpha^g$ that is the beginning of the green light period should be equal to 48. From the shockwave equations based on the SG events respectively logged manually, reported by the GPS-based algorithm, and reported by the GPS+IMU-based algorithm, the values of $-\beta^s/\alpha^s$ are respectively -5.12, -6.25 and -0.31, and the values of $-\beta^g/\alpha^g$ are respectively 52.15, 57.87, and 50.73.

To further verify the obtained shockwave equations, Table 5.3 lists traffic parameters that can be estimated from the shockwave equations. The parameters include the average vehicle

arrival rate and the length of the red light period. The average arrival rate is estimated by $|\alpha^s| \times l/h$ where $l = 1$ in the experiment. As suggested in [15], a reasonable vehicle head space h is from $6.10m$ to $9.14m$. Based on the SG events respectively logged manually, reported by the GPS-based algorithm, and reported by the GPS+IMU-based algorithm, the average vehicle arrival rates are respectively 20.67, 19.51 and 22.21 if $h = 6.10$, and respectively 13.59, 12.83 and 14.6 if $h = 9.14$. We can see that the vehicle arrival rates are around the measured one. The length of the red light period can be estimated by $\left| \frac{\beta^g}{\alpha^g} - \frac{\beta^s}{\alpha^s} \right|$. The estimated value for three data sets are respectively 57.3, 64.2 and 51.1.

The results indicate that the framework proposed in this work can properly depict the shockwave equations, and however, the GPS+IMU-based algorithm is usually the best one. At the end of this section, we give three comments to briefly conclude the experiment results. First, the shockwave equations obtained by the proposed framework have been verified by several traffic parameters. Second, IMU are useful to improve the accuracy of the SG event detection. Last, and the most important, our experiments show that 1.2% penetration rate is enough to discover the shockwave equations. In short, the proposed framework for shockwave model discovery is practical and economic.

5.2 Simulation

To verify the proposed concepts, extensive simulations are performed on VISSIM [16] traffic simulator to generate vehicle traffic in a signalized single-lane road segment of $650m$ for 600 seconds. Each vehicle moves in a range of speed from $48km/h$ to $58km/h$ before encountering

a red signal. It slows down its speed to $0km/h$ after the light turns to red and starts to accelerate its speed after a green signal. In addition, it behaves according to a car following model given in [16]. The traffic light system is operated with the cycle length, the red and green phase length set to 150 seconds, 47 seconds and 103 seconds, respectively. The traffic light phase sequence starts with a red light period followed by a green light period. Since the stop/go shockwaves may not imply the knowledge of vehicle speeds, and vice versa, the stop/go events detected by different speeds: $0km/h, 1km/h, 2km/h$ and $3km/h$ are tested for different stop/go shockwave models. The traffic information including vehicle arrival/relief rates and the start time of red/green light periods are mined from different stop/go shockwave models under various levels of traffic intensity from 18 to 24 vehicles per minute. In addition, to evaluate the penetration rate issues on the accuracy, different sampling numbers of events from 3 to 8 in the same traffic light cycle are used for shockwave identification. For example, if there are 18 events detected in the same traffic light cycle, then the penetration rate for 3 sampled events is $\frac{3}{18} \doteq 16.6\%$.

5.2.1 Time-Position Tracking

Fig. 5.6 illustrates a time-position graph of the propagation of stop-go shockwaves in the simulations with an average vehicle arrival rate of 24 vehicles per minute. In the figure, the stop and go events detected by $0km/h$ are marked by triangles and circles, respectively. It can be noted that the stop/go shockwaves occurring with a similar slope every traffic light cycle.
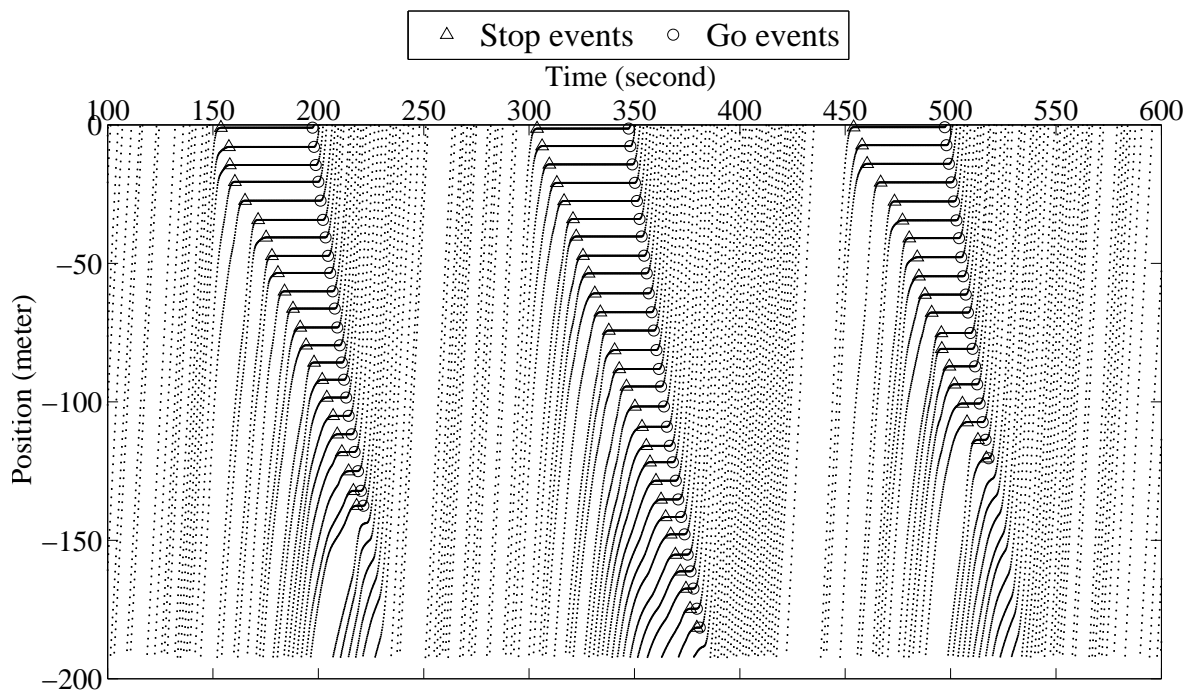
31

Figure 5.6: The propagation of stop-go shockwaves in the simulations.

5.2.2 Simulation for LSM

In this subsection, we talk about our simulation result of using LSM to identify stop-go shockwave model.

Vehicle Flow Information

We first validate the vehicle flow information including the vehicle arrival and relief rate. Fig. 5.7 is the average *Root Mean Square Errors* (RMSE) of estimated vehicle arrival rates and mined from stop shockwaves by stop events detected by different speeds. The x-axis represents the number of sampling events; the y-axis represents the RMSE of the estimated

vehicle arrival rates. The lines marked with diamonds, squares and triangles represent the estimated vehicle arrival rates under average vehicle arrival rates of $18, 20$ and 24 vehicles per minute, respectively. The vehicle arrival rate is estimated by $|\alpha^s| \times L/H$ where $L = 1$ since the road segment is single-lane, and H is the average headspace calculated from the simulation data ranged from $6.5m$ to $6.8m$. Theoretically, H can be set from $6m$ to $9m$. The RMSE is calculated by the errors between the estimated vehicle arrival rates and the actual vehicle arrival rates during the simulations. It can be noted that the RMSEs are all within 4 vehicles per minute, and the results show that 4 stop events in a cycle can achieve an RMSE of 3.2 vehicles per minute, and this is good enough for estimation of vehicle arrival rate. The *Probability Density Function* (PDF) of errors of the vehicle arrival rates is represented by a heavy line in Fig. 5.8. Note that, in our simulations, the PDF shows similar distribution for different stop shockwaves. It shows a bell-shape like distribution with the center at 0 vehicle per minute and the width ranged from -5 to 5 vehicles per minute. The *Cumulative Distribution Function* (CDF) of errors of the vehicle arrival rates is represented by a red line in Fig. 5.9. It shows the 80 percent of arrival rate errors is between -2 and 2 vehicle per minute. In addition, the error distribution is with mean of 1.37 vehicles per minute and standard deviation of 1.16 vehicles per minute. This reflects the fact that $|\alpha^s|$ is the vehicle accumulating distance per unit time per lane. The results show that the proposed model is useful for the vehicle arrival rate estimation.

Fig. 5.10 is the average RMSE of estimated vehicle relief rates measured by the similar way in the vehicle arrival rate estimation. Same legends are used in the figure to represent
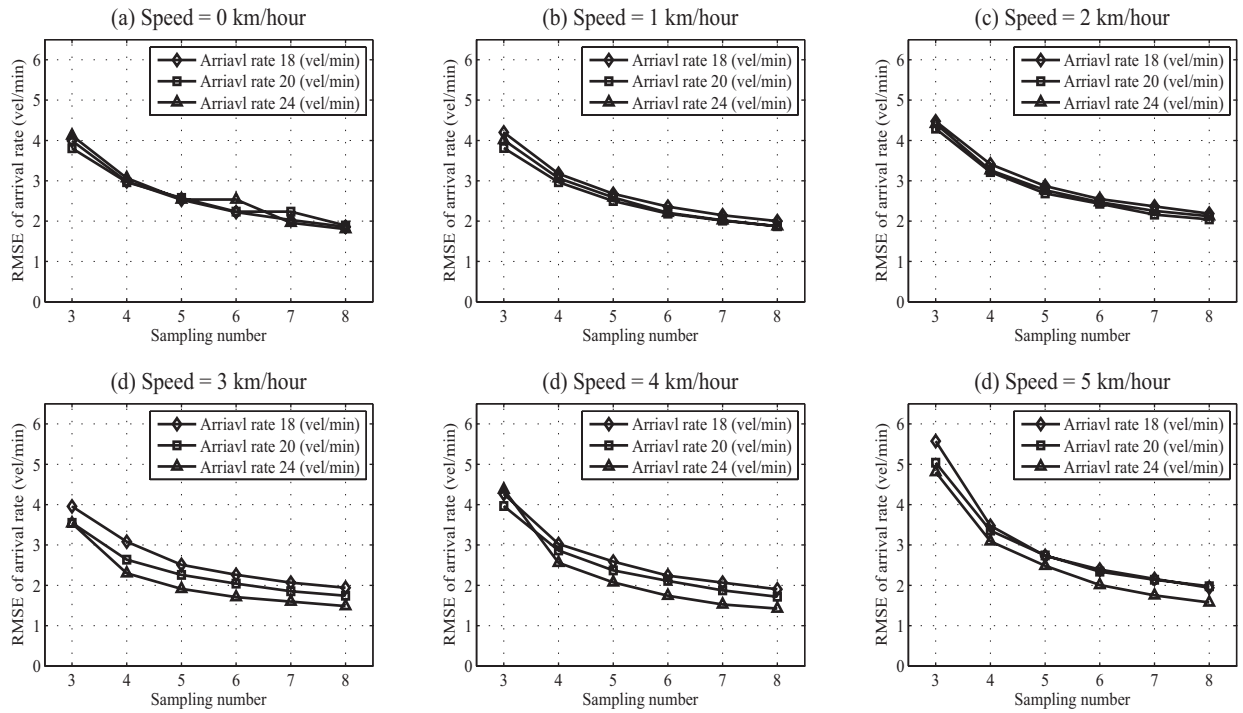
Figure 5.7: RMSE of estimated arrival rates of different stop shockwave models.

the vehicle relief rates under average vehicle arrival rates of 18, 20 and 24 vehicles per minute. The vehicle relief rate is estimated by $|\alpha^g| \times L/H$. There are two observations from the results. First, the vehicle relief rates mined from the go shockwaves with go events detected by speeds of $0km/h$ have the lowest RMSE among these four shockwave models. Errors become larger as the speed is getting higher. This is because when the go events are determined by a higher speed, it causes the go shockwave underestimate the vehicle relief rate. Second, it shows that there is no significant affection by the penetration rate and 4 go events in a cycle is enough to achieve a good estimation of vehicle relief rate which has an RMSE of 4 vehicles
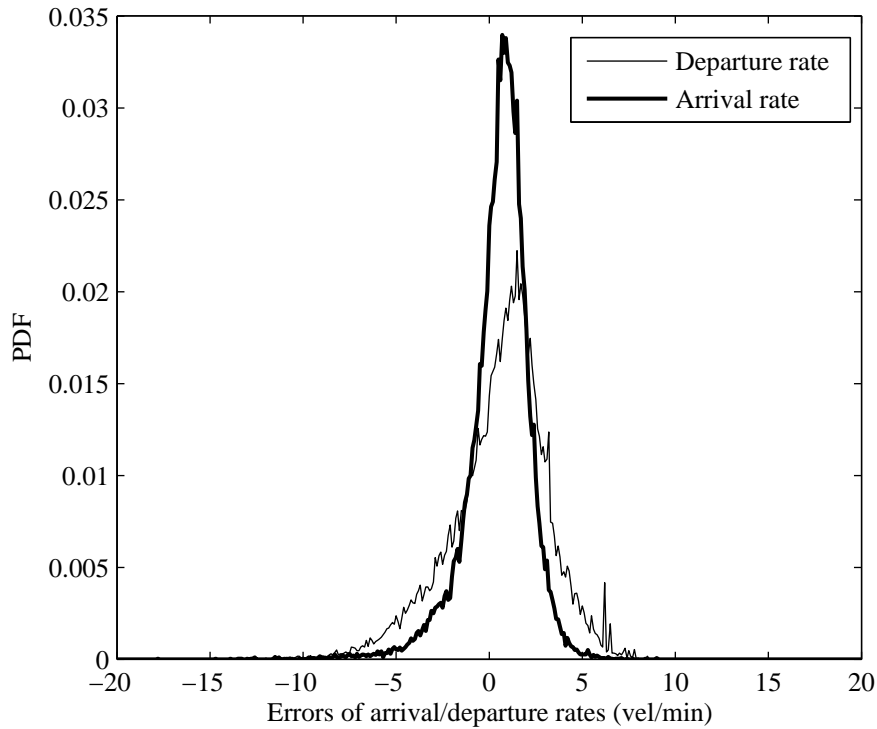
34

Figure 5.8: The PDF of the errors of the vehicle flow information. (a) the vehicle arrival rate. (b) the vehicle relief rate.

per minute. This is because vehicles all start to move immediately when a traffic light turns to green, i.e., the slopes between go events are very similar. The PDF of errors of the vehicle relief rates is depicted by a line in Fig. 5.8. It shows a similar bell-shape distribution as the vehicle arrival rate with mean of 2.12 vehicles per minute and standard deviation of 1.62 vehicles per minute. The CDF of errors of the vehicle relief rates is depicted by green line in Fig. 5.9, and it shows the 80 percent of errors is between -3.2 and 3.2 vehicle per minute. The results shows that $|\alpha^g| \times L/H$ can be used as a good estimator for the relief rate.
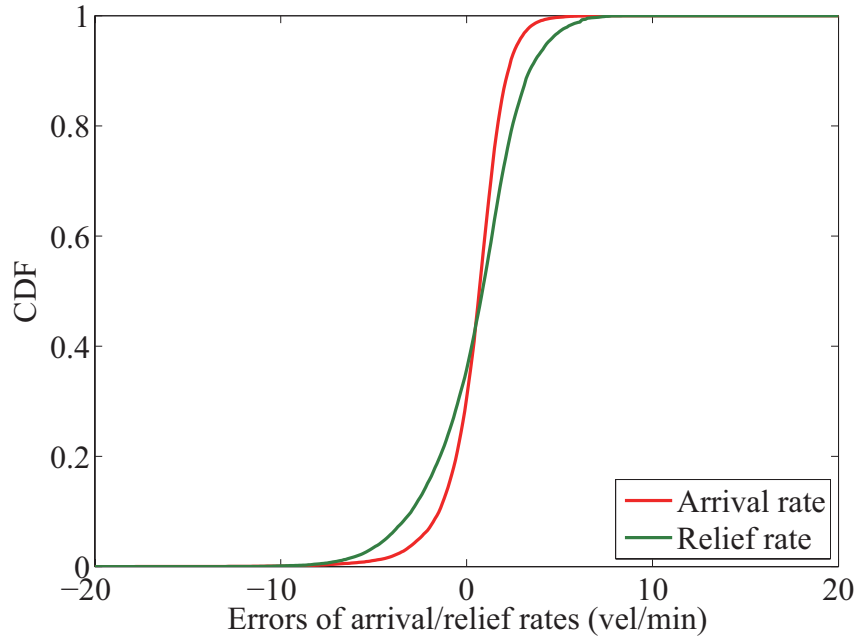
Figure 5.9: The CDF of the errors of the vehicle flow information. (a) the vehicle arrival rate. (b) the vehicle relief rate.

Traffic Light Transition Information

The traffic light transition information including the start time of a red light period and the start time of a green light period are validated. Fig. 5.11 is the average RMSE of estimated start times of red light periods from stop shockwave models with stop events detected by different speeds. The x-axis represents the number of sampling events; the y-axis represents the RMSE of the estimated start time of a red light period in seconds. The lines marked with diamonds, squares and triangles represent the estimated start time of a red period under average vehicle arrival rates of $18, 20$ and 24 vehicles per minute, respectively. The
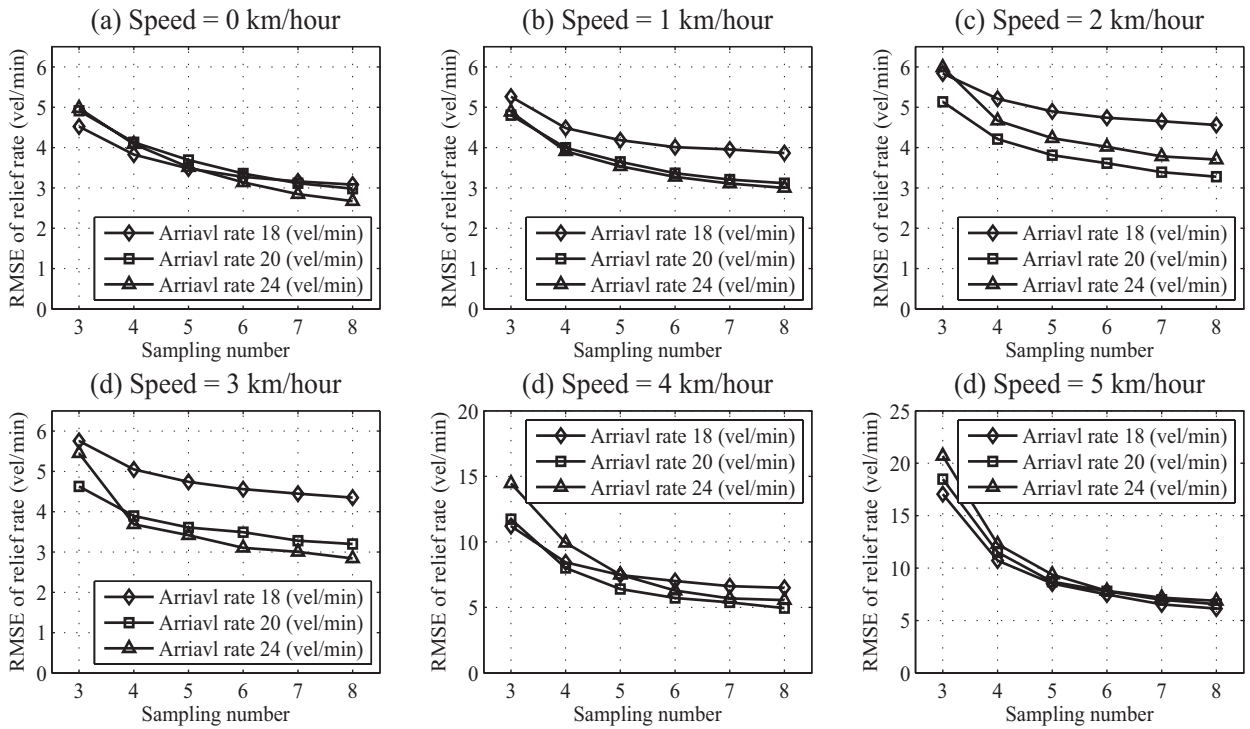
Figure 5.10: RMSE of estimated relief rates of different go shockwave models.

start times of red light periods in the simulations are at $0, 150, 300, 450, 600$ seconds. The

x-intercept of the stop shockwave equation $-\beta^s/\alpha^s$ is used as an estimator for the start time

of a red light period. The RMSE is calculated by the errors between the estimators and

the actual start times of red light periods in the simulations. The results of the four stop

shockwave models show that the RMSEs of estimated start times of red light periods are all

in the range of 6~8 seconds. The PDF of the RMSEs of errors of the start times of red light

periods is denoted by a dash line in Fig. 5.12, and the CDF is denoted by red line in Fig.

5.13. It can noted that the PDF is a bell-shape with the center at 5 seconds and width of

a range from -5 seconds to 25 seconds, and the 80 percents of RMSEs of the start times of red light periods is between -9 seconds and 9 seconds. This is because there may not be vehicles appearing at the time when a traffic light turns to red, and this delay time is related to the vehicle arrival rate.
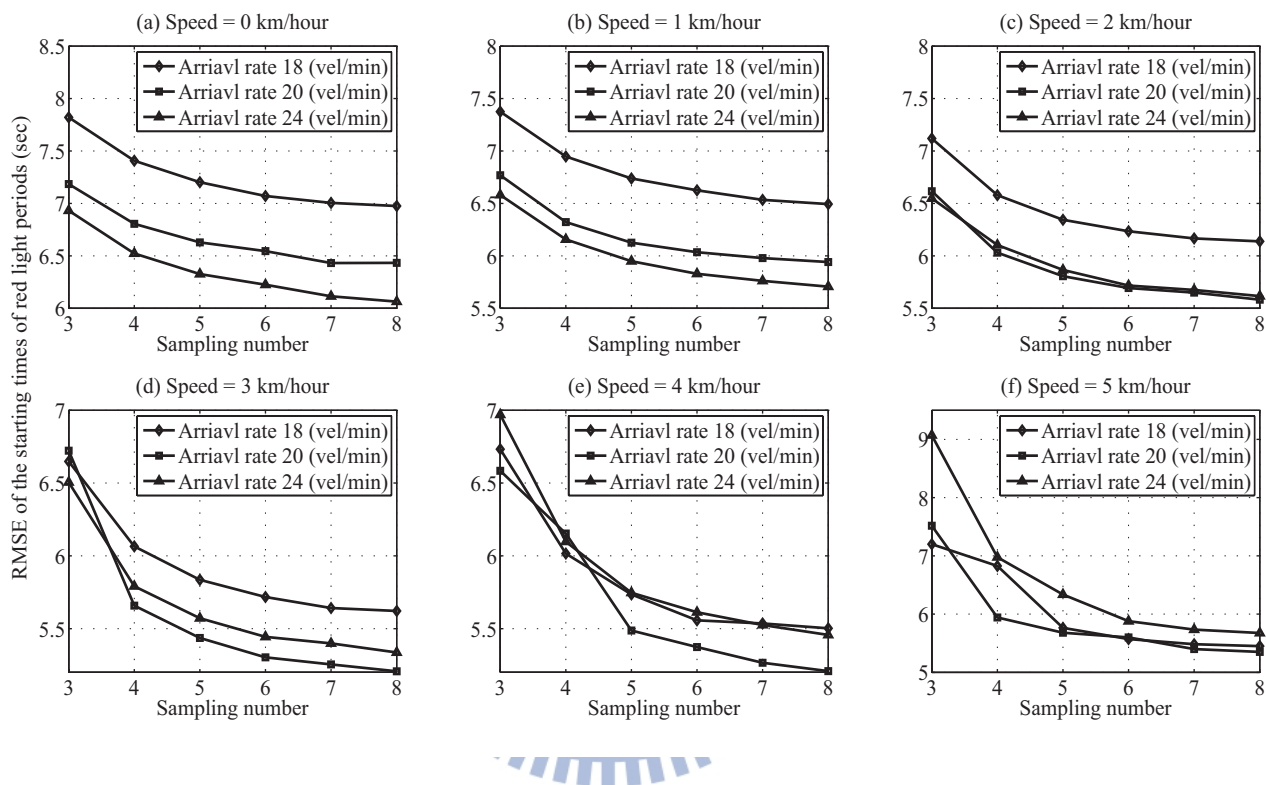

Figure 5.11: RMSE of estimated start times of red light periods of different stop shockwave models.

Fig. 5.14 is the average RMSE of estimated start times of green light periods from different go shockwave models. The similar legends are used for the representation of the estimated start time of a green light period. The start times of green light periods in the
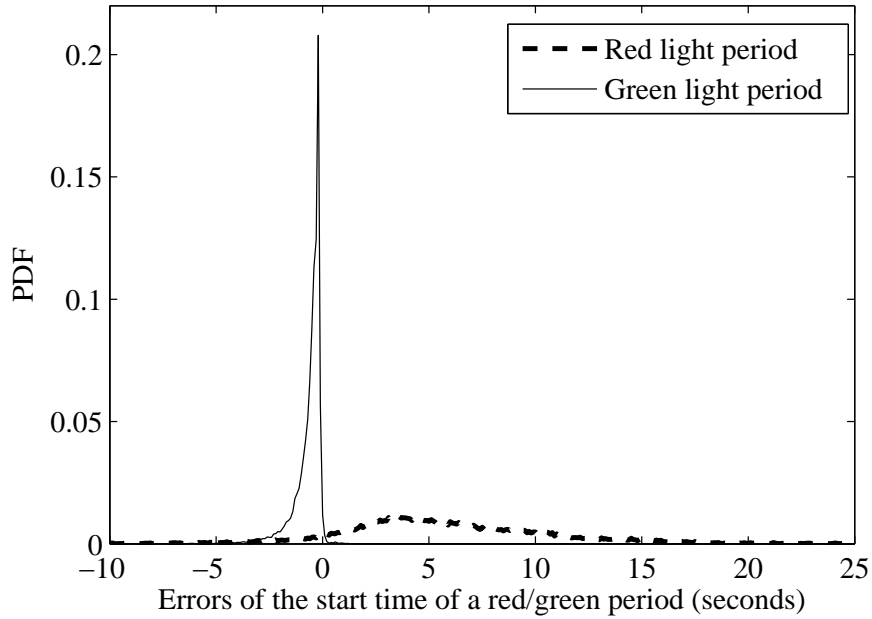
Figure 5.12: The PDF of the errors of the start times of red and green light periods.

simulations are at $51, 201, 351, 501$ seconds. Similarly, $-\beta^g/\alpha^g$ is used as an estimator for the

start time of a green light period. Note that only 4 go events is enough for a good estimation

of the start time of a green light period. The results of the four go shockwave models show

that the RMSEs of estimated start times of green light periods are all within 2 seconds. The

PDF of the RMSEs of errors of the start times of green light periods is denoted by a solid

line in Fig. 5.12, and CDF is denoted by green line in Fig. 5.13. It can noted that the PDF

is a bell-shape with the center at -0.5 seconds and width of a range from -5 to 1 seconds,

and the Fig. 5.13 shows the 80 percents RMSEs is between -1.3 to 0. This is because there

is a delay when a vehicle starts to move after a traffic light turns to green. The results show
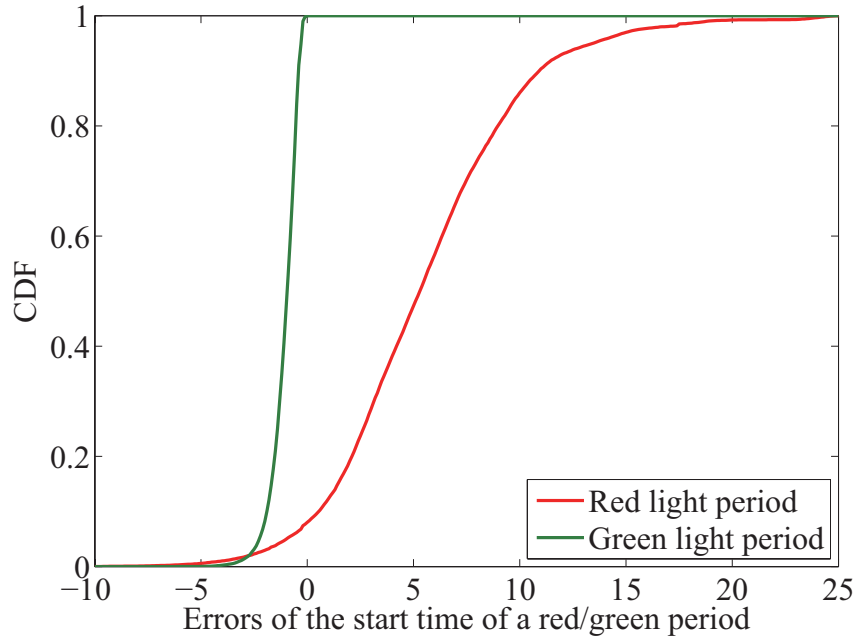
Figure 5.13: The CDF of the errors of the start times of red and green light periods.

that $-\beta^g/\alpha^g$ can be used as a good estimator for the start time of a green light period.

Summary

In summary, the simulations reveal four things about the performance on traffic information mining from the stop/go shockwave models. First, both stop/go shockwave models are useful to estimate the vehicle arrival/relief rate. Second, the stop shockwave model may not be used in the mining of the start time of a red light period, in contrast, the go shockwave model is very useful in the mining of the start time of a green light period. We may use the go shockwave occurring in the cross direction of the road to help in the mining of the start time of a red light period. Third, it is better to use stop/go events detected by 0 speed for
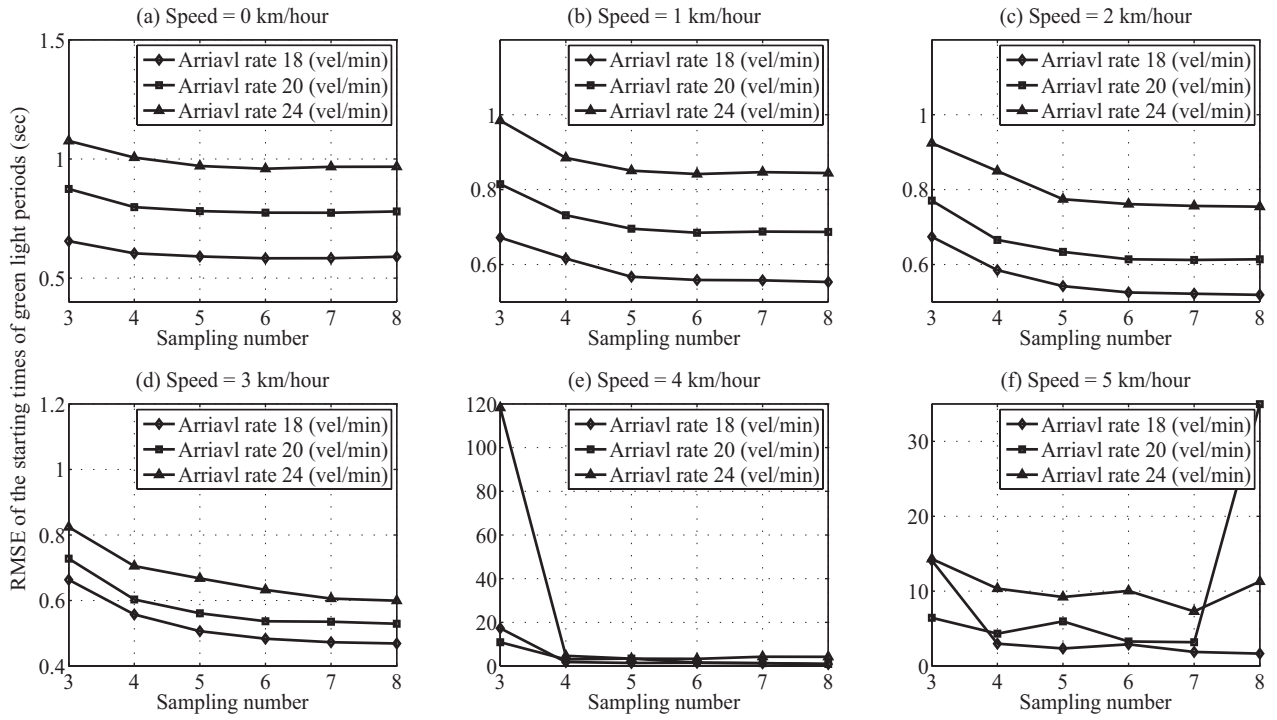
Figure 5.14: RMSE of estimated start time of green light periods of different go shockwave models.

shockwave identification. Most importantly, it only requires 4 stop/go events to obtain a good result for the traffic information estimation. If the folding technique is applied to fold events within 10 cycles and the vehicle arrival rate is 24 vehicles per minute, the penetration rate can be reduce as low as $\frac{4}{24 \times 10} \doteq 1.6\%$.

Chapter 6

Conclusions

In this thesis, we propose a smartphone-based probe car system that utilizes mobile sensing to pervasively detect and grade road abnormality such as potholes, speed bumps, expansion joints manhole covers, etc. To overcome vibration factors such as orientation of smartphones, phone racks, sensor chips of smartphones, types of vehicles, and driving speed, several mechanisms are proposed including a vertical component extraction algorithm and an abnormality detection algorithm by the standard deviation of vertical vibration. The result indicates that our algorithm can eliminate these factors and grade the road abnormality.

Bibliography

[1] M. J. Lighthill and G. B. Whitham, "On kinematic waves I. flood movement in long rivers," *Proceedings of the Royal Society of London. Series A, Mathematical and PhysicalSciences*, vol. 229, no. 1178, pp. 281–316, May 1955.

[2] ——, "On kinematic waves II. a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London. Series A, Mathematical and PhysicalSciences*, vol. 229, no. 1178, pp. 317–345, May 1955.

[3] X. Wu, H. X. Liu, and D. Gettman, "Identification of oversaturated intersections using high-resolution identification of oversaturated intersections using high-resolution," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 626–638, Jan 2010.

[4] N. Geroliminis and A. Skabardonis, "Identification and analysis of queue spillovers in city street networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. pp, no. 99, pp. 1–9, May 2011.

[5] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568–583, Aug 2010.

[6] X. Ban, P. Hao, and Z. Sun, "Real time queue length estimation for signalized intersections using travel times from mobile sensors," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1133–1156, Dec 2011.

[7] A. Skabardonis and N. Geroliminis, "Real-time monitoring and control on signalized arterials," *Journal of Intelligent Transportation Systems*, vol. 12, no. 2, pp. 64–74, Apr 2008.

[8] H. X. Liu, X. Wu, W. Ma, and H. Hu, "Real-time queue length estimation for congested signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 4, pp. 412–427, Aug 2009.

[9] A. Skabardonis and N. Geroliminis, "An empirical analysis on the arterial fundamental diagram," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 255–266, June 2010.

[10] J. C. Herrera and A. M. Bayen, "Traffic flow reconstruction using mobile sensors and loop detector data," in *the 87th TRB Annual Meeting*, 15 November 2007, pp. 1 – 18.

[11] P. Izadpanah, B. Hellinga, and L. Fu, "Automatic traffic shockwave identification using vehicles trajectories," in *the 88th Annual TRB Conference*, 11 - 15 January 2009, pp. 1 – 12.

[12] S. W. on the Highway, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, Feb 1956.

[13] G. Stephanopoulos, P. G. Michalopoulos, and G. Stephanopoulos, "Modelling and analysis of traffic queue dynamics at signalized intersections," *Transportation Research Part A: General*, vol. 13, no. 5, pp. 295–307, Oct 1979.

[14] P. G. Michalopoulos, G. Stephanopoulos, and G. Stephanopoulos, "An application of shockwavetheory to traffic signal control," *Transportation Research Part B: Methodological*, vol. 15, no. 1, pp. 35–51, Oct 1981.

[15] J. M. Ishimaru and M. E. Hallenbeck, "Flow evaluation design technical report," Washington State Department of Transportation, Tech. Rep., March 1999.

[16] G. M. S. Mosseri, M. A. Hall, and J. J. Meyers, "VISSIM micro-simulation modeling of complex geometry and traffic control: a case study of ocean parkway, NY," in *Institute of Transportation Engineers 2004 annual meeting and exhibit compendium of technical papers*, 1-4 August 2004, pp. 1–10.