The following publication E. Ua-areemitr, A. Sumalee and W. H. K. Lam, "Low-Cost Road Traffic State Estimation System Using Time-Spatial Image Processing," in IEEE Intelligent Transportation Systems Magazine, vol. 11, no. 3, pp. 69-79, Fall 2019 is available at https://doi.org/10.1109/MITS.2019.2919634.

Low-cost Road Traffic State Estimation System Using Time-spatial Image Processing

Ekalux UA-AREEMITR, Agachai SUMALEE, and William H.K. LAM

***Abstract*—Road traffic mobility can be described by its Level of Services (LoS). A major challenge in traffic state and LoS estimation is the limitation of observed traffic data. To derive the traffic state of a road network, a sensor network needs to be installed. Most stationary sensing techniques involve high investment in terms of the sensor installation, data communication and computational resources. This paper proposes a low-cost image processing system for road traffic state estimation using time-spatial image (TSI) processing. The TSI is an image processing technique for transforming a series of video images into a single image. Therefore, the TSI can reduce memory resources compared with the traditional methods. A camera can be exploited for traffic-state estimation through integration with TSI generating and processing modules. In addition, traffic state variables such as space-mean-speed, flow and density can be estimated. Empirical results are provided based on several experiments to show that TSI processing is a viable low-cost approach to traffic state estimation.**

***Index Terms*—: Traffic state estimation, Space-mean-speed, Traffic flow, Traffic density, Level of service, Time-spatial image processing, Low video frame rate processing**

I. Introduction

TRAFFIC state estimation is an essential process of describing the traffic state on a particular road segment. The traffic state is estimated based on a number of traffic variables e.g. flow, speed, and density. The estimation results are fundamental requirements for improving traffic control, operation, and planning. In addition, providing real-time traffic information to road users can improve road service capabilities in terms of speed and volume. The road users can make use of the real-time information to avoid congestion and minimize their travel time [1].

In terms of traffic data collection, two types of traffic sensing technologies are used: stationary sensors and floating data. For taking measurements, stationary sensors such as inductive loop detectors [19], radar detectors [20] and closed-circuit television cameras [4] are widely used due to their ability to collect traffic data from most of the passing vehicles at a particular reference point or detection zone. In addition, the sensors can provide traffic state information by lane. A major drawback of stationary sensors, however, is the cost of sensor installation, which significantly affects the number of sensors installed on each road segment. Moreover, the installation and maintenance processes for inductive loop detectors interrupt the traffic operation.

Floating data or mobile data are alternative traffic state estimation approaches that include cellular phone signals [2], GPS [3], and Bluetooth data [24]. The floating data can be considered as low-cost traffic data since the data can be collected from ubiquitous smartphones. Furthermore, sensor installation is unnecessary for most forms of floating data. The traffic data derived from an individual vehicle are travel time and speed. These data may not represent the actual traffic state at the macroscopic level, however, since the floating data alone are unable to estimate the traffic flow and density directly.

Due to the limitations of floating data, it is easier for traffic operators and road users to perceive and interpret the traffic state from stationary sensors. In particular, CCTV cameras are usually installed on most road segments for road traffic monitoring, especially on expressways, owing to the advantages of relatively low installation costs and low traffic disruption during maintenance. With the advancement of image processing and computer vision techniques, the methodologies of traffic data extraction using video images have been developed [4]. Table I summarizes the latest traffic sensing technologies, particularly video detection systems based on CCTV cameras. The information in Table I is adapted from the Traffic Monitoring Guide [18] in order to compare types of traffic data in different dimensions.

In general, the existing image processing techniques for traffic data extraction can be classified into 2 types: background-subtraction-based approaches [5, 6, 7, 9] and motion-detection-based approaches [8, 10, 11, 12]. High-quality video images might be required for collecting accurate traffic data in this way. In order to produce the required high-

Manuscript received March 31, 2018; The work described in this paper was jointly supported by research grants from the Research Committee of the Hong Kong Polytechnic University (Project No. 1-ZVFJ) and the Research Institute of Sustainable Urban Development of the Hong Kong Polytechnic University (Project No.1-ZVBX and 1- ZVBZ).

Ekalux Ua-areemitr is with The Hong Kong Polytechnic University, Kowloon, Hong Kong, China (e-mail: ekalux.uaareemitr@connect.polyu.hk)

Agachai Sumalee is with The Hong Kong Polytechnic University, Kowloon, Hong Kong, China. (e-mail: agachai.sumalee@polyu.edu.hk).

William H.K. Lam is with The Hong Kong Polytechnic University, Kowloon, Hong Kong, China, and also with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China (e-mail: william.lam@polyu.edu.hk).

TABLE I
LITERATURE REVIEW AND COMPARISON OF TRAFFIC SENSING TECHNOLOGIES

	Traffic Sensing Technology	Flow	Density	Speed	Vehicle Classification	Lane-based Data	VDO Image Resolution	Communication Bandwidth	Sensor Purchasing Cost	Processing Time
Mobile Data	Floating-car data from cell phone signal system[2]	X	X	X			N/A	H	H	H
	Floating-car data from global positioning system (GPS) [3]			X			N/A	M	L	L
	Floating-car data from Bluetooth system[24]			X			N/A	M	L	L
Stationary Data	Inductive Loop [19]	X		X*		X	N/A	L-M	L	L
	Microwave Radar [20]	X	X	X	X	X	N/A	M	L-M	M
	Video Detection System (Background subtraction) [6]	X		X	X	X	H	M	L-M	L
	Video Detection System (Motion detection) [10, 11]	X		X	X		M-H	M	L-M	M-H
	Video Detection System (Autoscope) [5]	X	X	X	X	X	H	L	M-H	L
	Video Detection System (TSI) [13, 14, 15, 16, 17]	X			X	X	M-H	L	L	L
	Video Detection System (Proposed system)	**X**	**X**	**X**	**X**	**X**	**L**	**L**	**L**	**L**

N/A = not applicable, X* = required multiple sensors, X = applicable, L = Low, M = Moderate, H = High

quality video images, it is also necessary to have access to high quality video cameras, sufficient communication bandwidth, and sufficient processing time. In order to compromise the extra requirements caused by high-quality video images, a group of researchers [13-17] has developed a method of extracting traffic data extraction based on a time-spatial image technique (TSI).

The TSI is an adapted image processing technique used to transform a series of video images into a single time-spatial image [13]. By setting a virtual detecting line on a monitoring road section, this technique can continue capturing the image details on the line and discard other information. The line information captured from a series of video images over time can then be constructed as a stack of lines called a TSI, which includes the data of vehicles that travel across the virtual detection line. It is worth noting that TSI significantly reduces the required memory storage compared to video images for an equal amount of time. Hence, this technique could minimize data communication costs and be suitable for decentralized processing.

TSI has been adopted for traffic data extraction by some researchers [14-17]. Li et al. [14] introduced a methodology for estimating traffic states with TSI. The method involved installing a video camera to observe a road junction and the results showed that the time-spatial images in free-flow traffic conditions were totally different from in congestion cases. In their study, Li et al. were able to identify congested traffic states with this technique. In another study, Yue [15] used TSI for the purpose of counting vehicles. By implementing a morphological technique and hole filling on the edge of the detected image, the system was able to extract and count the vehicles on an urban road. In 2016, Zhang et al. [16] extended the work initiated in [15] by adding a foreground extraction technique to provide more accurate vehicle shape estimations for vehicle counting purposes. Li et al. [17] also used TSI processing for vehicle classification. According to the previous studies, TSI processing is capable of estimating traffic conditions and traffic state information. However, there is a lack of research about vehicle speed and density estimation and real-time traffic state estimation using the TSI technique.

To this end, this paper proposes a low-cost road traffic state estimation system based on the TSI technique. The system is capable of estimating lane-based traffic variables (e.g. traffic flow, space-mean-speed, and density) in real-time with minimum requirements in terms of sensor costs, data storage, communication bandwidth, and processing time. As a potential solution to the need for a low-cost traffic state estimation system, the original works on TSI are extended by proposing a method for speed and density estimation. Moreover, the use of low-quality video images and low video framerates are investigated in order to find the limitations of the proposed system compared to a commercial speed-trap video sensor. It can be seen from Table I that the traffic information provided by the proposed system is similar to the information provided by Autoscope. However, the proposed system has lower resource requirements in terms of video resolution and sensor costs. This paper provides an in-depth examination of the system performance in terms of estimation accuracy and data transmission cost compared with the traditional video processing method.

This paper is organized as follows: Section II describes the detailed procedures of traffic state estimation using TSI. In Section III, empirical studies are carried out to evaluate the performance of the proposed system. Finally, section IV provides the conclusions of the study.

II. TRAFFIC STATE ESTIMATION USING TSI

As illustrated in Fig. 1, the proposed system consists of 4 main parts: TSI initialization, TSI generating, vehicle tracking and traffic state estimation. The first part involves determining a virtual detection line for generating TSI. The TSI generating process is for video data pre-processing and TSI construction.. In the vehicle tracking part, a TSI is processed to estimate vehicle numbers, occupancy time, vehicle entry time, vehicle exit time and time headway. Finally, the information from vehicle tracking can be further used for traffic state estimation.

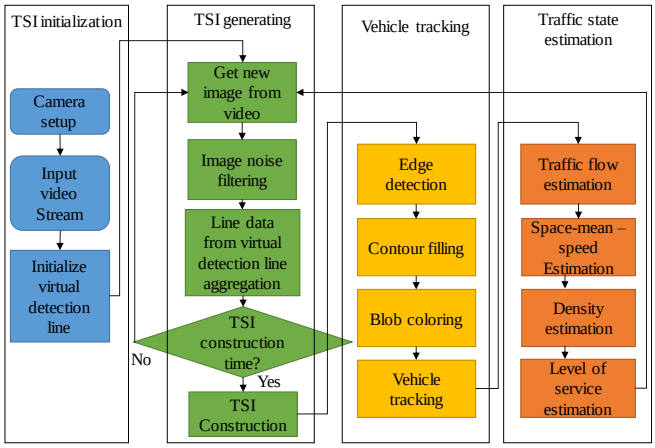


Fig. 1. Overall TSI processing framework

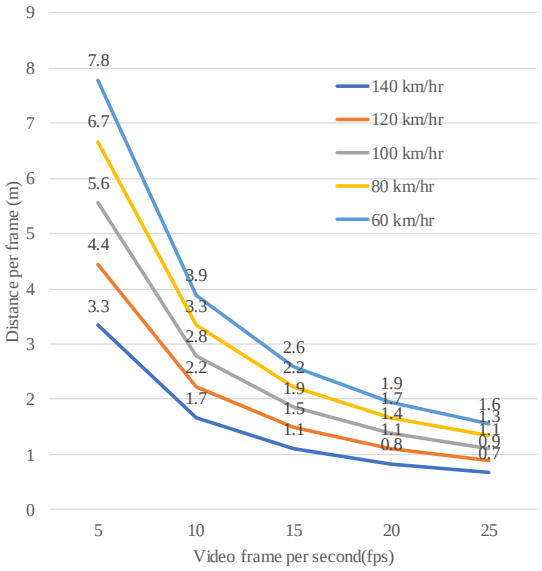


Fig. 2. Minimum detection length per video fps based on different maximum free-flow speeds

A. TSI Initialization

1) Camera setup

Before the TSI generating process, several camera settings and image pre-processing procedures are required for enhancing the quality of the video scene. Sometimes a vehicle may obscure the following vehicle due to the vehicle height. Therefore, this study has a restrict assumption on camera installation: the camera height and angle setting must be carefully adjusted and the vehicles in the video should not be obscured.

The camera should be installed in a high building or a gantry with a narrow camera angle to the vertical axis or similar angles to the top down angle. For image preprocessing, noise reduction is required [21]. Gaussian noise is adopted to reduce salt and pepper noise in the video data. In this study, the camera angle is restricted and carefully installed on a building to capture the traffic data.

2) Virtual detection line setup

A digital image is represented by a two-dimensional matric, where the X-axis and Y-axis represent image width and height in a pixel units, respectively (e.g. $W*H$). Therefore, an image from a video source at time T can be denoted as a color intensity matric $F_{T,W*H}$. An element $\vec{f}_{T,ij}$ is a vector representing the intensity of a pixel in the 3 primary color domains of red, green, and blue $\vec{f}_{T,ij} = \begin{bmatrix} r_{T,ij} & g_{T,ij} & b_{T,ij} \end{bmatrix}^T$ where i and j are location of the pixel in x-axis and y-axis respectively. The intensity value of each color is quantized into 256 values (i.e. $0 \le r_{T,ij} \le 255$) where zero represents the lowest color intensity or the darkest color, and vice versa.

To set a virtual detection line D^n where n is an index of the detection line D^n, two points at both ends of a line, (i_1, j_1) and (i_2, j_2), must be chosen from the $F_{T,W*H}$. For example, a virtual detection line D^1 can be drawn along an observing lane as a straight line as can be seen in Fig. 3a. The actual detection length of D^1 is denoted as l_{D^1} in meter units.

The only limitation of setting the virtual detection line is the length of the line. The virtual detection line must be longer than the minimum length which can be estimated based on 2 factors: the video frame per second (fps) and the maximum speed on the observed road. For defining the minimum length of the detection line, the maximum distance of a vehicle traveling at maximum speed in a video frame must be determined.

$$s_{\max} = \frac{v_{\max}}{3.6 * \lambda} \quad (1)$$

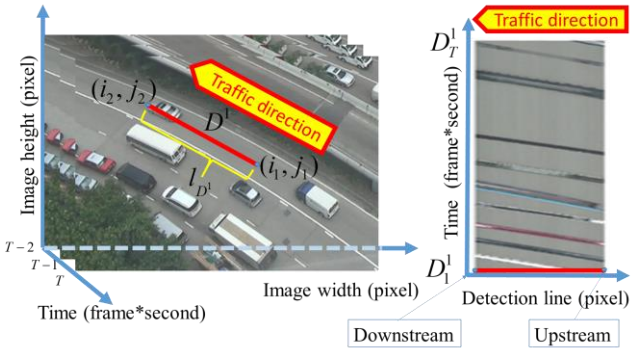
where $s_{\max}$ is the maximum distance of a vehicle traveling at maximum speed in a video frame (m/frame), $v_{\max}$ is the maximum speed (km/hr) and λ is the video frame per second (frame/s). Finally, the minimum length of the detection line is equal to $s_{\max}$.

Fig. 2 demonstrates the minimum length of the detection line based on various video fps and maximum vehicle speeds. In this study, the video fps is 25 and the maximum speed is 80 km/hr. Therefore, the detection line length l_{D^1} in Fig. 3a is approximately 24 m which is longer than the **0.9** m. The effect of different length detection lines will be discussed in section VI.

The color intensity value at each pixel on the detection line D_T^1 at time T will be used for constructing TSI and for road traffic data extraction in the next following sections.

B. TSI Generating Process

Time-spatial imagery (TSI) is a technique used to generate a two-dimensional image constructed with lines data over the period. TSI can be represented as a compact Time-Space diagram. Let P represent the number of video frames in a TSI construction period (frame). Therefore, $\Psi_{D_T^n * P}$ represents a



(a) (b)

Fig. 3. Virtual detection line setting (a) and TSI generated from $D1$ (b)

two-dimensional image of a TSI generated from a virtual detection line D_T^n while $t = \{1, 2, 3, ..., P\}$.

$$\Psi_{D_t^n * P} = \begin{bmatrix} D_1^n & D_2^n & . & . & D_P^n \end{bmatrix}^T \quad (2)$$

where (2) may be called the transformed image of a video stream period or TSI. The virtual detection line setting ensures that every vehicle in the detection lane is well recorded in the TSI apart from vehicles which changed lanes. The system captures the line data D_T^n of each time-frame of $t = \{1, 2, 3, ..., P\}$, then constructs the TSI as $\Psi_{D_T^n * P}$ as can be seen in Fig. 3b.

In this study, **frame-seconds** will be used as a unit of time instead of seconds for estimating road traffic information. As can be seen in Fig. 3b, the frame-second represents a unit of time corresponding to the video format image per second or frame per second (fps).

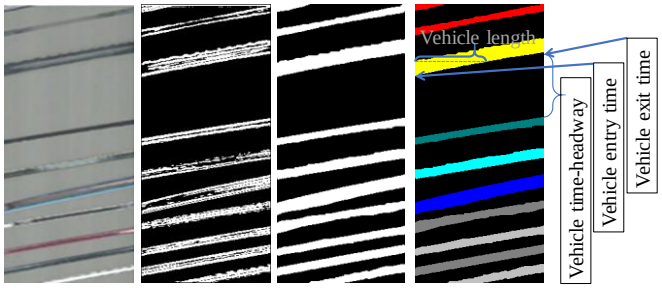
$$t = \frac{P}{\lambda} \quad (3)$$

where P is the total number of frames or TSI height (frame), t represents time (s) and λ is video frame per second (frame/s). For example, a video with a resolution of 25 fps generates a TSI of 25 pixels per second. This means that a minute of P frame will be 1,500 pixels long in the Y-axis for a 25fps video. Therefore, the units of a TSI in the Y-axis are frame-seconds, which can be transformed into seconds by using fps directly.

As can be seen in Fig. 3a, the D^1 detection line is established along the detection area. The right side and left side of the D^1 's TSI are the upstream and downstream, respectively. A vehicle that passed the detection zone from upstream to downstream will appear as a complete slope line from left to right in D^1 's TSI in Fig. 3b. A vehicle changing lane either out of or into D^1 will result in an incomplete slope line from left to right.

C. *Vehicle Tracking*

To estimate the traffic state from a TSI, individual vehicles need to be identified. The TSI is an image constructed from several line data D_T^n where $T = \{1, 2, 3, ..., P\}$ from (2) as



(a) (b) (c) (d)

Fig. 4. TSI (a), edge detection image of TSI (b), contour filling of TSI (c) and Blob colouring of TSI (d)

can be seen in Fig. 4a. First, the edge detection technique is implemented to detect the distinctiveness of an object's contour [22]. However, the texture within a TSI is changed due to the time dimension. Therefore, TSI edge detection $\mathrm{E}_{D_T^n * P}$ can be considered as an estimating edge based on the time dimension instead of the 2 dimensions of height and width which would be used in an ordinary 2-D image as can be seen in Fig. 4b. The $\mathrm{E}_{D_T^n * P}$ can be constructed using (4), (5) and (6).

$$\mathrm{E}_{D_T^n * P} = \begin{bmatrix} E_1^n & E_2^n & . & . & E_P^n \end{bmatrix}^T \quad (4)$$

$$E_t^n = D_t^n - D_{t-1}^n \quad (5)$$

$$e_{t,x}^n = \begin{cases} 0 & ;(d_{T,x}^n - d_{T-1,x}^n) < k \\ 1 & ; \quad otherwise \end{cases} \quad (6)$$

where E_t^n represents the edge detection line generated from line data subtraction in color of $D_t^n - D_{t-1}^n$. Each element $en_{t,x}$ is a binary value representing edge detection while $\{e_{T,1}^n, e_{T,2}^n, .., e_{T,X}^n\} \in E_t^n$, $d_{T,X}^n = \{0, 1, 2, 3, ..., 255\}$ is an average pixel value from $\overline{f}_{T,X}$ (grayscale value) to minimize the complexity of the image processing and $\{d_{T,1}^n, d_{T,2}^n, .., d_{T,X}^n\} \in D_t^n$.

Then the location of the vehicles in the TSI can be estimated. The dilate, contour filling and blob coloring technique [23] are adopted to identify the estimated distinctiveness of the vehicle. The dilated image will enlarge the edge detection line by an image filter of 3x3 instead of by 1 pixel to join the contours of the same vehicle together. Contour filling will be used for filling any holes in the edge detection process. Then blob coloring will show the distinct vehicles in the TSI as can be seen in Fig. 4c and Fig. 4d.

$$Blob\_coloring(Contour\_filling(\mathrm{E}_{D_T^n xP})) \quad (7)$$

where the vehicle location in the time axis within the TSI will be stored in a look-up-table.

The vehicle tracking data from Fig. 4c are vehicle counting, vehicle entry time, vehicle exit time, and time headway. The data sets will be stored in a data look up table for each time interval. The data table can be further used in the following

section for traffic state estimation including traffic flow, time-mean speed and density.

D. Traffic State Estimation

This section consists of 4 sub-sections including traffic flow estimation, space-mean-speed estimation, density estimation and level of service estimation. The flow and space-mean-speed can be derived from a TSI input while the density will be estimated indirectly from the time headway and space-mean-speed information.

1) Traffic flow estimation

Traffic flow is the number of vehicles passing a reference point per time unit. Since TSI is an image generated from a reference detection line for a period of time, the system can simply count the number of vehicles in the TSI and estimate the flow.

$$\bar{q} = \frac{3600\lambda}{T}\sum_{j=1}^{M} P_{n,j} \tag{8}$$

where $\bar{q}$ represents average flow (veh/hr/lane), λ is frame rate of the video (frame/s), M is number of vehicle in a TSI, $P_{n,j}$ is a vehicle index j in the TSI generated from detection line n, and T is the height of the TSI (pixels).

2) Traffic space-mean-speed estimation

Normally, there are lane markings along the road network, whereby the length between each lane marking is constant. The detection line length can be set up based on the lane markings' constant lengths.

The following assumptions are:

- All vehicles pass the detection zone without changing lane
- Each vehicle speed is constant within the detection zone.
- The vehicle must not stop at the virtual detection line longer than the TSI generating time. Otherwise, the result would be 0 km/hr.

The estimated speed from the TSI is a space-mean-speed since the TSI is a group of lines over time. The space-mean-speed is an average speed over the period.

$$V = M(\sum_{i=1}^{M}\frac{1}{v_i})^{-1} \tag{9}$$

where V is the space-mean-speed (km/hr), M is the number of vehicles (veh), and v_i represents the spot speed of an individual vehicle i (km/hr). In order to estimate v_i, the time occupancy must be estimated. The vehicle entry time and vehicle exit time are recorded in the TSI in frame-seconds. As mentioned in the previous section, the Y-axis represents the time. Therefore, the speed during each vehicle's occupancy in the detection area can be computed.

$$v_i = \frac{3.6 * l_{Dn} * \lambda}{|\Delta L_i|} \tag{10}$$

where l_{Dn} represents the length of detection line Dn (m), λ is video frames per second (frame/s), and ΔL_i is the occupancy time or the difference in the Y-axis of the same vehicle in two TSIs (pixels). Finally, the space-mean-speed can be calculated.

$$V = M(\sum_{i=1}^{M}\frac{|\Delta L_i|}{3.6 * l_{Dn} * F})^{-1} \tag{11}$$

3) Traffic density estimation

TSI contains information of vehicle **pixel-time-headway** as well. As can be seen in Fig. 4d, the gap between two consecutive vehicles can be estimated as time-headway.

$$\overline{h_p} = \frac{1}{M}\sum_{n=1}^{M-1}(L_n - L_{n+1}) \tag{12}$$

where $\overline{h_p}$ is the average pixel-time-headway (pixels), M represents the number of vehicles in each lane (veh*lane), and L_n is the location in the Y-axis of vehicle n. In order to transform the pixel-headway into time-headway, the frame-rate of the source video is applied for dividing the pixel-headway.

$$\overline{h_t} = \frac{\overline{h_p}}{3600 * \lambda} \tag{13}$$

where $\overline{h_t}$ denotes time-headway (hr), and λ is represented by video frames per second (frame/s). The estimated time-headway can be directly used for calculating density.

$$k = \frac{3600 * M * \lambda}{V\sum_{n=1}^{M-1}(L_n - L_{n+1})} \tag{14}$$

where k is the density (veh/km/lane), and $\bar{v}_s$ represents the calculated space-mean-speed (km/hr) calculated in the previous section.

4) Level of Service (LoS) estimation

Basically, the density can be used for estimating the level of service as listed in Table II. In practice, road users do not need the precisely estimated data to plan their journey. For example, the Hong Kong Transport Department provides the traffic speed on a tri-color basis which is similar to most countries. Moreover, Google Traffic maps also illustrate the traffic state using tri-color representation. Therefore, the level of services in this study will be categorized into 3 distinct groups. The levels of service A and B are grouped as free-flow cases. The moderate traffic condition is grouped as C and D levels of service, while high congestion consists of levels of service E and F.

TABLE II
LEVEL OF SERVICE BASED-ON DENSITY ON FREEWAY SEGMENT

HMC based		Combined from HCM	
Level of service	Density (veh/km/lane)	Level of service	Density (veh/km/lane)
A	0-7	A and B	0-11
B	7-11		
C	11-16	C and D	11-22
D	16-22		
E	22-28	E and F	>22
F	>28		

Source: Special Report 209: Highway Capacity Manual, Fourth Edition, Copyright 2000 by the Transportation Research Board, National Research Council, Washington, DC.

One of the traffic state characteristics is changing restrictions on the traffic state. A free-flow state cannot be suddenly changed into a high-congestion state, and vice-versa. However, a low-congestion traffic state can be developed to a free-flow or high-congestion state while both free-flow and high-congestion states can be developed to the low-congestion state only. This leads to a weight traffic level of service estimation.

A high-congestion case can be determined by the preprocessing stage. The jam density is 28 veh/km/lane while the density in free-flow and low-congestion cases is varied.

$$k_t = \alpha k_{t-1} + (1-\alpha)k_t$$
$$\alpha = \begin{cases} 0 \;; |k_{t-1} - k_t| > \phi \\ 0.5; \;\; otherwise \end{cases} \quad (15)$$

where k_t is density (veh/km/lane), t represents time index, α is weight, and ϕ is the supervised learning density that will trigger the sudden change from free-flow to high-congestion. In this case, the trained ϕ can be represented as 11 veh/km/lane. The system will not include the weight to recalculate the ϕ when $|k_{t-1} - k_t|$ is equal to or lower than ϕ. On the other hand, the high values of ϕ will lead to data misinterpretation for free-flow to high-density cases. The abnormal difference of density in two consecutive times can be smoothed with (15).

III. Empirical Study

In this section, the test site is described. Several experiments are carried out to evaluate the TSI processing performance in several aspects, e.g. data communication cost, flow estimation accuracy and space-mean-speed estimation accuracy.

A. Test Site Description

A 4-lane road segment of the Chatham Road South located in front of The Hong Kong Polytechnic University, Block Z campus is used as the test site in this study. The special characteristic of this road segment is that there is traffic congestion on two slow lanes during peak hours. Some vehicles change lane on this road segment. On the other hand, the traffic state of the other two fast lanes is rarely congested, even in the peak hours.

A video camera was installed on the 11th floor of Block Z facing toward the traffic scene as can be seen in Fig. 3a. This

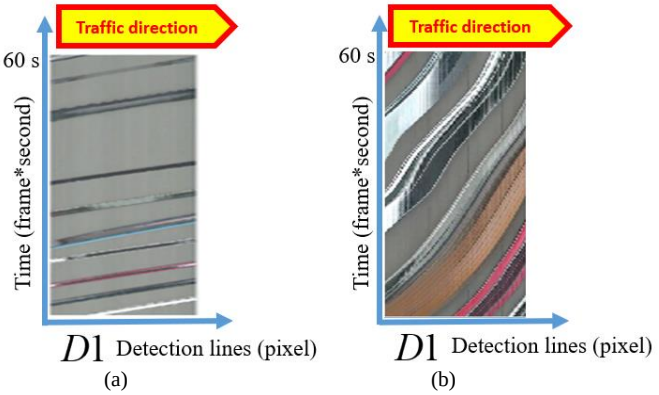


Fig. 5. Time Spatial Images generated from $D1$ during (a) free-flow, and (b) congestion traffic condition

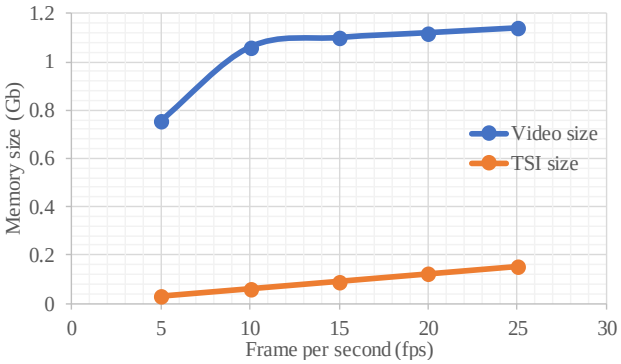


Fig. 6. Traditional video size comparison with TSI size from various video fps sources

study uses the video data recorded during one hour of the morning peak period. The light is varied during this period. The virtual detection lines were set on each lane. Therefore, the empirical results include various traffic conditions.

The length of each detection line is 24 m and the estimated interval is every 60 s including 60 s line data collection and ~0.005 second simultaneous processing during the TSI construction period.

The system is tested using different fps video inputs. The original video was transformed into several fps videos. The commercial speed-trap video sensor using 25-fps videos as input is considered as actual data for the TSI result validation.

B. TSI Characteristics

From the previous section, a virtual detection line captures and stores a line of each frame in the array data along a time-frame period. The arrays of each time-frame can be constructed into a TSI. As shown in Fig. 5a, the X-axis of the TSI represents the captured line at each frame while the Y-axis represents the time of the captured line in frame-seconds. For example, the frame rate of the video stream is 25 frames per second. The system acquires 25 pixels of height from each second. If the system sets the TSI image height equal to 1,500 pixels, the TSI generating time is a little over 30 seconds including the computational time. The TSI height corresponds to the time of the TSI constructing period as (3).

In cases where there are no vehicles passing the virtual detection line, the plain background of the pavement texture is captured in the TSI. The vehicles will be recorded in the TSI and the vehicles' shapes can be changed due to the vehicle behavior during different states and the location of different virtual detection lines.

The vehicles traveling in the free-flow traffic conditions tend to pass the detection line at a high speed as can be seen in Fig. 5a. The slope lines in $\Psi_{D1_T\,xP}$ represent vehicle speeds. The slope is gentle and constant most of the time because the vehicle speed is considerably high.

The TSIs generated during traffic congestion cases are shown in Fig. 5b. The slope of the lines in $\Psi_{D1_T\,xP}$ is uncertain owing to the variation of speeds on the detection line $D1$.

The length of the virtual detection line affects the resolution of the estimation. A longer detection line provides more estimation details which insignificantly affects the estimation

results. However, a longer detection line might introduce a higher rate of vehicles changing lanes than a shorter line.

C. Data Communication Cost and Processing Time

Most of the traffic monitoring systems that process the traffic state data at a back-end server need to transmit the video stream to the server. On the other hand, the TSI approach transmits a single image over a time period. The required memory is, therefore, significantly reduced, compared with the back-end processing or centralized systems as can be seen in Fig. 6. The TSI technique uses approximately one tenth of the data size required by other systems. For 5-fps videos, the memory size is significantly reduced.

To implement the system on a real-time basis, multi-thread processing can be performed. The line data collection procedure will not be interrupted by the traffic state estimation process, which will perform its tasks periodically. The computational time is less than 0.5 s using a computer with Intel Core i7-7700HQ, 16 GB RAM. Therefore, the estimation interval depends on the video fps resolution plus the processing time (less than a second).

D. Traffic State Estimation Results Using Various Frames per Second

Suppose that the results from a traditional video-based speed-trap using an Autoscope sensor provide a ground truth dataset. The accuracy of the Autoscope system is more than 98% [5]. The Mean Absolute Percentage Error (MAPE) can be estimated as:

$$MAPE(\%) = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{16}$$

where n is the total number of data points, A_t is the data generated from Autoscope using 25-fps videos, and F_t is the estimated value from the proposed system.

The recorded entry and exit times are used for speed and flow estimation. The 25-fps video is used for result validation. As can be seen in Fig. 7, the TSI technique outperforms the traditional approach in terms of flow estimation. The video fps directly influences the accuracy of the traditional approach. This is because the proposed system estimates the flow data by counting vehicles in the TSI, while the speed-trap method may encounter a problem of vehicle misdetection due to the low video fps. The lower video fps means a lower sampling rate. With the low sampling rate, several vehicle entry times or exit times are not registered in the speed-trap system. For the space-mean-speed, the traditional approach provides higher accuracy with the higher video fps. However, the TSI method provides better results when the video fps is lower than 15-fps. For video at 5-fps, the TSI method is still able to estimate accurate results while the results of the traditional method are poorly estimated.

E. Level of Service (LoS) Results

The proposed system can roughly classify the level of services into 3 groups: “A and B”, “C and D” and “E and F”. From the calculated results, the accuracy of LoS classification

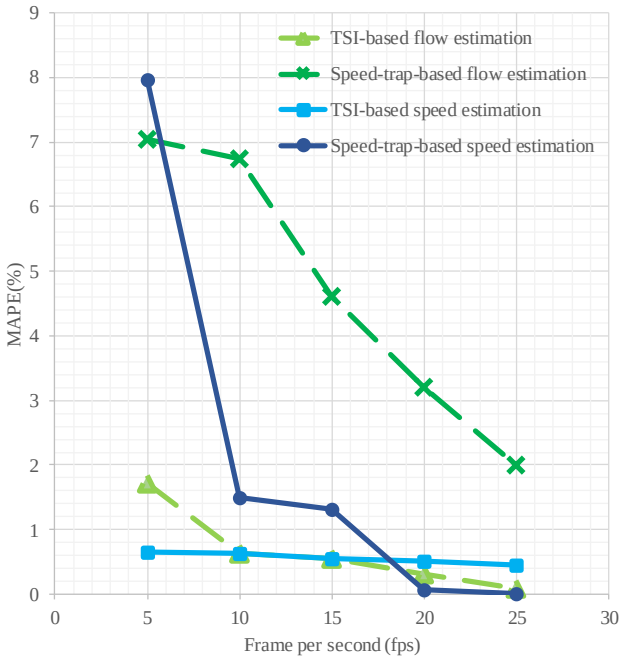


Fig. 7. MAPE comparison of flow and space-mean-speed estimation using TSI technique and the traditional video-based speed-trap technique

is fairly high at 97% for “A and B”, 94% for “C and D” and 93% for “E and F” from a dataset of one hour. However, higher traffic congestion leads to higher estimation error. First, errors can be caused by LoS misclassification. For instance, the calculated speed in heavy congestion cases may lead to the LoS for D or E which are not classified into the same LoS group. Second, the LoS misclassification can be caused by incomplete vehicle trajectories in a TSI. A vehicle trajectory in a TSI can be extended to the next TSI in the traffic congestion state.

There are several ways to improve the accuracy in this respect. One method is to change the TSI to Time-Spatial Stream (TSS) processing. TSS is the same concept as TSI but TSI height is dynamic whereas TSS processing provides an opportunity to capture the road traffic information with lesser deflected results. However, the dynamic TSI height in the TSS approach consumes more computational time during the TSI generating process of estimating the vehicle occupancy within the virtual detection line. Moreover, temporal TSI processing can be used to estimate the vehicles that are separated into two consecutive TSIs. This approach needs to match the same vehicle in the two TSIs at the current time step and previous time step. Nonetheless, both approaches require more computational cost.

F. Future research on vehicle classification

Vehicle classification can be developed in future research based on vehicle length estimation. As can be seen in Fig. 4d, the vehicle length information that is also recorded in the TSI assumes that the vehicles do not change lanes while they are passing the VDL.

To estimate the actual length, the positions of the vehicle front and rear bumper need to be identified in pixel units, which then need to be transformed into real-world dimension units

based on the VDL length and the actual length of the VDL on the road segment.

IV. Conclusion

This paper proposes an efficient time-spatial image (TSI) processing technique for estimating real-time road traffic information. The TSI processing module can be integrated with CCTV cameras which have been installed on the road network for traffic surveillance purposes. Therefore, the unoccupied CCTVs can be opportunistically used to estimate the road traffic state and traffic information, such as flow, space-mean-speed, and density. The TSI shows promising results in terms of data communication and processing costs. Moreover, the TSI can efficiently classify the traffic state of the road network into various levels of service for the illustration of congestion levels on the test site.

The system can be delivered as several real-world applications, e.g. speed map and density map, since these speed-density maps normally use a tri-color system to represent the traffic states.

However, there are several requirements for further improvement of TSI processing, such as the consideration of traffic speed and vehicle length in the TSI relationship, and vehicle classification using TSI. In addition, the current TSI processing system can be enhanced with some image processing techniques, e.g. shadow removal algorithms and 3-dimensional object learning from TSI. Finally, Time-Spatial Stream (TSS) processing might be worthy of further investigation in order to explore its potential for improving the system accuracy by involving the temporal effect of partial vehicle trajectory in TSI.

V. Acknowledgements

The work described in this paper was jointly supported by research grants from the Research Committee of the Hong Kong Polytechnic University (Project No. 1-ZVFJ) and the Research Institute of Sustainable Urban Development of the Hong Kong Polytechnic University (Project No.1-ZVBX and 1- ZVBZ).

References

[1]. R. Arnott, A. D. Palma, and R. Lindsey, "Does providing information to drivers reduce traffic congestion?," *Transportation Research Part A: General*, vol. 25, no. 5, pp. 309–318, 1991.

[2]. F. Calabrese, M. Colonna, P. Lovisolo, D. Parata and C. Ratti, "Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome", IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 1, pp. 141-151, 2011.

[3]. R.P. Schäfer, K.U. Thiessenhusen, E. Brockfeld, and P. Wagner. "A traffic information system by means of real-time floating-car data." (2002).

[4]. V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *Image and Vision Computing*, vol. 21, no. 4, pp. 359–381, 2003.

[5]. P. Michalopoulos, "Vehicle detection video through image processing: the Autoscope system," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1, pp. 21–29, 1991.

[6]. J.C. Tai and K.T. Song, "Background segmentation and its application to traffic monitoring using modified histogram," *IEEE International Conference on Networking, Sensing and Control, 2004*.

[7]. L. Wei and D. Hong-Ying, "Real-time Road Congestion Detection Based on Image Texture Analysis," *Procedia Engineering*, vol. 137, pp. 196–201, 2016.

[8]. M. H. Ali, S. Kurokawa, and A. Shafie, "Autonomous Road Surveillance System: A Proposed Model for Vehicle Detection and Traffic Signal Control," *Procedia Computer Science*, vol. 19, pp. 963–970, 2013.

[9]. J. Wu, Z. Liu, J. Li, C. Gu, M. Si, and F. Tan, "An algorithm for automatic vehicle speed detection using video camera," *2009 4th International Conference on Computer Science & Education*, 2009.

[10]. D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russell, "Towards robust automatic traffic scene analysis in real-time," *Proceedings of 12th International Conference on Pattern Recognition*.

[11]. Z. Yang, H. Meng, Y. Wei, H. Zhang, and X. Wang, "Tracking Ground Vehicles in Heavy-traffic Video by Grouping Tracks of Vehicle Corners," *2007 IEEE Intelligent Transportation Systems Conference*, 2007.

[12]. P. Kumar, S. Ranganath, H. Weimin, and K. Sengupta, "Framework for Real-Time Behavior Interpretation From Traffic Video," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 43–53, 2005.

[13]. A. Albiol, V. Naranjo, and I. Mora, "Real-time high density people counter using morphological tools," *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*.

[14]. L. Li, L. Chen, X. Huang, and J. Huang, "A Traffic Congestion Estimation Approach from Video Using Time-Spatial Imagery," *2008 First International Conference on Intelligent Networks and Intelligent Systems*, 2008.

[15]. Y. Yue, "A Traffic-Flow Parameters Evaluation Approach Based on Urban Road Video," *International Journal of Intelligent Engineering and Systems*, vol. 2, no. 1, pp. 33–39, 2009.

[16]. Y. Zhang, C. Zhao, and Q. Zhang, "Counting vehicles in urban traffic scenes using foreground time-spatial images," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 61–67, Jan. 2017.

[17]. C. Li, K. Ikeuchi, and M. Sakauchi, "Acquisition of traffic information using a video camera with 2D spatio-temporal image transformation technique," *Proceedings 199 IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems (Cat. No.99TH8383)*.

[18]. "Traffic Monitoring Guide", Federal Highway Administration, U.S. Department of Transportation, Feb. 13, 2018. [Online]. Available: https://www.fhwa.dot.gov/policyinformation/tmguide/tmg_fhwa_pl_17_003.pdf

[19]. B. Coifman, "Estimating travel times and vehicle trajectories on freeways using dual loop detectors," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 4, pp. 351–364, 2002.

[20]. J. Wang, E. Case, and D. Manor, "The Road Traffic Microwave Sensor (RTMS)," *The 3rd International Conference on Vehicle Navigation and Information Systems*.

[21]. J.S. Lee, "Digital Image Enhancement and Noise Filtering by Use of Local Statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 2, pp. 165–168, 1980.

[22]. J. Canny, "A Computational Approach to Edge Detection," *Readings in Computer Vision*, pp. 184–203, 1987.

[23]. J.H. Ahn, and J.H. Kim, "A Stable Hand Tracking Method by Skin Color Blob Matching," PSR Journal, Pacific Science Review, 12(2), pp.146-151, 2010.

[24]. Bachmann, M. Roorda, B. Abdulhai and B. Moshiri, "Fusing a Bluetooth Traffic Monitoring System With Loop Detector Data for Improved Freeway Traffic Speed Estimation", Journal of Intelligent Transportation Systems, vol. 17, no. 2, pp. 152-164, 2013

Ekalux Ua-areemitr received a B.Eng. degree in computer engineering from King Mongkut's University of Technology Thonburi, Bangkok, Thailand, in 2010. He is currently working towards the Ph.D. degree in transportation engineering at The Hong Kong Polytechnic University, Hong Kong, China.

His research interests include network sensors, intelligent transportation systems, artificial intelligence and image processing.

Agachai Sumalee received a B.Eng. degree in civil engineering from KMITL-Ladkrabang in 1999 and a M.Sc.(Eng.) and Ph.D. degrees in transportation planning and engineering from the University of Leeds, Leeds, U.K., in 2000 and 2004, respectively. He is a Professor with the Department of Civil and Environmental Engineering at The Hong Kong Polytechnic University, Kowloon, Hong Kong.

His research areas include transit planning, intelligent transport systems, network modeling and optimization, and transport economics. Prof. Sumalee is currently the Editor-in-Chief of Transportmetica B: Transport Dynamics; the Editor of the Journal of Transport Policy, an Associate Editor of Networks and Spatial Economics; and an Editorial Board Member of several prestigious journals.

William H. K. Lam received B.S. and M.S. degrees from the University of Calgary, Calgary, AB, Canada, in 1978 and 1980, respectively, and a Ph.D. degree in transportation engineering from the University of Newcastle upon Tyne, Newcastle, U.K., in 1992. He is currently a Chair Professor in civil and transportation engineering with the Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong. He is also with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China.

His current research interests include transport network modeling, traffic simulation, public transport, and pedestrian studies. Prof. Lam is a member of the International Advisory Committee of the International Symposium on Transportation and Traffic Theory (ISTTT) and a member of the International Scientific Committee of the International Symposium on Transportation Network Reliability (INSTR). He is also the Founding Editor-in-Chief of Transportmetrica A and the Coeditor-in-Chief of the Journal of Advanced Transportation.