Real-time Traffic Pattern Analysis and Inference with Sparse Video Surveillance Information

Yang Wang[1,2], **Yiwei Xiao**[2], **Xike Xie**[1], **Ruoyu Chen**[2], **and Hengchang Liu**[1]
[1] School of Computer Science and Technology, University of Science and Technology of China
[2] School of Software Engineering, University of Science and Technology of China
{angyan, xkxie, hcliu}@ustc.edu.cn, {xiaoyw, chenry}@mail.ustc.edu.cn

Abstract

Recent advances in video surveillance systems enable a new paradigm for intelligent urban traffic management systems. Since surveillance cameras are usually sparsely located to cover key regions of the road under surveillance, it is a big challenge to perform a complete real-time traffic pattern analysis based on incomplete sparse surveillance information. As a result, existing works mostly focus on predicting traffic volumes with historical records available at a particular location and may not provide a complete picture of real-time traffic patterns. To this end, in this paper, we go beyond existing works and tackle the challenges of traffic flow analysis from three perspectives. First, we train the transition probabilities to capture vehicles' movement patterns. The transition probabilities are trained from third-party vehicle GPS data, and thus can work in the area even if there is no camera. Second, we exploit the Multivariate Normal Distribution model together with the transferred probabilities to estimate the unobserved traffic patterns. Third, we propose an algorithm for real-time traffic inference with surveillance as a complement source of information. Finally, experiments on real-world data show the effectiveness of our approach.

1 Introduction

The proliferation of urban video surveillance systems gives prominence to advanced traffic services, from personalized route optimization to macrolevel traffic administration. Such surveillance data applications include urban vehicle driving optimization [Gonzalez *et al.*, 2007; Schmitt and Jula, 2006; Leduc, 2008], road network traffic flow analysis [Bas *et al.*, 2007; Liu *et al.*, 2013; Suzuki and Nakamura, 2006], and intelligent transportation realization [Zhang *et al.*, 2011; Wang, 2010; Lu *et al.*, 2009].

Most traffic analysis with such surveillance systems assumes a dense camera distributions and so as a full coverage of road network traffic surveillance. However, the sparsity of camera distribution can hardly be avoided in real scenarios, due to the high deployment overheads and the dynamic nature of urban road networks. On the other hand, there is an increasing need for utilizing such camera surveillance data for traffic inference, especially for most small and medium-sized cities of China, where floating vehicles are not well employed and therefore the solutions with historical trajectory collections are not applicable. For instance, in Suzhou, a leading city in deploying urban traffic monitoring systems, there are 107 intersections in the industrial park while only 41.1% intersections are equipped with traffic surveillance systems.

There also exist extensive studies on forecasting traffic conditions with time series analysis. However, such seemingly related techniques fall short in making predictions with the incomplete information caused by the partial coverage of traffic surveillance. We can summarize existing works of this field into two categories, *neural network based methods* [Yasdi, 1999; Dia, 2001] and *spatial topology based methods* [De Fabritiis *et al.*, 2008; Min and Wynter, 2011].

Neural network based methods predict a road segment's traffic flow by matching the real time information with historical patterns. In particular, Yasdi et al. [Yasdi, 1999] utilize Jordan architecture based neural networks and Dia et al. [Dia, 2001] develop time-lag recurrent neural networks. Nevertheless, such predictions are done based on the long-term data accumulation of the target road segment. Therefore, it is not clear how to extend them to support road networks that are only partially surveilled.

Spatial topology based methods mostly consider the correlation between interconnected road segments. Particularly, Fabritiis et al. [De Fabritiis *et al.*, 2008] predict short-term (15 to 30 minutes) road travel speeds of the Italian motorway network with historical and real-time floating-car-collected traffic speed information, and Min et al. [Min and Wynter, 2011] define a time lag function to describe the spatiotemporal correlations of a road segment and its counterparts in the neighborhood. Without exception, these spatial topology methods cannot abstract the correlations between road segments which are not surveilled either.

In summary, previous works on traffic volume prediction take the completeness of the road network surveillance as an essential ingredient. It is not clear how prediction can be made if the some parts of the surveillance is missing. In this work, we tackle the challenge in three folds. First, we model the traffic volume of the entire road network with transferred transition probabilities from a third-party GPS dataset. Second, we use a model that takes transition probabilities as

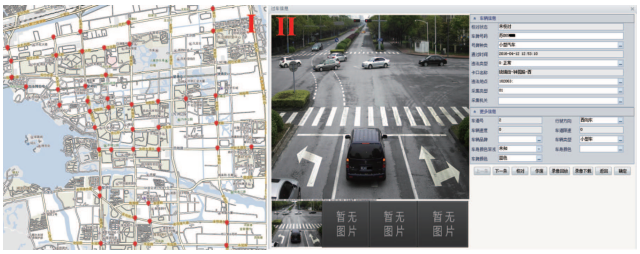

Figure 1: Examples (I. Surveillance Camera Distribution; II. Before-/After Camera Data Processing of Vehicles)

inputs to make the incomplete surveillance space approximately complemented. The model follows multivariate normal distributions, which are considered as a general stochastic method for capturing travel behaviours on road networks [Lam and Xu, 1999]. Also, with extensive data analysis, we observe that the transferred probabilities rest with time varying traffic contexts, e.g., workdays and weekends, rush and non-rush hours. It means that the utilization of the probabilities has to contend with specific spatiotemporal contexts. Third, we propose a novel approach to accurately infer the real-time traffic volume with only partially camera-equipped road networks. The results are cross-validated with real data.

Our work is based on a real project, SkyEye, in collaboration with Suzhou Traffic Police Department. To our best knowledge, this is the first work on inferring real-time traffic volumes with sparse road video cameras. The distribution of traffic surveillance cameras and exemplified surveillance information is shown in Figure 1. The GPS data is collected from 4,303 taxicabs of a year, i.e., from May, 2015 to June, 2016. We also collected all the video surveillance information with 44 camera-equipped road intersections of two-month, including the April and May of 2016. Experiments show that our proposal can improve the estimation accuracy by up to 30%.

The rest of this paper is organized as follows. Section II introduces preliminaries and formalizes the problem. Section III investigates our technical proposals including the inference algorithm. Section IV presents empirical studies and Section V concludes the paper.

2 Problem Definition

In this section, we formally define basic concepts as well as the problem studied in the work.

Definition 1 (Road Network.) *A road network can be modeled as a directed graph $G(\mathcal{V}, \mathcal{E})$, where the vertex set $\mathcal{V}$ denotes all the road intersections and the edge set $\mathcal{E}$ refers to all road segments. Given two road intersections $v_i, v_j \in \mathcal{V}$, $e_{ij} \in \mathcal{E}$ is the road segment between v_i and v_j.*

In SkyEye, traffic surveillance is implemented by cameras that equipped on the road intersections. The captured image and videos are then analyzed and parsed, as shown in Figure 1. Accordingly, the traffic volume statistics can be obtained. The intersections can thus be classified into two sets, camera-equipped intersections $\mathcal{V}_c$ and camera-free intersections $\overline{\mathcal{V}_c}$, satisfying $\mathcal{V}_c \cup \overline{\mathcal{V}_c} = \mathcal{V}$ and $\mathcal{V}_c \cap \overline{\mathcal{V}_c} = \emptyset$.

Definition 2 (Intersection Traffic Volume.) *Regarding intersection v_j, its traffic volume information in a specific time interval Δt can be formulated by $V_j(\Delta t)$, meaning the traffic volume of the intersection v_j during Δt is V_j vehicles per time unit.*

Definition 3 (Trajectories.) *Considering a moving object o, e.g., a taxicab, the trajectory of o within a given time interval Δt can be formulated by $T_o(\Delta t) = \{v_1(t_1), v_2(t_2), \cdots, v_k(t_k)\}_{t_i \in \Delta t}$, which means object o starts its journey at intersection v_1 of time t_1 and ends at intersection v_k of time t_k. Here, $v_i(t_i)$ refers to the snapshot of time t_i that trajectory T_o moving across intersection v_i.*

Notice that trajectories and traffic flows vary with different settings of time intervals Δt. We define trajectories and traffic volume in accordance with the time-varying urban traffic features, because it is commonly accepted that the traffic flows change regularly over time in practice [Wang *et al.*, 2014; 2016; Zheng *et al.*, 2010; Yuan *et al.*, 2013; 2012] [1].

Definition 4 (Inference with Incomplete Surveillance.) *Given road network $G(\mathcal{V}, \mathcal{E})$ and intersections under surveillance $\mathcal{V}_c$, for any given time interval Δt, our purpose is to design a method such that the traffic volume $V_i(\Delta t)$ of intersection $(v_i \in \overline{\mathcal{V}_c})$ without surveillance can be estimated.*

The quality of the estimation or inference is measured by Equation (1). Here, $\widehat{V}_i$ denotes the estimated value of traffic volume by our method. For a camera-equipped intersection, the accuracy is equal to 100%. For a camera-free case, the accuracy measure the ratio of the real value to the estimated value and is normalized to 1 by setting the dominator as summation of real value and the estimation error.

$$Inference\ \ Accuracy = \frac{V_i(\Delta t)}{V_i(\Delta t) + |\widehat{V}_i(\Delta t) - V_i(\Delta t)|} \quad (1)$$

3 Real-time Traffic Volume Inferring Algorithm

In this section, we propose the solution framework for the traffic analysis and inference problem. Figure 2 illustrates the architecture of our solution which consists of three major components: 1) training transition probabilities from third-party datasets; 2) unobserved traffic pattern completion with model-based approximations; 3) read-time traffic inference algorithms. We detail each step in the following subsections.

3.1 Time-varying Traffic Influence Models Between Neighboring Intersections

Now we model the transition probability between road intersections. Suppose intersection v_i is a m-way road intersections. To find the probability distribution of an incoming vehicle of v_i to other $m-1$ directions within a specific period,

[1] The setting of Δt should balance the tradeoff between the accuracies and temporal granularity. In our implementation, we slice the temporal information into slots of 30 minutes following the common settings. Notice that such a setting may be related to the results of inference accuracy but is orthogonal to the generalities of our proposals

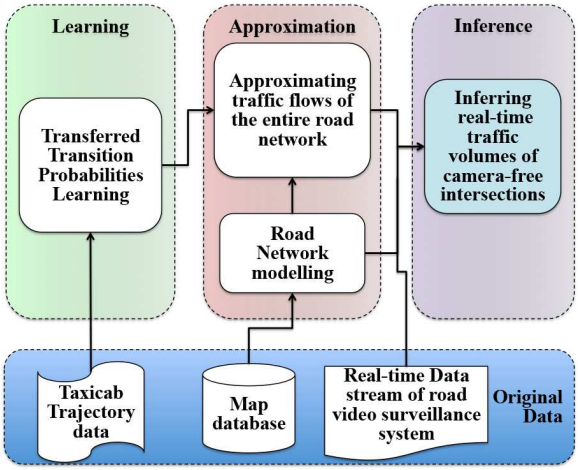

Figure 2: Algorithm overview

e.g., Δt, helps to infer the traffic patterns of v_i's neighboring intersections.

Let v_i and v_j be two adjacent intersections. The transition probability from v_i to v_j, denoted as $p_{\Delta t}(v_i, v_j)$, represents the probability of a vehicle moving from intersection v_i to v_j within time interval Δt, and thus enables quantifying the traffic volume of v_j, given that all v_j's neighboring intersections are known. In our implementation, we adopt an order-1 Markov transition probability matrix, $\mathbf{B}_{\Delta t}$, for summarizing transition probabilities of all pairs of road intersections in the urban road network, as shown in Equation (2).

$$\mathbf{B_{\Delta t}} = \begin{bmatrix} p_{\Delta t}(v_1,v_1) & \cdots & p_{\Delta t}(v_1,v_k) & \cdots & p_{\Delta t}(v_1,v_{|\mathcal{V}|}) \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ p_{\Delta t}(v_j,v_1) & \cdots & p_{\Delta t}(v_j,v_k) & \cdots & p_{\Delta t}(v_j,v_{|\mathcal{V}|}) \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ p_{\Delta t}(v_{|\mathcal{V}|},v_1) & \cdots & p_{\Delta t}(v_{|\mathcal{V}|},v_k) & \cdots & p_{\Delta t}(v_{|\mathcal{V}|},v_{|\mathcal{V}|}) \end{bmatrix} \tag{2}$$

We assume an acyclic road network such that the transition probability from an intersection to itself equals 0, i.e., $p_{\Delta t}(v_j, v_j) = 0$. Also, for two intersection v_j and v_k, if they are not adjacent, the corresponding transition probabilities equal to 0, i.e., $p_{\Delta t}(v_j, v_k) = p_{\Delta t}(v_k, v_j) = 0$. According to our analysis, the traffic patterns in weekdays and weekends are significantly different. So, the traffic influence is modeled for them separately.

The transition probabilities are trained with third-party GPS datasets in our implementation. With transition probabilities, the mutual influence between road intersections are established. In the sequel, we study how the trained probabilities are further used for estimating unobserved intersections.

3.2 Traffic Volume Analysis of Entire Road Network

As concluded in [Lam and Xu, 1999], the traffic volume of road networks follows a MND (Multivariate Normal Distribution). Recall that earlier, we introduce the time-varying properties of traffic volumes. So, we assume, for Δt and v_j, the traffic volume is $\widehat{V_j(\Delta t)}$, $j = 1 \cdots |\mathcal{V}|$, where V_j can be calculated by:

$$\widehat{V_j(\Delta t)} = \begin{pmatrix} \mu_{v_j}(\Delta t) + \varepsilon_{v_j}(\Delta t) + \\ \sum_{v_k \in N(v_j)} p_{\Delta t}(v_k, v_j) * V_k(\Delta t) \end{pmatrix} \tag{3}$$

where $\varepsilon_{v_j}(\Delta t) \sim N(0, (\sigma_{v_j}^{\Delta t})^2)$, and $N(v_j)$ is the neighbor intersection set of intersection v_j.

For the entire road network and Δt, we use a vector $\mathbf{TV}$ to denote the traffic volumes of all intersections. After substituting elements with Equation (3), we have:

$$\mathbf{TV}(\Delta t) = \mathbf{M}(\Delta t) + \mathbf{B}_{\Delta t} * \mathbf{TV}(\Delta t) + \mathbf{E}(\Delta t) \tag{4}$$

where $\mathbf{M}(\Delta t) = [\mu_{v_1}(\Delta t), \mu_{v_2}(\Delta t), \cdots, \mu_{v_{|\mathcal{V}|}}(\Delta t)]^T$, and $\mathbf{E}(\Delta t) = [\varepsilon_{v_1}(\Delta t), \varepsilon_{v_2}(\Delta t), \cdots, \varepsilon_{v_{|\mathcal{V}|}}(\Delta t)]^T$. Based on the definition of MND, we can have:

$$\mathbf{TV}(\Delta t) \sim N \begin{pmatrix} (1 - \mathbf{B}_{\Delta t})^{-1}\mathbf{M}(\Delta t), \\ \begin{pmatrix} (1 - \mathbf{B}_{\Delta t})^{-1} * \\ \mathbf{\Upsilon}(\Delta t)((1 - \mathbf{B}_{\Delta t})^T)^{-1} \end{pmatrix} \end{pmatrix} \tag{5}$$

where $\mathbf{\Upsilon}(\Delta t)$ is the diagonal matrix with elements of $(\sigma_{v_j}^{\Delta t})^2$'s:

$$\mathbf{\Upsilon}(\Delta t) = \begin{bmatrix} (\sigma_{v_1}^{\Delta t})^2 & 0 & \cdots & 0 \\ 0 & (\sigma_{v_2}^{\Delta t})^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\sigma_{v_{|\mathcal{V}|}}^{\Delta t})^2 \end{bmatrix} \tag{6}$$

Note that we have already calculated the matrix $\mathbf{B}_{\Delta t}$ in the last subsection. So far, we have derived the equation of $TV(\Delta t)$ whose input parameters can be calculated by Equations (2) and (6).

As we have analyzed, the road intersections can be with or without camera surveillance. Let the number of intersections with camera be K. We use a $K \times |\mathcal{V}|$ indicator matrix $\mathbf{P}$ to indicate the existence of camera surveillance. The equation of $\mathbf{P}$ is as follows.

$$\mathbf{P} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1|\mathcal{V}|} \\ a_{21} & a_{22} & \cdots & a_{2|\mathcal{V}|} \\ \vdots & \vdots & \vdots & \vdots \\ a_{K1} & a_{K2} & \cdots & a_{K|\mathcal{V}|} \end{bmatrix} \tag{7}$$

If an intersection v_j is camera-equipped, the element a_{jj} is set to 1. In total, K columns are set to 1 because there are K intersections are camera-equipped. Defining as such, we can denote the observed traffic volume as below.

$$\begin{aligned} \mathbf{TV}_{OBS}(\Delta t) &= \mathbf{P} \times \mathbf{TV}(\Delta t) \\ &\sim N \begin{pmatrix} \mathbf{P}(1 - \mathbf{B}_{\Delta t})^{-1}\mathbf{M}(\Delta t), \\ \begin{pmatrix} \mathbf{P}(1 - \mathbf{B}_{\Delta t})^{-1} * \\ \mathbf{\Upsilon}(\Delta t)((1 - \mathbf{B}_{\Delta t})^T)^{-1}\mathbf{P}^T \end{pmatrix} \end{pmatrix} \end{aligned} \tag{8}$$

The purpose here is to estimate the unknown parameters $\mu_{v_i}(\Delta t)$ and $(\sigma_{v_i}^{\Delta t})^2$. To do that, we utilize maximum likelihood estimation:

$$L(TV_{OBS}(\Delta t)) \propto \left| \begin{array}{c} \mathbf{P}(1-\mathbf{B}_{\Delta t})^{-1} * \\ \boldsymbol{\Upsilon}(\Delta t)((1-\mathbf{B}_{\Delta t})^T)^{-1}\mathbf{P}^T \end{array} \right|^{-\frac{1}{2}} * \exp \left\{ \begin{array}{c} -\frac{1}{2}(\mathbf{TV}_{OBS}(\Delta t) - \mathbf{P}(1-\mathbf{B}_{TP})^{-1}\mathbf{M}(\Delta t))^T * \\ \left(\begin{array}{c} \mathbf{P}(1-\mathbf{B}_{\Delta t})^{-1} * \\ \boldsymbol{\Upsilon}(\Delta t)((1-\mathbf{B}_{\Delta t})^T)^{-1}\mathbf{P}^T \end{array} \right)^{-1} \\ *(\mathbf{TV}_{OBS}(\Delta t) - \mathbf{P}(1-\mathbf{B}_{\Delta t})^{-1}\mathbf{M}(\Delta t)) \end{array} \right\} \tag{9}$$

Very likely, we cannot estimate all unknown $\mu_{v_i}(\Delta t)$ and $(\sigma_{v_i}^{\Delta t})^2$ since the number of unknown parameters is more than the number of observations. Therefore, we group the intersections in nearby regions and make those camera-free intersections have the same $\mu_{v_i}(\Delta t)$ and $(\sigma_{v_i}^{\Delta t})^2$ value to the nearest camera-equipped intersections to them. By solving the maximum likelihood in Equation (9), we can have the estimation results $\widehat{\mathbf{M}(\Delta t)}$ and $\widehat{\boldsymbol{\Upsilon}(\Delta t)}$.

3.3 Inferring Real-time Traffic Volumes of Camera-free Intersections

We have shown how unknown parameters can be estimated and how the MND for the road network traffic volume can be modeled. Now, we describe the mechanism which can infer real-time traffic volumes of camera-free intersections with the approximated MND. We first show how to use the calculated $TV(\Delta t)$ to infer the real-time traffic volumes of camera-free intersections. For simplicity, we formally define $\mathbf{r}$ and $\boldsymbol{\Gamma}$:

$$\begin{cases} \mathbf{r}(\Delta t) = (1-\mathbf{B}_{\Delta t})^{-1}\widehat{\mathbf{M}(\Delta t)} \\ \boldsymbol{\Gamma}(\Delta t) = (1-\mathbf{B}_{\Delta t})^{-1}\widehat{\boldsymbol{\Upsilon}(\Delta t)}((1-\mathbf{B}_{\Delta t})^T)^{-1}) \end{cases} \tag{10}$$

then we have $\mathbf{TV}(\Delta t) \sim N(\mathbf{r}(\Delta t), \boldsymbol{\Gamma}(\Delta t))$. After generating the MND, we can proceed to calculate the real-time traffic volumes of intersections without fixed video cameras by computing conditional expectation of MND. We first re-arrange the matrix $\mathbf{TV}(\Delta t)$ as:

$$\mathbf{TV}_{RE}(\Delta t) = \left[\begin{array}{c} \mathbf{TV}_{MISSING}(\Delta t) \\ \mathbf{TV}_{OBS}(\Delta t) \end{array} \right] \tag{11}$$

where there are $|\mathcal{V}| - K$ and K elements in $\mathbf{TV}_{MISSING}(\Delta t)$ and $\mathbf{TV}_{OBS}(\Delta t)$ respectively, and the goal is to compute $E[\mathbf{TV}_{MISSING}(\Delta t)\,|\mathbf{TV}_{OBS}(\Delta t)]$. According to the re-arranged matrix $\mathbf{TV}_{RE}(\Delta t)$, we also re-arrange matrix $\mathbf{M}(\Delta t)$ and $\boldsymbol{\Upsilon}(\Delta t)$ as $\mathbf{M}_{RE}(\Delta t)$ and $\boldsymbol{\Upsilon}_{RE}(\Delta t)$ respectively, then partition $\mathbf{r}_{RE}(\Delta t)$ and $\boldsymbol{\Gamma}_{RE}(\Delta t)$ as:

$$\mathbf{r}_{RE}(\Delta t) = \left[\begin{array}{c} \mathbf{r}_1(\Delta t) \\ \mathbf{r}_2(\Delta t) \end{array} \right] \tag{12}$$

$$\boldsymbol{\Gamma}_{RE}(\Delta t) = \left[\begin{array}{cc} \boldsymbol{\Gamma}_{11}(\Delta t) & \boldsymbol{\Gamma}_{12}(\Delta t) \\ \boldsymbol{\Gamma}_{21}(\Delta t) & \boldsymbol{\Gamma}_{22}(\Delta t) \end{array} \right] \tag{13}$$

where there are $|\mathcal{V}| - K$ and K elements in $\mathbf{r}_1(\Delta t)$ and $\mathbf{r}_2(\Delta t)$, and $\boldsymbol{\Gamma}_{11}(\Delta t)$, $\boldsymbol{\Gamma}_{12}(\Delta t)$, $\boldsymbol{\Gamma}_{21}(\Delta t)$ and $\boldsymbol{\Gamma}_{22}(\Delta t)$ are $(|\mathcal{V}| - K) * (|\mathcal{V}| - K)$, $(|\mathcal{V}| - K) * K$, $K * (|\mathcal{V}| - K)$ and $K * K$ sub-matrices respectively. Based on the definition of conditional expectation of MND, we have:

$$E[\mathbf{TV}_{MISSING}(\Delta t)\,|\mathbf{TV}_{OBS}(\Delta t)] = \mathbf{r}_1(\Delta t) + \left\{ \begin{array}{c} \boldsymbol{\Gamma}_{12}(\Delta t) * (\boldsymbol{\Gamma}_{22}(\Delta t))^{-1} * \\ (\mathbf{TV}_{OBS}(\Delta t) - \mathbf{r}_2(\Delta t)) \end{array} \right\} \tag{14}$$

3.4 TISV Algorithm

Now, we present the solution that consists of two parts.

The first part is to extract the time-varying traffic influence models between neighboring intersections and to extract the holistic time-varying traffic patterns of the entire road network. The travel patterns are obtained by mining the historical trajectory dataset of taxicabs and by mining the dataset of fixed road surveillance cameras. This part only needs to be executed once and the results can be used to infer the traffic volumes of urban camera-free intersections in real-time in the second part. Note that the first part should be executed for all time slots of a day.

Algorithm 1 The TISV Algorithm

procedure TRAFFICVOLUMEINFERRING(Δt)
  Initialization: Real-time observed traffic volume dataset Θ of camera-equipped intersection set Ψ
  $\mathbf{TV}_{OBS}(\Delta t) \leftarrow \Theta$
  $\mathbf{r}_{RE}(\Delta t) \leftarrow rearrange\ \mathbf{r}(\Delta t)\ with\ \Psi$
  $\boldsymbol{\Gamma}_{RE}(\Delta t) \leftarrow rearrange\ \boldsymbol{\Gamma}(\Delta t)\ with\ \Psi$
  $K \leftarrow |\Psi|$
  $\left(\begin{array}{c} \mathbf{r}_1(\Delta t) \\ \mathbf{r}_2(\Delta t) \end{array} \right) \leftarrow Divide\ \mathbf{r}_{RE}(\Delta t)\ with\ K$
  $\left(\begin{array}{cc} \boldsymbol{\Gamma}_{11}(\Delta t) & \boldsymbol{\Gamma}_{12}(\Delta t) \\ \boldsymbol{\Gamma}_{21}(\Delta t) & \boldsymbol{\Gamma}_{22}(\Delta t) \end{array} \right) \leftarrow Divide\ \boldsymbol{\Gamma}_{RE}(\Delta t)\ with\ K$
  $a \leftarrow (\boldsymbol{\Gamma}_{22}(\Delta t))^{-1}$
  $b \leftarrow \mathbf{TV}_{OBS}(\Delta t) - \mathbf{r}_2(\Delta t)$
  $\mathbf{TV}_{MISSING}(\Delta t) \leftarrow \mathbf{r}_1(\Delta t) + \boldsymbol{\Gamma}_{12}(\Delta t) * a * b$
  for $i = 1\ to\ |\mathcal{V}| - K$ **do**
    $Return\ \widehat{V_j(\Delta t)}$
  end for
end procedure

The second part is the Traffic Volume Inferring with Sparse Video Surveillance Cameras algorithm (**TISV** in short), which is triggered if the real-time traffic volume information is to be inferred. In particular, to infer the traffic volumes of camera-free intersections, we first obtain the traffic volume information of all camera-equipped intersections, then re-arrange and re-calculate the corresponding sub-matrices of the MND on-the-fly. Finally, the expected traffic volume values for all camera-free intersections can be calculated. The

pseudo code of the TISV algorithm is given in Algorithm 1. The parameters in Algorithm 1 are as defined previously.

4 Experiments

We present the experimental setup in Section 4.1 and show the results in Section 4.2.

4.1 Experimental Setup

Road Networks. Our experiment is carried out on the main urban area of Suzhou Industrial Park covering 45.5 square kilometers with a road network of 107 road intersections and 298 road segments. Among the 107 intersections, 44 are equipped with fixed video surveillance systems.

Taxicab Systems. Historical records of 4,303 taxicabs are collected for training transition probabilities. All taxicabs are equipped with a GPS-based navigation system and 3G/4G network and upload their ID, location, speed, direction, etc. in every two minutes.

Verification. The inference performance is measured by cross-validation. We randomly choose 35% camera-equipped intersections as verifying intersections. Assuming the surveillance results of the chosen intersections are missing, we can infer them by the information of other camera-equipped intersections and then compare with the ground truth. An example region is shown in Figure 3, where camera-equipped intersections are marked by red dots, and the verifying intersections are illustrated by blue dots. We use the dataset of taxicabs from May 1, 2015 to April 30, 2016 to train the transition probabilities, then use the surveillance data of the same period to build the holistic traffic estimation models. We then randomly select another 10 days, including both workdays and weekends, to test the accuracy of inference of the real-time traffic volumes for the verifying intersections. In all testing cases, we consider time slots from 7 am to 8 pm, for each day.

Competitors. We evaluate the performance of our TISV algorithm by comparing it with the **OKA** (One Kilometer Average), **LR** (Linear Regression), **K-means** and **LSTM** (Long Short-Term Memory) strategies. The OKA strategy assumes the real-time traffic volume of a camera-free intersection equals to the average traffic volume of all camera-equipped intersections within one kilometer. The LR strategy establishes an linear regression model for each camera-equipped intersection, and assumes its nearest camera-free intersection follows the same linear regression model. The K-means strategy partitions all intersections into clusters and assume the traffic volume of a camera-free intersection equals to the one of its cluster center. The LSTM strategy approximates the traffic flow of a camera-free intersection by its nearest camera-equipped intersection.

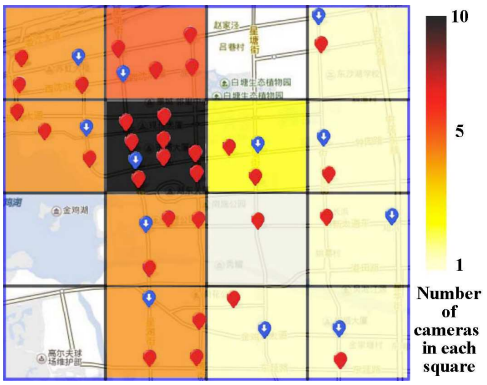

Figure 3: An example region with verifying intersections

4.2 Results

We first evaluate the effectiveness of our proposed TISV algorithm by comparing its accuracy with several competitors. Then, we made detailed testing to show how the proposal performs with varying factors. In the sequel, we use AR to denote inference accuracy for brevity.

Effectiveness

Impact of day type. We show the effectiveness of our proposal in Figure 4. It can be observed that the accuracy of TISV method is steadily above 70%. Compared with the baseline methods, i.e., OKA, LR, K-means and LSTM, our solution can increase the accuracy by 31.19%, 29.36%, 31.62% and 8.68%, respectively. The LSTM method performs better than other baselines. But the performance is unstable, i.e., for weekends. By contrast, our method is stable towards day types. The reason behind is that our method addresses the influences of traffic volumes in a global view, whereas other methods only consider the local effect of the traffic volume, e.g., on the neighborhood camera-equipped intersections.

Impact of time period. We examine the performance with respect to the effect of time slots in Figures 5 and 6, respectively. In particular, we show the results on weekdays in Figure 5, and the results on weekends in Figure 6.

From the two figures, we can observe that our solution outperforms all competitors and achieves an accuracy as high as 75%. For traffic volume estimation, different predicability is achieved due to the different travel patterns, which is consistent with the observations of Figure 4. More, we find that the inference accuracy works better in rush hours. For example, in Figure 5, the accuracy can be as high as about 80% at $17pm$. It is because that during rush hours the travel patterns are with more regularities. Thus, it enables a better estimation on traffic volume inference. Finally, we discover the performances of five solutions in morning rush hours are much better than those of afternoon. It is because of the centralized distribution of traffic flows during rush hours. Also, this is consistent with the skewed distribution of business districts and factories where a large population work in a relatively small region. However, the traffic flow during afternoon rush hours are more decentralized due to the more evenly distributed residential regions.

Analysis

We test the accuracy of our proposed TISV algorithm with varying factors for the real-time intersection traffic volume inference. In particular, we investigate: i) the impact of the number of camera-equipped intersections; ii) the impact of the traffic volume; and iii) the impact of the density of camera-equipped intersections.

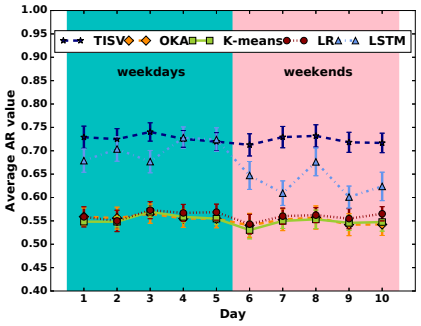

Figure 4: Impacts of weekdays and weekends

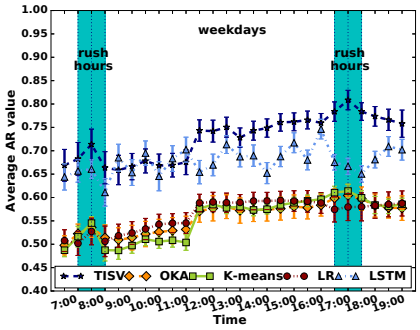

Figure 5: Impacts of time periods in weekdays

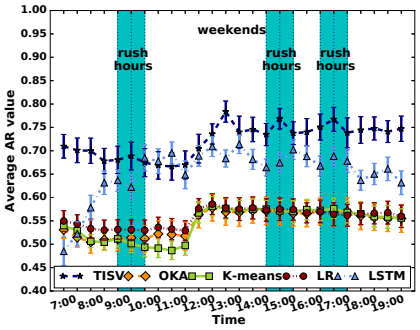

Figure 6: Impacts of time periods in weekdays

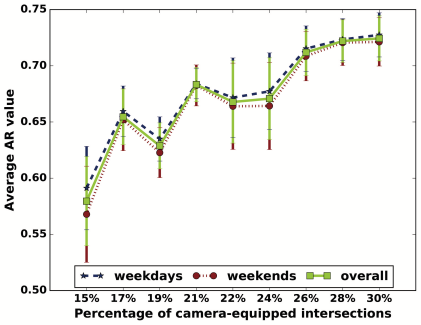

Figure 7: Impacts of percentages of camera-equipped intersections

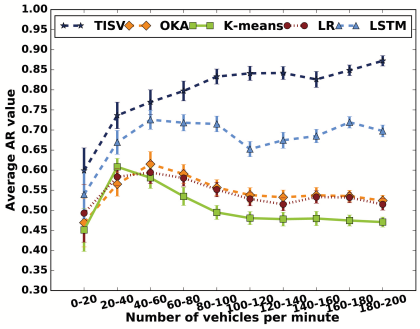

Figure 8: Impacts of traffic volumes

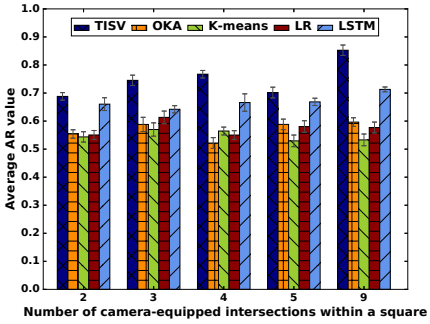

Figure 9: Impacts of densities of camera-equipped intersections

Impact of percentage of camera-equipped intersections. We investigate the effect of the percentages of camera-equipped intersections by varying the percentage of camera-equipped intersections from 15% to 30%. The results are reported in Figure 7. It can be observed that the accuracy ratio of our solution increases roughly with the increase of the percentage of camera-equipped intersections. Because the MND can be better modelled if with larger number of camera-equipped intersections. Also, it reflects that the performances of our solution for weekdays and weekends are close, which means our solution is general to the day types.

Impact of traffic volume. We study the effect of real time volume of camera-free intersections on the prediction accuracy in Figure 8. It can be observed that the performance of our solution increases in accordance with the increase of actual traffic volumes of camera-free intersections and that our solution outperforms the other four in all settings. In particular, the accuracy ratio of our algorithm has an increase of more than 25% while the actual traffic volumes of camera-free intersections increase from 0-20 vehicles per minute to 180-200 vehicles per minute. Also, we can observe that the performances of the OKA, K-means, LR and LSTM strategies are relatively better while the actual traffic volumes of camera-free intersections are between 20 and 60 vehicles per minute. For most urban intersections, the traffic volume is within such a range during the day time, and the traffic flows follow stable patterns as well. Compare with the competitors, our solution considers the mutual influence between intersections which results in a better estimation.

Impact of density of camera-equipped intersections. Finally, we evaluate the performances of our method as well as competitors while changing the density of camera-equipped intersections. To do that, we partition the road network into equal-sized squares, as exemplified in Figure 3. The density of camera-equipped intersections is classified into five levels, i.e., 2, 3, 4, 5, 9 per square, respectively. The results are reported in Figure 9. We can observe that with the increase of the number of camera-equipped intersections, the inference accuracy increases accordingly. By contrast, the competitors' performance does not change. This is because our solution approximates the traffic volume model from a global view, whereas all others only consider intersections in the neighborhood.

5 Conclusion

Recent technology advances provide us new opportunities to improve traditional real-time traffic surveillance system. In this paper, we model the time-varying traffic influence model between neighboring intersections with transition probabilities, then use MNDs to approximate the traffic volumes of the entire road network. We further propose a novel method to infer the possible real-time traffic volume of a camera-free intersection with the real-time traffic volume information of sparsely distributed camera-equipped intersections. Performance evaluations from real datasets demonstrate the effectiveness of our proposal.

Acknowledgments

We gratefully acknowledge anonymous reviewers for reading this paper and giving valuable comments. This paper is partially supported by the NSFC (No.61672487, No.61472384, No.61772492, No.61772492, No.U1301256), Jiangsu Natural Science Foundation (No.BK20151239, No.BK20171240, No.BK20171240), and Fundamental Research Funds for the Central Universities. Xike Xie is supported by the CAS Pioneer Hundred Talents Program.

References

[Bas *et al.*, 2007] Erhan Bas, A Murat Tekalp, and F Sibel Salman. Automatic vehicle counting from video for traffic flow analysis. In *2007 IEEE Intelligent Vehicles Symposium*, pages 392–397. Ieee, 2007.

[De Fabritiis *et al.*, 2008] Corrado De Fabritiis, Roberto Ragona, and Gaetano Valenti. Traffic estimation and prediction based on real time floating car data. In *2008 11th International IEEE Conference on Intelligent Transportation Systems*, pages 197–203. IEEE, 2008.

[Dia, 2001] Hussein Dia. An object-oriented neural network approach to short-term traffic forecasting. *European Journal of Operational Research*, 131(2):253–261, 2001.

[Gonzalez *et al.*, 2007] Hector Gonzalez, Jiawei Han, Xiaolei Li, Margaret Myslinska, and John Paul Sondag. Adaptive fastest path computation on a road network: a traffic mining approach. In *Proceedings of the 33rd international conference on Very large data bases*, pages 794–805. VLDB Endowment, 2007.

[Lam and Xu, 1999] William HK Lam and G Xu. A traffic flow simulator for network reliability assessment. *Journal of advanced transportation*, 33(2):159–182, 1999.

[Leduc, 2008] Guillaume Leduc. Road traffic data: Collection methods and applications. *Working Papers on Energy, Transport and Climate Change*, 1(55), 2008.

[Liu *et al.*, 2013] Honghai Liu, Shengyong Chen, and Naoyuki Kubota. Intelligent video systems and analytics: A survey. *IEEE Transactions on Industrial Informatics*, 9(3):1222–1233, 2013.

[Lu *et al.*, 2009] Rongxing Lu, Xiaodong Lin, Haojin Zhu, and Xuemin Shen. Spark: a new vanet-based smart parking scheme for large parking lots. In *INFOCOM 2009, IEEE*, pages 1413–1421. IEEE, 2009.

[Min and Wynter, 2011] Wanli Min and Laura Wynter. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, 19(4):606–616, 2011.

[Schmitt and Jula, 2006] E Schmitt and Hossein Jula. Vehicle route guidance systems: classification and comparison. In *2006 IEEE Intelligent Transportation Systems Conference*, pages 242–247. IEEE, 2006.

[Suzuki and Nakamura, 2006] Kazufumi Suzuki and Hideki Nakamura. Trafficanalyzer-the integrated video image processing system for traffic flow analysis. In *Proceedings of the 13th ITS World Congress, London, 8-12 October 2006*, 2006.

[Wang *et al.*, 2014] Yang Wang, Liusheng Huang, Tianbo Gu, Hao Wei, Kai Xing, and Junshan Zhang. Data-driven traffic flow analysis for vehicular communications. In *INFOCOM, 2014 Proceedings IEEE*, pages 1977–1985. IEEE, 2014.

[Wang *et al.*, 2016] Yang Wang, Erkun Yang, Wei Zheng, Liusheng Huang, Hengchang Liu, and Binxin Liang. A realistic and optimized v2v communication system for taxicabs. In *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, pages 139–148. IEEE, 2016.

[Wang, 2010] Fei-Yue Wang. Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications. *IEEE Transactions on Intelligent Transportation Systems*, 11(3):630–638, 2010.

[Yasdi, 1999] Ramin Yasdi. Prediction of road traffic using a neural network approach. *Neural computing & applications*, 8(2):135–142, 1999.

[Yuan *et al.*, 2012] Nicholas Jing Yuan, Yu Zheng, and Xing Xie. Segmentation of urban areas using road networks. *Technical Report*, 2012.

[Yuan *et al.*, 2013] Jing Yuan, Yu Zheng, Xing Xie, and Guangzhong Sun. T-drive: Enhancing driving directions with taxi drivers' intelligence. *Knowledge and Data Engineering, IEEE Transactions on*, 25(1):220–232, 2013.

[Zhang *et al.*, 2011] Junping Zhang, Fei-Yue Wang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1624–1639, 2011.

[Zheng *et al.*, 2010] Yu Zheng, Jing Yuan, Wenlei Xie, Xing Xie, and Guangzhong Sun. Drive smartly as a taxi driver. In *Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing (UIC/ATC), 2010 7th International Conference on*, pages 484–486. IEEE, 2010.