Probabilistic Modeling of Dynamic Traffic Flow across Non-overlapping Camera Views

[1]Ching-Chun Huang, [2]Wei-Chen Chiu, [1]Sheng-Jyh Wang, and [2]Jen-Hui Chuang
Dept. of [1]Electronics Engineering/ [2]Computer Science, National Chiao Tung University, Taiwan
E-mail: chingchun.huang3@gmail.com; walon.cs97g@nctu.edu.tw

Abstract

In this paper, we propose a probabilistic method to model the dynamic traffic flow across non-overlapping camera views. By assuming the transition time of object movement follows a certain global model, we may infer the time-varying traffic status in the unseen region without performing explicit object correspondence between camera views. In this paper, we model object correspondence and parameter estimation as a unified problem under the proposed EM (Expectation-Maximization) based framework. By treating object correspondence as a latent random variable, the proposed framework can iteratively search for the optimal model parameters with the implicit consideration of object correspondence.

1. Introduction

One core technique in modern intelligent transport systems is to automatically monitor the traffic flow that reflects the characteristics of traffic status. In general, existing methods for traffic flow monitoring can be roughly categorized into non-vision-based and vision-based approaches. For non-vision-based methods, most methods use an automatic vehicle localization system (AVLS) to collect the temporal and spatial data of the monitored vehicles. Unfortunately, an AVLS system has the drawback of being very expensive in installation and maintenance. Moreover, the privacy issue also hinders the popularization of AVLS–based applications [1]. On the contrary, with the increasing usage of surveillance cameras, vision-based traffic flow monitoring has become more popular. By combining vision-based traffic flow estimation with the Google Map or the CCTV Live-Cam service, cloud-computing based ITS services are becoming feasible.

Many computer vision techniques have been widely used in traffic applications [2]. However, almost all the existing vision-based traffic surveillance systems suffer from the limited field of views (FOV's). These systems can only infer the observed traffic status at the camera location. For a wide-area traffic flow surveillance system, we may have to estimate the traffic state in the unseen zones out of the FOV's in order to infer a "semi-global" traffic flow.

In this paper, we propose a method to infer the traffic flow in the unseen zones by modeling the probabilistic distribution of transition time traveling between non-overlapped FOV's. In principle, a longer transition time implies a heavier traffic status, and vice versa. To dynamically determine the transition-time distribution, performing object correspondence between non-overlapped FOV's is a possible solution. Unfortunately, efficient and accurate correspondence is still an open challenge in computer vision. In [3,4], Javed et al. proposed a different approach, in which they learned the inter–camera color difference and the transition properties between cameras. Based on the object appearance and space–time cues, they matched the tracks of targets between non-overlapping camera views in a probabilistic sense. In [5], Song et al. matched the appearance and biometric features for tracking targets with a camera network. By using a feature graph that contains the feature vector and the similarity score observed over space and time, trajectories of moving people can be estimated by finding the optimal path in this graph. However, since vehicles usually share similar appearance, shape, and motion, this approach doesn't work as well in tracking moving vehicles.

On the other hand, a few research works try to model the motion in the unseen zones. Sheikh et al. [6] presented a framework for the association of objects across camera views in a planar scene. They used the homography transformation to relate the geometry between cameras and assumed the object's trajectory follows a polynomial kinetic model. By simultaneously computing the maximum likelihood estimates of the homography and the trajectory parameters, they establish the object associations. However, the polynomial kinetic model is mainly for continuous motion and is not suitable for describing abrupt motion change, like vehicle stop caused by traffic signals. In [7], Makris et al. assumed a single mode transition time distribution between FOV's and they exhaustively searched for the mode by computing cross–correlation between the arrival time and departure time of objects across camera views. However, this single-mode constraint is somewhat too tight for the modeling of traffic flow. In stead, Tieu et al. [8] proposed an entropy-based method to estimate the transition-time distribution and object association. They claimed the optimal object association corresponds to the transition-time distribution with the minimal entropy. Though this approach is simple and flexible, the minimum-entropy assumption may not hold for practical applications.

In this paper, we propose the use of a Gaussian mixture model (GMM) for the description of the transition-time distribution and develop an Expectation–Maximization (EM) based framework to determine the distribution of transition time with the implicit consideration of object correspondence. Our experiment showed that the GMM model is indeed more appropriate than the single-mode model and the minimum-entropy model in the modeling of the traffic flow across non-overlapping camera views.

2. Target selection and traffic flow model

Our goal is to monitor the traffic status between two non–overlapping FOV's linked by a road. In general, since there could be many intersections along the road, vehicles may leave or enter the road in–between. To achieve robust analysis, we chose the monitored targets to be the buses that follow a reliable route across the camera views. In our system, we recorded the entry time and exit time of all the buses in the non–overlapping camera views. Without knowing the exact correspondence between the entry time and exit time, the recorded times are the only observations we used to infer the traffic status. A segment of the target selection and entry/exit time records are shown in Figure 1.

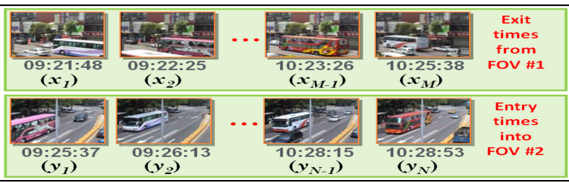

Fig. 1. A result of target selection. We choose buses as the targets and record the exit time ($\{x_m\}_{m=1\sim M}$) and entry time. ($\{y_n\}_{n=1\sim N}$).

Although every vehicle has its own behavior pattern, the traffic flow tends to follow a global trend due to the interactions among vehicles and the environments, like car following, car queuing, bus stops, and traffic signals. Hence, if we treat the transition time as a random variable, we may assume the transition-time distribution follows certain kind of global model that reflects the statistical behavior of the traffic flow. In our system, we propose the use of the Gaussian Mixture model (GMM) to approximate the transition-time distribution. Here, based on the central limit theorem, we assume the overall uncertainty caused by various kinds of factors over a continuous traffic flow can be well modeled by a Gaussian distribution, while the factors that contribute to the interrupt of traffic flow, like the bus stops and the traffic signals, cause the multiple modes of the distribution. With the GMM model, our system objective is to derive the parameters of this GMM model in order to describe the status of traffic flow. Moreover, since the traffic flow status changes dynamically over time, we split the observation time into overlapped time windows and perform the traffic flow analysis over each window individually, as illustrated in Figure 2.

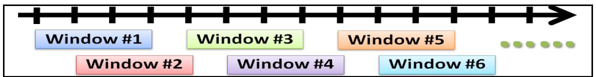

Fig. 2. Illustration of overlapped observation periods.

3. Problem formulation and EM solution

Given two non-overlapping camera views, to determine the model parameters and to further infer the traffic status in-between, we formulate the problem as an optimization process expressed as

$$\theta^* = \arg\max_{\theta} p(\theta \mid Z), \qquad (1)$$

where θ denotes the model parameters and Z= ($\{x_m\}_M$, $\{y_n\}_N$) denotes the observations that includes the exit time set ($x_1, x_2,..., x_M$) extracted from one FOV and the entry time set ($y_1, y_2, . . . , y_N$) extracted from the other FOV. However, it is difficult to directly model the probability model P(θ|Z) in (1) owing to the lack of physical connection between θ and Z. To compensate the gap between the model parameters θ and the observations Z, we introduce the correspondence C that represents the mappings from $\{x_m\}_M$ to $\{y_n\}_N$. Here, we denote C = $\{c_m(x_m)\}_M$, where $c_m(x_m)$ indicates the unknown mapping from x_m to some element in $\{y_n\}_N$. That is, if a bus leaves one FOV at x_m, travels through the road for some period, and then enters the other FOV at y_k, we express this mapping as $c_m(x_m)=y_k$ and define the transition time to be $t_m = c_m(x_m) - x_m$. Once if all the mappings are determined, by collecting all the transition times to form a data set T = $\{t_m\}$, the model parameters θ can be estimated.

Since we do not have the correspondence C yet, we treat C as latent information and reformulate (1) as (2) by marginalizing out all possible correspondence C.

$$\theta^* = \arg\max_{\theta} \sum_{C} p(\theta, C \mid Z). \quad (2)$$

Although a direct approach to computing this optimization problem is generally intractable, the EM algorithm provides a mean to deduce θ* by iteratively performing the following two steps until the convergence condition is met:

1. E Step: Calculate the expected log likelihood function Q($\theta \mid \theta^{(old)}$):

$$Q(\theta \mid \theta^{(old)}) = \sum_{C} log(p(Z, C \mid \theta)) p(C \mid Z, \theta^{(old)}). \quad (3)$$

   where the expectation is taken with respect to the posterior distribution p(C|Z, $\theta^{(old)}$) over all possible correspondence C, given the data Z and the estimated $\theta^{(old)}$ at the previous iteration.

2. M step: Find the Maximum Likelihood estimate $\theta^{(new)}$ by maximizing Q($\theta \mid \theta^{(old)}$):

$$\theta^{(new)} = \arg\max_{\theta} Q(\theta \mid \theta^{(old)}). \quad (4)$$

However, in the E step, since the number of possible correspondences explodes enormously, it is difficult to compute the expectation directly. Instead, we apply the Markov chain Monte Carlo (MCMC) sampler to draw R correspondence samples $\{C^r\}_R$ from p(C|Z, $\theta^{(old)}$) and approximate the expectation in the E step based on these correspondence samples. Particularly, we use the Gibbs sampling to generate R samples, as illustrated in *Algorithm 1*. While sampling, we prefer the entry/exit times to be one-to-one correspondence in order to match the physical constraint. However, in practice, we allow an entry time to be matched with more than one exit time to speed up the Gibbs sampling process. In detail, in the (r+1)th iteration, we draw the mth mapping $c_m^{(r+1)}$ based on the conditional probability

$$p(c_m \mid c_1^{(r+1)}, \ldots, c_{m-1}^{(r+1)}, c_{m+1}^{(r)}, \ldots, c_M^{(r)}, Z, \theta^{(old)})$$

$$\sim p_{GMM}(t_m \mid \theta^{(old)}) \exp(-\alpha \sum_{j=1, j\neq m}^{M} \delta[c_m, c_j]). \quad (5)$$

In (5), p_{GMM} denotes the GMM distribution model, the exponential term represents the penalty over many-to-one mappings, α is an adjustable penalty constant, and δ[p,q] is defined as

$$\delta[p,q] = \begin{cases} 1 & \text{if } p = q \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$

In our system, we consider the R correspondence samples $\{C^r\}_R$, and assume the set of observation pairs $\{(x_m, c_m^{(r+1)}(x_m) = y_n\}_M^R$ are independent of each other. Hence, the likelihood Q($\theta \mid \theta^{(old)}$) in (3) can be approximated as

$$Q(\theta \mid \theta^{(old)}) = \sum_{r=1}^{R} log(p(Z, C^r \mid \theta)) = \sum_{r=1}^{R}\sum_{m=1}^{M} log(p_{GMM}(t_m^r \mid \theta)), \quad (7)$$

where $t_m^r = c_m^{(r)}(x_m) - x_m$. Next, in the M step, the optimization problem is interpreted as finding the optimal GMM parameters $\theta^{(new)}$ by regarding all the transition time samples $\{t_m^r\}_M^R$ as the observation. Here, we use a mixture of K Gaussians to approximate the global transition-time distribution. This distribution is parameterized by the weights w_j, the mean μ_j, and the variance σ_j^2 for the jth Gaussian, with j = 1, . . . ,K. To estimate these parameters given the set of transition times $\{t_m^r\}_M^R$, we adopt the typical EM algorithm for GMM which can be formulated as

$$w_j^{(i+1)} = \frac{1}{RM} \sum_{r=1}^{R}\sum_{m=1}^{M} P(j \mid t_m^r) \quad (8)$$

$$\mu_j^{(i+1)} = (\sum_{r=1}^{R}\sum_{m=1}^{M} P(j \mid t_m^r) t_m^r) \Big/ (\sum_{r=1}^{R}\sum_{m=1}^{M} P(j \mid t_m^r)) \quad (9)$$

$$(\sigma_j^2)^{(i+1)} = (\sum_{r=1}^{R}\sum_{m=1}^{M} P(j \mid t_m^r)(t_m^r - \mu_j^{(i+1)})^2) \Big/ (\sum_{r=1}^{R}\sum_{m=1}^{M} P(j \mid t_m^r)) \quad (10)$$

where

$$P(j \mid t_m^r) = (w_k^{(i)} P(t_m^r \mid j; \mu_j^{(i)}, \sigma_j^{(i)})) \Big/ (\sum_{k=1}^{K} w_k^{(i)} P(t_m^r \mid k; \mu_k^{(i)}, \sigma_k^{(i)})). \quad (11)$$

After a few iterations over (3) and (4), the optimal θ* can be obtained. In Figure 3, we illustrate the process of the proposed algorithm within a time window.

Algorithm 1: Gibbs sampler (GS)

Initialize $C^0 = c_{(1:M)}^{(0)}$ = Null correspondence

For r = 0 to R − 1 do

Sample $c_1^{(r+1)} \sim p(c_1 \mid c_2^{(r)}, c_3^{(r)}, \ldots, c_M^{(r)}, Z, \theta)$.

Sample $c_2^{(r+1)} \sim p(c_2 \mid c_1^{(r+1)}, c_3^{(r)}, \ldots, c_M^{(r)}, Z, \theta)$.

: : :

Sample $c_m^{(r+1)} \sim p(c_m \mid c_1^{(r+1)}, \ldots, c_{m-1}^{(r+1)}, c_{m+1}^{(r)}, \ldots, c_M^{(r)}, Z, \theta)$.

: : :

Sample $c_M^{(r+1)} \sim p(c_M \mid c_1^{(r+1)}, \ldots, c_{M-1}^{(r+1)}, Z, \theta)$.

Output R correspondence samples $\{C^r\}_R = \{C^1, \ldots, C^R\}$.

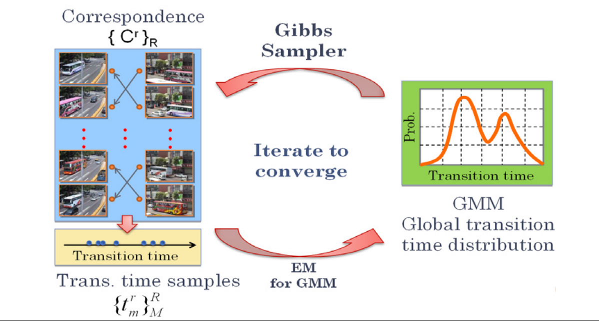

Fig. 3. The Proposed EM algorithm for a time–window.

4. Results and conclusion

Our experimental environment is shown in Figure 4. Two cameras were mounted at two overpasses that are a few blocks apart. FOV's of these two cameras are non-overlapping and the scenes are linked by Kuang Fu Road (the red line), a main street in Hsinchu, Taiwan. Besides, there are three intersections with traffic lights and a few bus stops in-between. The traffic videos were taken from 9:45 in the morning till 18:50 in the evening. We divide the whole time period into 37 one-hour time windows, with 45 minutes overlapping with the previous window.

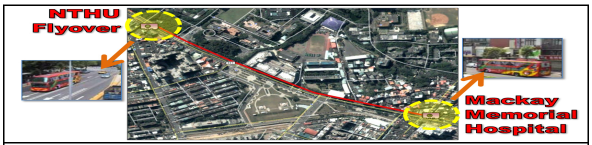

Fig. 4. The experimental environment. All results are at http://140.113.238.220/~chingchun /projects.html

We manually checked the test videos, selected the frames with the presence of buses, and built the correspondence of buses between two camera views as our ground truth. Besides, the entry/exit times of buses were recorded as our input data. In the near feature, we plan to add automatic vision-based bus detection into our framework. In our experiments, we applied the proposed EM algorithm over each time window. For the first time window, we initialise the GMM model to contain K broadly separated, widely spread, and equally weighted Gaussian functions. For instance, for the case of K = 3, we set the parameters to be μ = {50, 150, 250}, σ^2 = {200, 200, 200}, and w = {0.33, 0.33, 0.33}. For the initial setting of the subsequent time windows, we simply propagate the correspondence result of the overlapped period from the previous window to the current window.

To evaluate our method and compare with the methods proposed by Makris [7] and Tieu [8], we calculated the earth mover's distance (EMD) between the estimated transition-time distribution and the ground truth. For the ground truth, Tieu's method, and our method, the object correspondences are first transferred into transition-time samples. The samples are converted into the transition-time distribution via kernel density estimation. In Figure 5, we compare the distributions based on different methods. Besides, the entropies of the estimated transition-time distributions are calculated. The experimental results showed that the optimal correspondence does not necessarily have the minimum entropy. In Table 1, the details of the calculated earth mover distance and entropy are listed. Our method produces the minimum EMD and well matches the ground truth.

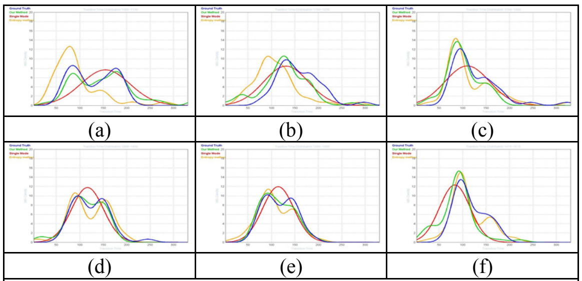

Fig. 5. Comparisons of transition time distributions in 6 different time-windows (x-axis: time, y-axis: Probability). Ground Truth (blue); Our method (green); Makris [7] (red) ; Tieu [8] (yellow).

Table. 1. Compare the earth mover distances (EMD) and entropies (Etrpy) of distributions (a)~(g) in Fig. 5.

	EMD				Etrpy			
	Truth	[7]	[8]	Our	Truth	[7]	[8]	Our
(a)	0	50.33	83.10	29.35	5.400	5.117	5.121	5.167
(b)	0	10.70	60.82	8.87	5.238	5.271	4.978	5.129
(c)	0	14.09	34.95	14.72	5.292	5.224	5.114	5.354
(d)	0	18.91	13.03	12.85	5.362	5.321	5.235	5.392
(e)	0	16.09	21.14	7.36	5.366	5.309	5.098	5.552
(f)	0	29.70	68.98	26.37	5.344	5.286	4.895	5.043

In conclusion, with the implicit consideration of object correspondence and the GMM modeling, we proposed an EM-based framework to probabilistically model the dynamic traffic flow between camera views. Our system provides a new thinking to well utilize the existing surveillance cameras for traffic monitoring. Experiments have shown that our approach performs well in a complicated traffic environment in real life.

References

[1] B. Fishbain, I. Ideses, D. Mahalel, and L.Yaroslavsky, "Real-time vision-based traffic flow measurements and incident detection," *Proceedings of SPIE*, 2009.

[2] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *Image and Vision Computing*, 2003.

[3] O. Javed, Z. Rasheed, K. Shafique, and M. Shah, "Tracking across multiple cameras with disjoint views," *IEEE Conference ICCV*, 2003.

[4] O. Javed, K. Shafique, and M. Shah, "Modeling inter-camera space–time and appearance relationships for tracking across non–overlapping views," *Computer Vision and Image Understanding*, 2008.

[5] B. Song and R. Chdhury, "Stochastic adaptive tracking in a camera network," *IEEE Conference ICCV*, 2007.

[6] Y. Sheikh, X. Li, and M. Shah, "Trajectory Association across Non–overlapping Moving Cameras in Planar Scenes," *IEEE Conference CVPR*, 2007.

[7] D. Makris, T. Ellis, and J. Black, "Bridging the gaps between cameras," *IEEE Conference CVPR,* 2004.

[8] K. Tieu, G. Dalley, and Grimson, "Inference of non–overlapping camera network topology by measuring statistical dependence," *IEEE Conference ICCV*, 2005.