Unsupervised Tracking With the Doubly Stochastic Dirichlet Process Mixture Model

Xing Sun, *Student Member, IEEE*, Nelson H. C. Yung, *Senior Member, IEEE*, and Edmund Y. Lam, *Fellow, IEEE*

***Abstract*—We present an unsupervised tracking algorithm for human and car trajectory detection, using what is called the temporal doubly stochastic Dirichlet process (TDSDP) mixture model. The TDSDP captures the global dependence and the variation of human crowds and cars in temporal domains without the Markov assumption, making it particularly suitable for long-term tracking. Moreover, TDSDP prior can estimate the number of trajectories automatically. We first define the TDSDP based on the Cox process and then explain how to construct a TDSDP mixture model from thinning multiple Dirichlet process mixtures (DPMs) with conjugate priors. Next, a Markov chain Monte Carlo sampling inference is presented. Experimental results on synthetic and real-world data demonstrate that the proposed TDSDP mixture is superior to the DPM and the dependent Dirichlet process (DDP) in terms of topic variation modeling. PETS2001 data set experiments show that TDSDP has more robust object tracking capability over DDP based on generalized Polya urn. Low-quality fish data set experiments indicate that the TDSDP excels at solving tracking problems with insufficient features.**

***Index Terms*—Nonparametric Bayesian, doubly stochastic, Gaussian process, unsupervised tracking.**

I. Introduction

Vehicle tracking data analysis with intelligent artificial tools plays a significant role in transportation systems. For example, Mao *et al.* [1] showed that tracking and prediction of vehicle positions are important in mining operations and other related applications in the intelligent transportation system area. Vehicle tracking has some challenging problems, such as resolving the vehicle occlusion [2] and finding the number of vehicles with occlusion [3]. Vasquez *et al.* [4] used Hidden Markov Model (HMM) to learn the incremental motion patterns for modeling the trajectory. Moreover, Morris *et al.* [5] employed the HMM to encode the motion paths, and it aimed to detect and predict abnormal trajectories. Nowadays, simple statistical methods seem to shoulder an almost unbearable burden to distinguish cars and human crowd trajectories in the more and more complex transportation systems. Estimation of the number of trajectories is another difficult task. In this paper, an unsupervised tracking algorithm is presented. It aims to detect trajectories in a fully unsupervised fashion and estimate the trajectory number in a scene with occlusions of humans and vehicles.

Nonparametric Bayesian models based on Dirichlet Process (DP) have been proposed to model topics over space and time [6]–[10]. The term "topic" is widely used in machine learning area, which represents the characteristics of each cluster. For instance, the topic of a vehicle in a single frame can be regarded as the position center of that vehicle or other feature parameters (e.g., average color). These nonparametric Bayesian models relax the assumption that the topic number is fixed or known. They assume that the topics pool behaves as a Dirichlet Process mixture model (DPMM) [11], [12] or their extensions [6]–[10]. In particular, Dependent Dirichlet Process (DDP) [13] is a pioneering method, which realizes random dependent measures and takes advantage of modeling the dependency of DP priors. Gelfand *et al.* [10] proposed a spatial DP, where the location and weighting of the topics are sampled from Gaussian Processes (GPs), hence creating topic dependency. However, these methods encounter a computational problem, that a huge inference task is generated by sampling massive auxiliary variables, making it difficult to employ these models to track the vehicles and the humans.

Lin *et al.* [6] proposed a novel DP prior model, which aims to model topic trajectories along the temporal domain based on the Poisson process. This can be regarded as the HMM incorporated with a DP prior, which can be used for enhancing the tracking algorithms based on HMM [4], [5]. It reduces inference complexity with the Markov chain but fails to achieve flexible object transition. This model will break the trajectories and lose its targets when it encounters occlusion.

Many well-known hierarchical models with DP priors also cannot solve this tracking problem. This is because the dependency is modeled based on a shared topic pool, and topics in the same cluster are identical along the temporal and spatial domains (multiple DP priors). Attempts include the hierarchical Dirichlet process [14], and the Dependent Normalized Random Measures [15]. An application example is the vehicle location tracking in videos, where we need to track the location of objects (topics) as well as their trajectories across frames (time). Some other approaches model dependency among nearby measures without the Markov assumption. For instance, Wang *et al.* [7] proposed a non-Markov topic model, which assumed that the duration of a topic follows the beta distribution. Dubey *et al.* [8] proposed a nonparametric extension, which relaxes the assumption that the number of topics is fixed and allows topics to share the duration models. These models use various distributions to model the topics' durations (temporal variations) and assume that the topics' locations (spatial variations) are fixed.

Many approaches have been proposed to analyze topic trajectories via modeling dependency in topic space. While DDP models based on GP [10], [13] are too complex to be implemented, some other models [6], [9] have been proposed to bypass the GP priors. However, these models make a stricter assumption to achieve the dependency modeling under Markov assumption [6], [16], or do not modeling the location variation and model dependency via overlapping regions [9]. The approaches [7], [8], which model topics dependency under the non-Markov framework, are unable to obtain the position variation of the vehicle or the human crowd.

A Temporal Doubly Stochastic Dirichlet Process (TDSDP) mixture model is proposed as a generalization of the DP priors, where the spatial-temporal dependent intensity is a Sigmoid Gaussian Process (SGP). This is used to model the varying intensity of the trajectories along frame stamps. In the TDSDP, the topic variation is modeled based on this varying intensity measurement, instead of the discrete Markov assumption. Subsequently, the mixture model of the TDSDP is constructed elegantly with a thinning procedure applied to multiple DPMMs. It allows us to handle massive transportation video data with

Manuscript received July 21, 2015; revised November 5, 2015; accepted January 11, 2016. Date of publication February 26, 2016; date of current version August 25, 2016. The Associate Editor for this paper was S. Sun.

X. Sun and E. Y. Lam are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pok Fu Lam, Hong Kong (e-mail: elam@eee.hku.hk).

N. H. C. Yung is with Yung and Partners Limited, Hong Kong.

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

Digital Object Identifier 10.1109/TITS.2016.2518212

a lower computational cost. For most cases, the topic number K is much smaller than the number of observations n, i.e., $K \ll n$. Classic DDP models need to sample a vast number of auxiliary variables in the order of $O(n)$. Our method only needs to sample latent variables in the order of $O(K)$ with the thinning procedure. This construction significantly reduces the computational cost. Finally, the TDSDP mixture model enables us to track varying human and vehicle positions along different video frames with one global SGP intensity prior. Moreover, it allows us to capture location variations, birth and disappearance of objects based on dependency modeling in all topic domains. It uses the continuous probability measure sampled from the global SGP prior to fit the distribution of the variation of objects and the global SGP prior enables us to achieve a robust result.

Moreover, many other GP models [17]–[19] can be used to support the extension of TDSDP for tracking and prediction. Take the object detection in videos as an example, the SGP prior of the TDSDP can be used to cluster object trajectories and predict object location in missing or future frames.

II. The TDSDP Mixture Model

A construction method of the TDSDP mixture model is proposed. Different from the single DSDP applied on spatial data [20], this TDSDP mixture model focuses on temporal modeling with multiple DP priors. That means different DP priors share a global SGP prior. A thinning procedure in [21] is used to generate a TDSDP mixture model as follows.

First, we sample GP functions Y_t from SGP prior GP$(\cdot)$ with mean function m_t and covariance function κ_t [17] at patch t, hence

$$Y_t \sim \text{GP}(m_t, \kappa_t). \tag{1}$$

Second, conditional TDSDP $D_t|Y_t$ at patch t is sampled from the DP prior DP$(\cdot)$ with thinned prior $\sigma(Y_t)\lambda_t^*$

$$D_t|Y_t \sim \text{DP}\left(\sigma(Y_t)\lambda_t^*\right) \tag{2}$$

where the thinned function $\sigma(Y_t)$ is the sigmoid GP function $\sigma(Y_t) = (1+e^{-Y_t})^{-1}$, and $\lambda_t^* = \alpha^* H_{t,0}$ is the conjugate prior [11] for the DPMM with the hyper-parameter α^* and initial base distributions $H_{t,0}$. Since the SGP function $\sigma(Y_t) \in (0,1)$, the thinned prior $\sigma(Y_t)\lambda_t^*$ can be regarded as thinning from the conjugate prior λ_t^*. Based on this thinning procedure, the base distributions H_t of patches are dependent given the global SGP functions $\sigma(Y_t)$, which follow $H_t \propto H_{t,0}\sigma(Y_t)$.

Third, the topic parameter $\boldsymbol{\theta_{t,i}}$ at patch t is independently sampled from conditional TDSDP $D_t|Y_t$

$$\boldsymbol{\theta}_{t,i}|D_t, Y_t \overset{iid}{\sim} D_t|Y_t \tag{3}$$

where $\theta \in \Theta \times T$, Θ is the feature space and T is the temporal space. Lastly, observation data $\boldsymbol{x_{t,i}}$ is sampled from the observation distribution $f(\boldsymbol{\theta}_{t,i})$ parameterized by $\boldsymbol{\theta_{t,i}}$

$$\boldsymbol{x}_{t,i}|\boldsymbol{\theta}_{t,i} \overset{iid}{\sim} f(\boldsymbol{\theta}_{t,i}). \tag{4}$$

III. Tracking Algorithm

The proposed tracking algorithm based on the TDSDP mixture model can be split into six steps, which are shown in Fig. 1 as different line types. An example in the PETS2001 dataset [22] has been used to illustrate the mechanism of this algorithm, which includes trajectory overlapping of cars and humans and fluctuating number of topics (cars and humans), in order to illustrate merits of the TDSDP mixture model. Estimating varying topics along frames, especially the overlapping

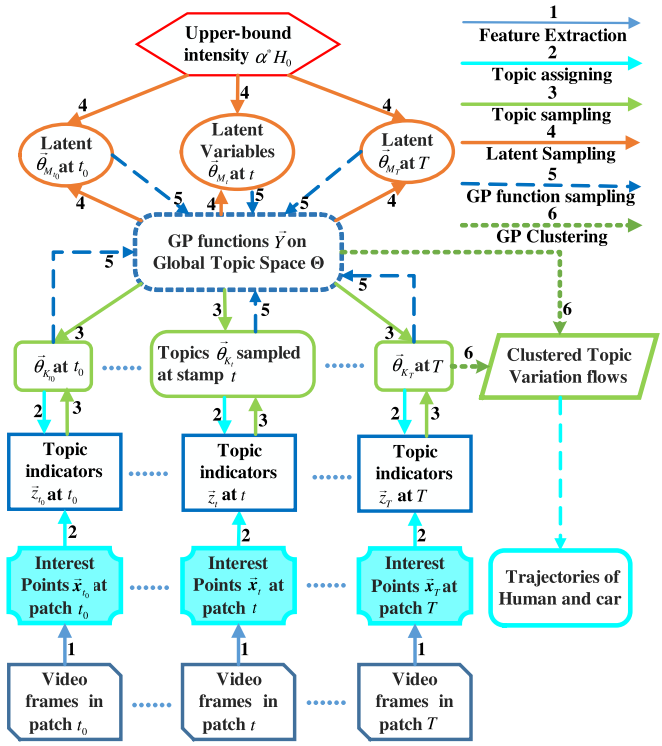

Fig. 1. Flow diagram of the TDSDP.

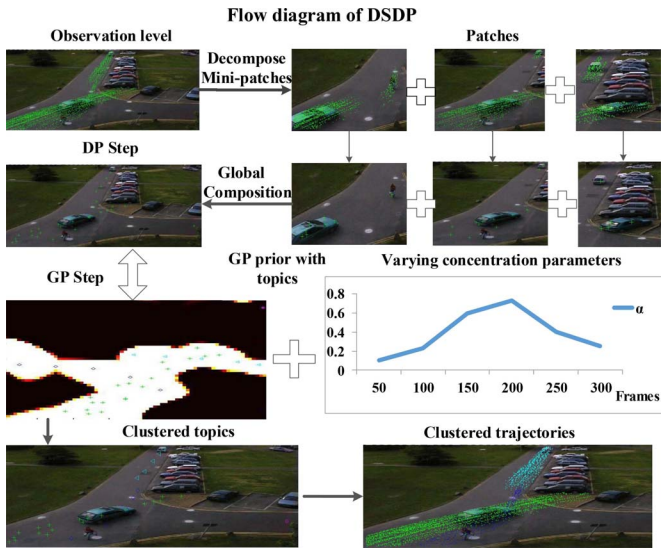

Fig. 2. Flow diagram of the TDSDP on the tracking example.

topic trajectories, demonstrates that the TDSDP enables us to capture topic variations with a global SGP prior.

A. Fast Feature Extraction Step (Blue Solid Lines)

This feature extraction [23] step captures all observations $\{\boldsymbol{x}_t\}_{t=1,\ldots,T}$ from T patches in the video, which consists of patch decomposition and interest point extraction. Extracted features are the points in the top panels of Fig. 2.

1a) Decompose the mini-patch at the observation level. Observation points are presented in the top panels as observed mixture trajectories in Fig. 2. Initial DP priors are applied to the data in

mini-patches independently, and patches are obtained by merging frames. We cut 350 frames into 20 patches. This partition can also be achieved by applying DPMM on a time stamp of data. To maintain the global dependency, none of these patches is overlapped. Since the algorithm is robust with continuous intensity, the mini-patch decomposition is more important for the computational cost in Section IV-D.

1b) Interest points are extracted from the clip using the Harris detector [23], and we apply the positions of the point as the feature based on the following two reasons. First, it can be calculated very quickly. which is important for an unsupervised tracking task with a massive unlabeled dataset. Furthermore, it illustrates the robust characteristics of our proposed tracking algorithm. Long-term tracking with inefficient feature is a difficult problem that occurs in the real world.

B. DP Priors Sampling Steps (Cyan and Green Solid Lines)

The inference algorithm is used for sampling multiple DP priors when a global SGP prior is given. This processing step is developed and extended from the sampling for the DSDP on spatial modeling, more information of which can be found from [20]. It includes two main steps of the tracking algorithm: topics assignment sampling and topic parameter sampling, which are represented by cyan and green solid lines in Fig. 1.

2) Topics assignment sampling step is used for sampling the topic indicator $z_{t,i}$ for the data $\boldsymbol{x}_{t,i}$.
3) The topic $\boldsymbol{\theta}_{t,k}$ is sampled for the cluster k at patch t when the SGP functions $\sigma(\boldsymbol{Y})$ and observations $\boldsymbol{x}_{t,i}$ assigned to the kth cluster are given.

A real-world example is shown in the second row of Fig. 2. The DP posterior is sampled in each patch with underlying SGP prior (or initial intensity). The varying mean measure H_t and concentration parameter α_t are sampled at patches.

First, we sample the DP topic indicators (cyan lines). When the SGP functions are given, we can sample the topic indicators of conditional TDSDP based on a classical DPMM sampling procedure. The probability of sampling a new topic indicator $z_{t,i}$ is proportional to the DP assignment probability (locally) at patch t and the SGP prior probability (globally):

$$\begin{aligned} &P(z_{t,i}=k|\boldsymbol{z}_{t,-i},\boldsymbol{Y}_t,\boldsymbol{\theta}) \\ &\quad\propto p(z_{t,i}=k|\boldsymbol{z}_{t,-i},\boldsymbol{x}_{t,-i})p(Y_{t,k}|z_{t,i}=k,\boldsymbol{Y}_{t,-k},\boldsymbol{\theta}) \\ &\quad\propto \begin{cases} n_{t,-i,k}\cdot\sigma\left(Y(\boldsymbol{\theta}_{t,k})\right)\cdot\ell(\boldsymbol{x}_{t,i}|\boldsymbol{\theta}_{t,k}), & k\leqslant K_t \\ \frac{\alpha^*}{M_t}\cdot\sigma\left(Y(\boldsymbol{\theta}^*)\right)\cdot\ell(\boldsymbol{x}_{t,i}|\boldsymbol{\theta}^*), & k>K_t \end{cases} \end{aligned} \tag{5}$$

where $n_{t,-i,k}$ is the number of data clustered into kth topic except $\boldsymbol{x}_{t,i}$ at patch t. The entire observations at patch t except $\boldsymbol{x}_{t,i}$ are denoted as $\boldsymbol{x}_{t,-i}$, $\boldsymbol{z}_{t,-i}$ and $\boldsymbol{Y}_{t,-k}$ have similar definitions. The number of topic variables, latent variables and data at patch t are denoted as K_t, M_t and n_t, respectively. DP assignment probability includes likelihood $\ell(\boldsymbol{x}_{t,i}|\boldsymbol{\theta}_{t,k})$ and mass function $n_{t,-i,k}$, which are only dependent on local information at patch t. Existing SGP function $\sigma(Y(\boldsymbol{\theta}_{t,k}))$ and new SGP function $\sigma(Y(\boldsymbol{\theta}^*))$ are sampled from the SGP prior.

Second, topic parameter $\boldsymbol{\theta}_{t,k}$ is updated during this topic sampling step. If $k\leqslant K_t$, then $G_k(\boldsymbol{\theta})$ is the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x}_{t,-i,k})$ given data partition $\boldsymbol{x}_{t,-i,k}$, where $\boldsymbol{x}_{t,-i,k}$ is the set of data assigned to kth cluster except $\boldsymbol{x}_i$ at patch t. Otherwise, $G_k(\boldsymbol{\theta})$ is the conjugate prior $H_0(\boldsymbol{\theta})$. Then a new topic $\boldsymbol{\theta}^*$ is sampled from $H_0(\boldsymbol{\theta})$, and its GP function $Y(\boldsymbol{\theta}^*)$ is sampled from the SGP prior. The Metropolis-Hastings acceptance ratio $a_{t,k}$ at patch t follows

$$a_{t,k}=\min\left(\frac{G_k(\boldsymbol{\theta}^*)\left(1+\exp\left(\mathrm{sgn}(k')Y(\boldsymbol{\theta}_{t,k})\right)\right)}{G_k(\boldsymbol{\theta}_{t,k})\left(1+\exp\left(\mathrm{sgn}(k')Y(\boldsymbol{\theta}^*)\right)\right)},1\right) \tag{6}$$

where $\mathrm{sgn}(\cdot)$ is a sign function and we set $k'=k-K_t+0.5$ to avoid uncertainty at $k'=0$. Furthermore, a random variable $r_{t,k}$ is uniformly sampled on [0, 1], if $a_{t,k}>r_{t,k}$, this proposal is accepted. Then $\boldsymbol{\theta}_{t,k}$ is set to $\boldsymbol{\theta}^*$ and $Y(\boldsymbol{\theta}_{t,k})$ is set to $Y(\boldsymbol{\theta}^*)$ in Eq. (5). Otherwise existing $\boldsymbol{\theta}_{t,k}$ and $Y_{t,k}$ are kept.

C. SGP Sampling Steps (Brown Solid and Blue Dotted Lines)

This step is used to infer the SGP prior when the DP priors are given. Sampling of latent variables and SGP functions are another two major steps for the tracking, which are represented as brown solid and blue dotted lines in Fig. 1, respectively.

4) Latent variables are initially sampled from upper bound intensity, which is fixed along iterations, and they are updated based on sampled SGP. Move, insert and reject operations for latent variables are similar to [24].
5) The SGP function inference step samples SGP functions with globally inferred topics and latent variables. The left panel in the third row of Fig. 2 shows the inferred SGP prior with global topics. The right panel shows the varying concentration parameter α_t along the patches.

Similar to the inference problem proven in [25], the inference of TDSDP also encounters a doubly-intractable problem. We cannot tractably and directly infer the SGP intensity. Inspired by the inference method for the SGP Cox process [24], we introduce latent variables sampled from the same measure space of topic parameters. SGP functions sampling of latent variable ensures all SGP functions are independent and inferences of the TDSDP are tractable. The likelihood of the TDSDP mixture model with multi-patches follows:

$$\begin{aligned} &p\left(\vec{\boldsymbol{\theta}}_K,M,\vec{\boldsymbol{\theta}}_M,\boldsymbol{Y}_{M+K}|\alpha^*,H_0,\boldsymbol{x},\boldsymbol{z}\right) \\ &\quad\propto(\alpha^*)^{K+M}\prod_{t=t_0}^{T}\prod_{k=1}^{K_t}\sigma(Y_{t,k})\ell(\boldsymbol{x}_{t,k}|\boldsymbol{\theta}_{t,k})H_0(\boldsymbol{\theta}_{t,k}) \\ &\qquad\cdot\mathcal{GP}(\boldsymbol{Y})\prod_{t=t_0}^{T}\prod_{m=K_t+1}^{K_t+M_t}\sigma(-Y_{t,m})H_0(\boldsymbol{\theta}_{t,m}). \end{aligned} \tag{7}$$

Equation (7) includes the posterior probability for the whole topic variables $\vec{\boldsymbol{\theta}}_K\triangleq\{\boldsymbol{\theta}_i\}_{i=1}^{K}$ and the prior probability for the whole latent variables $\vec{\boldsymbol{\theta}}_M\triangleq\{\boldsymbol{\theta}_i\}_{i=K+1}^{M+K}$. The set of entire GP functions is denoted as $\boldsymbol{Y}$, and $Y_{t,k}$ denotes the GP function at topic $\boldsymbol{\theta}_{t,k}$. The total number of topics K is the sum of topic numbers among patches, i.e., $K=\sum_{t=t_0}^{T}K_t$. Moreover, the number of total latent variables M follows a similar setting.

SGP functions of topics and latent variables are inferred with the log posterior probability, which takes the form

$$\ln p(\boldsymbol{Y}|\vec{\boldsymbol{\theta}})=-\sum_{i=1}^{K+M}\ln\left(1+e^{\mathrm{sgn}(i')Y_i}\right)-\frac{1}{2}\boldsymbol{Y}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}+c \tag{8}$$

where c is a constant, $i'=i-K+0.5$, and $\boldsymbol{\Sigma}\in\mathbb{R}^{(M+K)}$ denotes the covariance matrix of topics. It consists of kernel $\kappa(\boldsymbol{\theta},\boldsymbol{\theta}';\boldsymbol{\eta})$ with hyper-parameter $\boldsymbol{\eta}$, which is a squared-exponential (SE) covariance function [17].

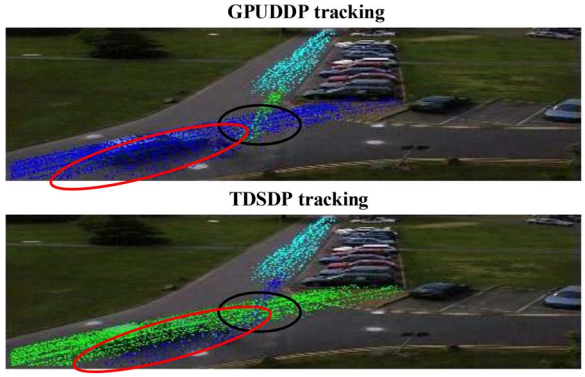

Fig. 3. Tracking results by GPUDDP and TDSDP.

D. Trajectories Clustering Step (Green Dotted Lines)

Here, a hierarchical SGP prior clustering step is provided.

6) The global SGP prior can be used for clustering trajectories without initial localization. The inferred SGP intensity with topics $\{\boldsymbol{\theta}_i\}_{i=1}^{K}$ is clustered into L independent SGPs, such as the overlapped topic trajectories in the left panel of Fig. 2. The trajectories of cars and humans in the right panel are clustered data points.

The probability of the assignment c_i for the topic $\boldsymbol{\theta}_i$ follows:

$$p(c_i = l) \propto \begin{cases} g_{-i,l} \sum\limits_{c_j = l} f(d_{ij}), & \text{if } l \leqslant L \\ \gamma, & \text{if } l = L + 1 \end{cases} \tag{9}$$

where i denotes topic index, L denotes the current number of topic trends, and d_{ij} is the distance between topics $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$. Based on the distance-based Chinese Restaurant Process [26], L can be automatically learned with the hyper-parameter γ controlling the probability of sampling a new cluster. The window decay function $f(d_{ij}) = \sigma(a - d_{ij})\text{rect}((d_{ij} - a/2)/a)$ contains the delay parameter a, the sigmoid function $\sigma(\cdot)$ and the rectangular function $\text{rect}(\cdot)$. This indicates that only topics with distance less than a are taken into account. Equation (9) indicates that the probability of topic $\boldsymbol{\theta}_i$ assigned to the lth cluster is affected by two factors: this cluster's topic density in the range of a from topic $\boldsymbol{\theta}_i$, and the kernel function calculated by topics $\{\boldsymbol{\theta}_i\}_{c_i=l}$ in the same cluster. Then a kernel classifier g [18] with the SE kernel $\kappa(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j; \boldsymbol{\eta})$ is used to make L GP gating network independent, which follows:

$$g\left(\{\boldsymbol{\theta}_j\}_{c_j=l, j\neq i}\right) = \frac{\sum_{j\neq i} \kappa(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j; \boldsymbol{\eta}) \delta(c_j, l)}{\sum_{j\neq i} \kappa(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j; \boldsymbol{\eta})}. \tag{10}$$

Fig. 3 shows the performance of the Generalized Polya Urn based Dependent Dirichlet Process (GPUDDP) [27] and the TDSDP, which are depicted in top and bottom panels, respectively. In the bottom panel, green and cyan points represent two cars, which meet each other half way. Blue points represent a human trajectory. Fig. 3 shows that the TDSDP models global topics with the SGP prior better than the GPUDDP. The GPUDDP depends on a birth and death mechanism. When a topic is lost in one frame, the following topics would be wrongly assigned (as shown in the top panel with red oval) or a new cluster is generated. This problem is the limitation of the DDP model based on the Markov assumption. In contrast, the proposed TDSDP with the SGP prior permits mistakes of the topic indicators in the overlapping area (shown in the bottom panel with a black oval). Moreover, topics can return to the correct clusters in later frames (the red oval). The experimental performance on insufficient features (only position and temporal stamp) illustrates the robustness of the TDSDP mixture model.

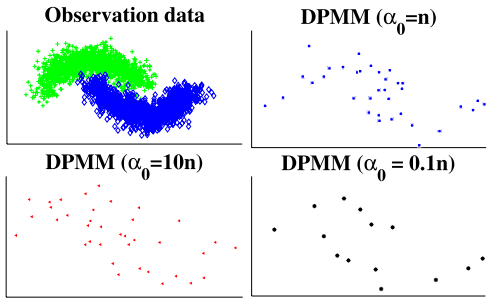

Fig. 4. Top-left panel shows observation data, and the other three panels show topic sampled by DPMM with hyperparameter $\alpha_0 = 10n$, n, and $0.1n$, respectively.

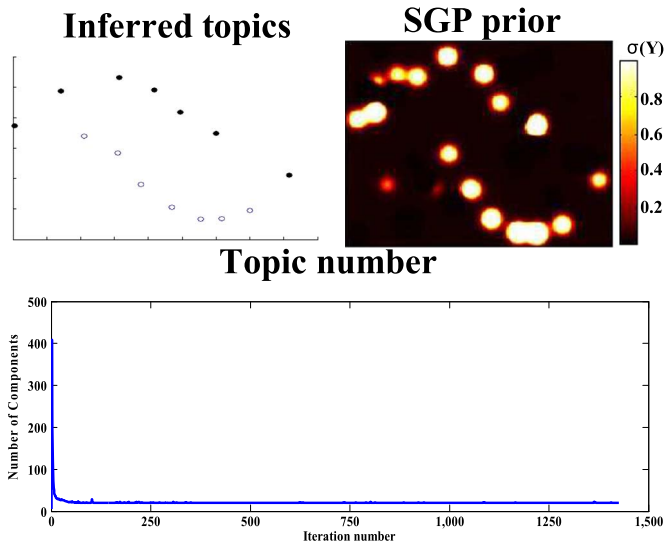

Fig. 5. Top panels show the inferred topics and SGP prior, and the bottom panel shows the topic number converged at $K = 14$.

IV. Experiments

A. Data Synthesis Experiment

We synthesize the overlapped data for two topic flows as shown in Fig. 4. The DPMM and the TDSDP are used to model the topic flows, and the results are shown in Figs. 4 and 5, where α is the concentration parameter, and the number of data points is denoted as n. In Fig. 4, the results are sensitive to different hyper-parameters for DPMM. Since updating of parameter α is unstable for overlapped data, the number of topics is also inconsistent. Topics inferred by DPMM are mixed together for $\alpha_0 = 10n$ and n, while topics will be separated far away when hyper-parameter α_0 is set smaller, $\alpha_0 = 0.1n$. As a result, it is difficult to cluster topic trend when α_0 seriously impacts the result. In contrast, the TDSDP mixture (single DP with the SGP prior) has a more robust result. This is due to a better fitting SGP prior in the top panels in Fig. 5 (hotter point $\boldsymbol{\theta}$ indicates its SGP function is larger). After convergence, a consistent number of the components estimated with thinned concentration parameters is shown in the bottom panel. This consistent and accurate topic inference enables us to model a topic trajectory better.

B. PETS2001 Dataset

We use the whole PETS2001 dataset [22] in this experiment. Similar to the setting presented in Section III, data points come from the interest points extracted in the video. To analyze the situation of

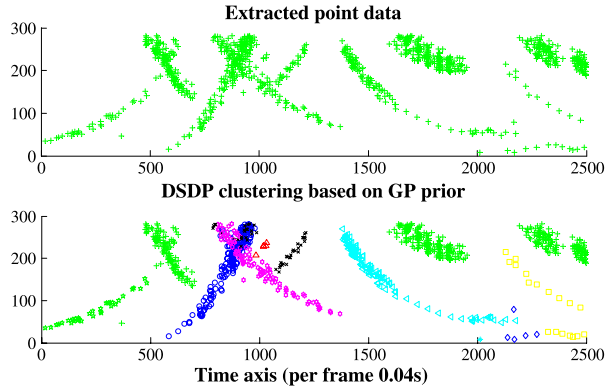

Fig. 6. Topic trajectories clustering for PETS2001.

TABLE I
RESULT COMPARISON

	GPUDDP	TDSDP	TDSDP	TDSDP
α_0	/	$0.1n$	$0.5n$	n
SFDA	0.925	0.905	0.955	0.940
ATA	0.282	0.810	0.914	0.870

insufficient features, only position and temporal information is used. Intuitively, without color or other features to rely upon, it becomes difficult to track object flows in the case of occlusion, especially for DDP based on the Markov assumption, where the birth and death mechanism causes the object trajectories to be interrupted easily. Since only position feature is used, we focus on estimating object flow trajectories, whereas the pedestrians always walking together are regarded as one object flow. Observation data (extracted interest points) and SGP prior clustering result for PETS2001 are shown in the top and bottom panels of Fig. 6, respectively. The horizontal axis shows the frame number and the vertical axis shows the first dimension of interest points. This experiment illustrates that TDSDP (DP priors share a SGP intensity measure) fits the dependency of topics better in the temporal domain (local DP priors) with global SGP prior.

Table I shows results of the proposed TDSDP and the Generalized Polya Urn based DDP tracking [27] with insufficient features. Sequence Frame Detection Accuracy (SFDA) and Average Tracking Accuracy (ATA) [28] are used to quantify the experimental results. SFDA measures the detection performance in each frame, and ATA quantifies the performance of a tracking algorithm for detecting objects across frames. The best result of the GPUDDP is shown in Table I, which is obtained by tuning parameter values for some samples from this dataset. The results of the TDSDP are shown with various hyper-parameters α_0 from $0.1n$ to n, where n is the average number of data points in each patch. Table I indicates that at the frame level, TDSDP and DDP methods have similar detection accuracy (i.e. 95.55% and 92.53%). However, TDSDP has a far superior tracking performance compared to the GPUDDP in terms of ATA. This result further illustrates that the robustness of TDSDP with the global SGP prior fits the varying topics better than the Markov assumption based DDP models, which have difficulties coping with interruption caused by occlusions. Unlike existing GP clustering based tracking algorithms, which require labeled object topics (i.e. labeled object centers in [19]). The TDSDP prior excels in providing a global object topic flow along temporal domains.

C. Low Quality Fish Tracking (LQFT) Dataset

A new low-quality dataset [29] for fish tracking is also tested here. Three sample frames with object detection are given in the top panel in Fig. 7, and tracked overlapping fish trajectories is shown in the bottom panel. Similar to the circumstance of the night transportation system, this surveillance video is captured with poor visibility. Moreover, fishes usually move in and out of the field of view rapidly, and the number of fishes is varying all the time. ATA of the TDSDP achieves 95.3% compared to 69.7% of the GPUDDP, while SFDAs of these models are similar (i.e. 98.47% and 96.5%). Experiments on this noisy and low-quality dataset illustrate the robustness of the proposed TDSDP mixture model.

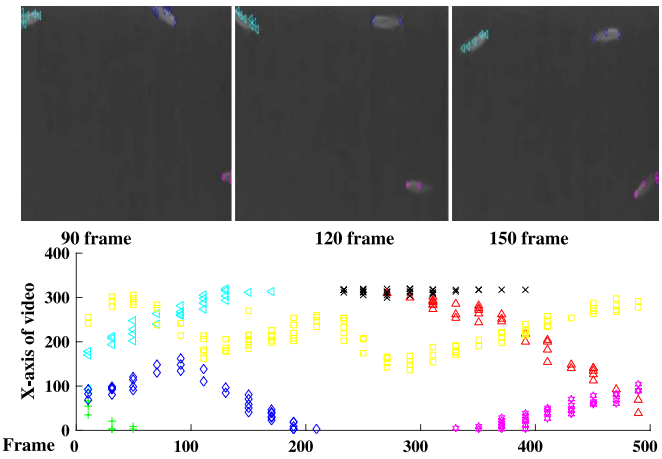

Fig. 7. Example frames and topic trajectories of the LQFT data set.

D. Computational Cost and Hyper-Parameter Learning

The computational cost includes two parts. The cost of the DP sampling step is $\mathcal{O}(n^2/c)$ and the cost of the SGP sampling step is $\mathcal{O}(c^3 K_t^3)$, in which c is the number of patches and n is the number of all data points. Strictly speaking, K_t is the number of topics and latent variables $(K+M)$ in each patch. In practice, K and M are of the same order [24], so we can regard K_t as the average number of topics in each patch. An optimal computation cost follows:

1) If the video is split into patches with the optimal number $c^* = K_t^{-1}\,(n/2)^{2/3}$, by substituting it into costs of DP and SGP steps above, the optimal cost follows $\mathcal{O}(n^{4/3} K_t)$.
2) If the frame number f is smaller than the optimal patch number c^*, i.e. $f < c^*$, then the computational cost is approximate the cost of the DP step: $\mathcal{O}(n^2 f^{-1})$.

Both costs are smaller than the DDP based models [13], [27], which are on the order of $\mathcal{O}(n^3)$. For the PETS2001, we set $f/c \approx 15$ and divide the video (2500 frames) into 150 patches.

Detailed inference of the hyper-parameter α^* follows:

$$\alpha^* \sim p\,(\alpha^* | K+M, n, \alpha_0, \beta_0) \tag{11}$$

where n is the number of data. While parameters $\alpha_0 = n/20$ and $\beta_0 = 1$ in [30], in our model, we set $\alpha_0 = 0.5n$ to ensure that the upper bound α^* is large enough. Moreover, Table I demonstrates that the TDSDP mixture model is robust with various α_0 varying from $0.1n$ to n. The initial hyper-parameter is sampled from the initial distribution $\alpha^* \sim \text{Gamma}(\alpha_0, \beta_0)$.

V. CONCLUSION

An unsupervised tracking algorithm is presented based on the Temporal Doubly Stochastic Dirichlet Process mixture model. The global intensity prior provides the TDSDP the ability to handle the complex transportation system, which includes occlusion of the human crowds

and vehicles. Moreover, the TDSDP enables modeling of the trajectories of human crowds and vehicles, whose number and positions are varying along temporal frames. Lastly, the thinning procedure enables the TDSDP to significantly reduce the computational cost.

REFERENCES

[1] M. Shan, S. Worrall, and E. Nebot, "Probabilistic long-term vehicle motion prediction and tracking in large environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 539–552, Jun. 2013.

[2] C. Pang, W. Lam, and N. Yung, "A novel method for resolving vehicle occlusion in a monocular traffic-image sequence," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 3, pp. 129–141, Sep. 2004.

[3] C. C. C. Pang, W. W. L. Lam, and N. H. C. Yung, "A method for vehicle count in the presence of multiple-vehicle occlusions in traffic images," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 441–459, Sep. 2007.

[4] D. Vasquez, T. Fraichard, and C. Laugier, "Incremental learning of statistical motion patterns with growing hidden Markov models," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 403–416, Sep. 2009.

[5] B. Morris and M. Trivedi, "Learning, modeling, and classification of vehicle track patterns from live video," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 425–437, Sep. 2008.

[6] D. Lin, E. Grimson, and J. W. Fisher, "Construction of dependent Dirichlet processes based on Poisson processes," in *Proc. NIPS*, 2010, pp. 1396–1404.

[7] X. Wang and A. McCallum, "Topics over time: A non-Markov continuous time model of topical trends," in *Proc. SIGKDD*, 2006, pp. 424–433.

[8] A. Dubey, A. Hefny, S. Williamson, and E. P. Xing, "A nonparametric mixture model for topic modeling over time," in *Proc. SIAM Int. Conf. Data Mining*, 2013, pp. 530–538.

[9] V. Rao and Y. W. Teh, "Spatial normalized gamma processes," in *Proc. NIPS*, 2009, pp. 1554–1562.

[10] A. E. Gelfand, A. Kottas, and S. N. MacEachern, "Bayesian nonparametric spatial modeling with Dirichlet process mixing," *J. Amer. Statist. Assoc.*, vol. 100, no. 471, pp. 1021–1035, Sep. 2005.

[11] C. E. Antoniak, "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," *Ann. Statist.*, vol. 2, no. 6, pp. 1152–1174, Nov. 1974.

[12] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *Ann. Statisti.*, vol. 1, no. 6, pp. 209–230, Mar. 1973.

[13] S. N. MacEachern, "Dependent Dirichlet Processes," Dept. Statist., Ohio State Univ. Columbus, OH, USA, Tech. Rep., 2000.

[14] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1566–1581, Dec. 2006.

[15] C. Chen, V. Rao, W. Buntine, and Y. Whye Teh, "Dependent normalized random measures," in *Proc. ICML*, 2013, pp. 969–977.

[16] C. Guo, S. Mita, and D. McAllester, "Robust road detection and tracking in challenging scenarios based on Markov random fields with unsupervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1338–1354, Sep. 2012.

[17] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006, pp. 715–719.

[18] C. E. Rasmussen and Z. Ghahramani, "Infinite mixtures of Gaussian process experts," in *Proc. NIPS*, 2002, pp. 881–888.

[19] J. Ross and J. Dy, "Nonparametric mixture of Gaussian processes with constraints," in *Proc. ICML*, 2013, pp. 1346–1354.

[20] X. Sun, N. H. C. Yung, E. Y. Lam, and H. K.-H. So, "Computationally efficient hyperspectral data learning based on the Doubly Stochastic Dirichlet process," *IEEE Trans. Geosci. Remote Sens.*, to be published.

[21] D. R. Cox, "Some statistical methods connected with series of events," *J. Royal Statist. Soc. Ser. B*, vol. 17, no. 2, pp. 129–164, 1955.

[22] PETS2001 Dataset. [Online]. Available: ftp://ftp.pets.rdg.ac.uk/pub

[23] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Proc. Vis. Surveillance Perform. Eval. Tracking Surveillance*, 2005, pp. 65–72.

[24] R. P. Adams, I. Murray, and D. J. MacKay, "Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities," in *Proc. ICML*, 2009, pp. 9–16.

[25] I. Murray, Z. Ghahramani, and D. J. C. MacKay, "MCMC for doubly-intractable distributions," in *Proc. UAI*, 2006, pp. 359–366.

[26] D. M. Blei and P. I. Frazier, "Distance dependent Chinese restaurant processes," *J. Mach. Learn. Res.*, vol. 12, pp. 2383–2410, 2011.

[27] W. Neiswanger, F. Wood, and E. Xing, "The dependent Dirichlet process mixture of objects for detection-free tracking and object modeling," in *Proc. AISTAT*, 2014, pp. 660–668.

[28] R. Kasturi *et al.*, "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 319–336, Feb. 2009.

[29] LQFT Dataset. [Online]. Available: www.eee.hku.hk/~sunxing/Data

[30] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *J. Amer. Statist. Assoc.*, vol. 90, no. 430, pp. 577–588, 1995.