Noname manuscript No.
(will be inserted by the editor)

Trajectory Analysis and Semantic Region Modeling Using Nonparametric Hierarchical Bayesian Models*

Xiaogang Wang · Keng Teck Ma · Gee-Wah Ng · W. Eric L. Grimson

Received: date / Accepted: date

Abstract We propose a novel framework of using a nonparametric Bayesian model, called Dual Hierarchical Dirichlet Processes (Dual-HDP) [2], for unsupervised trajectory analysis and semantic region modeling in surveillance settings. In our approach, trajectories are treated as documents and observations of an object on a trajectory are treated as words in a document. Trajectories are clustered into different activities. Abnormal trajectories are detected as samples with low likelihoods. The semantic regions, which are subsets of paths commonly taken by objects and are related to activities in the scene, are also modeled. Under Dual-HDP, both the number of activity categories and the number of semantic regions are automatically learnt from data. In this paper, we further extend Dual-HDP to a Dynamic Dual-HDP model which allows dynamic update of activity models and online detection of normal/abnormal activities. Experiments are evaluated on a simulated data set and two real data sets, which include $8,478$ radar tracks collected from a maritime port and $40,453$ visual tracks collected from a parking lot.

Keywords Visual surveillance · Activity analysis · Trajectory analysis · Scene modeling · Abnormality detection · Nonparametric hierarchical Bayesian models · Clustering · Gibbs sampling.

* Part of this work was published in [1]

Xiaogang Wang
the Department of Electronic Engineering, the Chinese University of Hong Kong
E-mail: xgwang@ee.cuhk.edu.hk

Keng Teck Ma
National University of Singapore E-mail: ktma@nus.edu.sg

Gee-Wah Ng
DSO National Laboratories E-mail: ngeewah@dso.org.sg

W. Eric L. Grimson
the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology E-mail: welg@csail.mit.edu

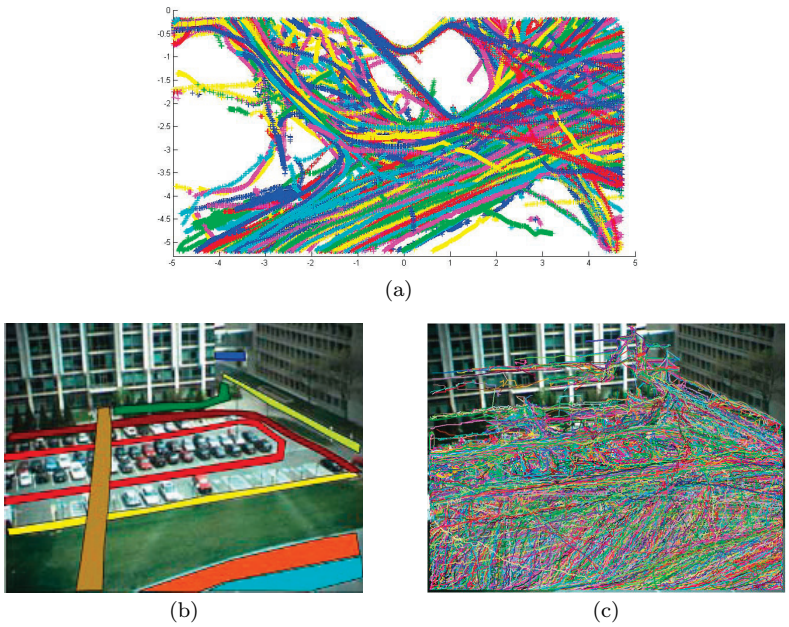


Fig. 1 Trajectories in our two data sets. (a) Radar tracks collected from a port. (b) Some examples of paths in the parking lot scene. They are manully drawn for the illustration purpose. (c) Tracks collected from a parking lot scene within one week.

1 Introduction

Activity analysis has long been one of the foci of research in surveillance. Over the past decade significant works [3]-[21] have been reported on this topic. Many of these approaches assumed that objects and/or their constituents were first detected and tracked throughout the scene and activities were modeled as sequences of movements of objects. In near-field settings, the features of gestures, poses, and appearance play an important role in explaining activities. However, in many far-field settings (i.e. wide outdoor areas), the captured videos are of low resolution and poor quality or even no videos are available (e.g. in some maritime surveillance, only radar signals are available). In these scenarios, it is difficult to compute more complicated features. Usually only positions of objects are recorded along the tracks, which are called trajectories. The majority of visible activities are distinguished by the patterns of objects moving from one location to another. In this work, our surveillance system is based on trajectory analysis.

Activities are closely related to scene structures, such as paths, entry and exit points, since they regularize the motion of objects. Some examples of paths can be found in Figure 1(b). One the one hand, these structures can be identified from trajectories related to particular activities [10, 11, 22, 17, 23, 19, 20, 24]. On the other hand, the knowledge of scene structures help to classify trajectories into activities, since it provides prior information on activities happening in a scene. In this paper the two related problems of activity analysis and scene modeling will be jointly solved. Developing algorithms automatically solving these two problem without human intervention

will save a lot of man power in surveillance applications. In same cases, it is very difficult for human beings to tell how many activity categories and scene structures exist in the scene, especially when the data sets are in large scale and activities are complicated (see examples shown in Figure 1 (a) and (c)), and therefore the assistance of computing algorithms becomes even more important[1]. The knowledge of the learnt scene structures is very useful in many surveillance tasks. It can support activity descriptions with spatial context, such as "a car moving off the road" and "a person waiting at a bus stop", and improves low-level tracking and classification [25]. For example, if an object disappears, but not at an exit point, then it is likely to be a tracking failure instead of a true exit. In classification, people can leverage the fact that vehicles are much more likely than pedestrians to move on the road.

In surveillance, it is easy to collect a huge amount of data over weeks, months or even years, as more and more cameras are installed in urban settings. People expect that the surveillance system can deal with a huge amount of data, process the data with as little human effort as possible, store and query data in an efficient way, and provide statistical summaries of activities. So an unsupervised or semi-supervised framework is preferred. In this paper, we propose an surveillance system with the following features:

- Cluster trajectories into different activities without supervision.
- Detect abnormal trajectories.
- Model semantic regions, which are subsets of paths in the scene.
- Online dynamically update the models of activities. Instead of keeping all the data for clustering purposes as existing methods do, in our system old trajectories can be replaced overtime.
- Handle a huge amount of data.

We propose a framework of using a nonparametric Bayesian model, Dual Hierarchical Dirichlet Processes (Dual-HDP), which was proposed in [2], for trajectory analysis. Dual-HDP advances the existing language processing model, Hierarchical Dirichlet Processes (HDP) [26] . HDP is a nonparametric Bayesian model. It clusters words often co-occurring in the same documents into one topic and automatically decides the number of topics. Dual-HDP co-clusters words and documents, and it automatically decides the numbers of both word topics and document clusters. Dual-HDP is similar to the nonparametric model, called Nested Dirichlet Process, proposed by Rodriguez et al. [27]. It is also closely related to the Transformed Dirichlet Process proposed by Sudderth et al. [28] applied to object recognition. Under our framework, trajectories are treated as documents and the observations (positions and moving directions of objects) on the trajectories are treated as words. Topics model the semantic regions, which are subsets of paths commonly taken by objects, in the scene, and trajectories are clustered into different activities. HDP can only cluster observations into semantic regions. Since trajectories have different combinations of semantic regions, Dual-HDP has an extra layer of hierarchical Dirichlet processes to model the clusters of trajectories on the top of semantic regions. In this paper we further extend Dual-HDP to dynamic Dual-HDP which allows the models of activities and semantic regions to be online dynamically updated. Dynamic Dual-HDP is related to the dynamic topic model proposed by Blei et al. [29], which was a parametric model assuming that the number of topics are fixed.

[1] Some scene structures cannot be identified from the appearance of the scene, such as the path crossing the grass field in Figure 1 (b). In some cases, the background image of the scene is even not available (for example, as shown in Figure 1 (a), in maritime surveillance, only radar tracks are available).

We evaluate our approach on a simulated data set and two real data sets (see Figure 1), which include 8,478 radar tracks collected from a port under maritime surveillance and 45,453 video tracks collected from a parking lot scene. In maritime surveillance, trajectory analysis is a natural way to analyze activities especially when only radar signals are available. Without expert knowledge, it is very difficult for humans to discover transportation structures on the sea, such as shipping fairways, since the appearance of the scene does not help. The tracks from the parking lot scene are obtained from far-field videos recorded by a fixed camera. We use the Stauffer-Grimson tracker [12] to obtain tracks in this data set. Both data sets have tracking errors. For example, a long track may be broken into segments because of occlusion and different objects are linked as one track because of incorrect data association.

2 Related Work

Probabilistic approaches were widely applied to object detection, tracking and event detection in visual surveillance [15,30–35]. Nillius et al. [34] used a Bayesian network to associate the identities of isolated tracks. Oliver et al. [31] used Coupled HMM to model the interaction between two objects. However, there are few studies on using graphical models to cluster tracks of objects into motion patterns. Pang et al. [35] proposed a Bayesian filtering framework to group targets which are moving together in similar directions and are close in space. It was evaluated on a very small data set only including four trajectories. This approach did not cluster whole trajectories since the group identities of targets might change dynamically. It only grouped targets moving at the same time, which means that trajectories were temporally aligned. In order to group targets observed at different time using this approach, trajectories have to be first aligned, which is one of the major difficulties in clustering trajectories since targets might be misdetected during some time windows and trajectories might be broken or associated incorrectly because of tracking errors. Our models do not require the alignment of trajectories.

Many of the existing trajectory analysis approaches cluster trajectories and detect abnormal trajectories by defining the pairwise similarities between trajectories. The proposed trajectory similarities or distances include Euclidean distance [36,37], Hausdorff distance and its variations [23,19], and Dynamic Time Warping (DTW) [38]. Some approaches required that two trajectories are aligned when computing their distance. An alignment method, long common subsequence (LCSS) were proposed in [39]. Based on the computed similarity/distance matrix, some standard clustering algorithms such as spectral clustering [40], agglomerative and divisive hierarchical clustering [41], and fuzzy c-means [20] were used to group trajectories into different activity categories. A comparison of different distance measurements and clustering methods can be found in [42]. These similarity/distance-based approaches have several drawbacks. First, there is no global probabilistic framework to model activities happening in the scene. They have an *ad hoc* nature especially on the definitions of distance measures. Abnormal trajectories are usually detected as those with a larger distance to other trajectories. Their abnormality detection lacks a probabilistic explanation. Second, they usually do not provide a solution to the number of clusters. They often require that the cluster number is known in advance. Third, some approaches required that two trajectories were temporally aligned when their distance was computed, which is difficult because of misdetection and tracking errors. Fourth, calculating the similarities between all pairs

of samples is computationally inefficient, with a complexity of $O(N^2)$ in both time and space, where N is the number of trajectories. Although some approaches, such as the Nyström method, approximate some clustering methods, such as spectral clustering, with fewer samples [43], if the selected samples are not representative enough, the approximation result may not be good. Some other approaches were proposed in recent years. For example, Anjum and Cavallaro [44] extracted a set of representative features from trajectories, clustered feature vectors using mean-shift, and merged similar adjacent clusters. Zhang et al. [45] used a quadratic curve to fit a trajectory. The parameters of the quadratic curve form a feature vector to represent the trajectory. Under this representation, motion patterns within each spatial block were learned by the Gaussian mixture model and motion patterns were clustered by a graph-cut algorithm. Saleemi et al. [46] modeled the motion patterns of objects in the form of a multivariate nonparametric probability density function of spatiotemporal variables. The model was learned using kernel density estimation.

Many other trajectory clustering approaches [47–50] have been proposed and applied to motion segmentation and object counting. In their applications, these approaches assumed that the trajectories were temporally aligned and the correspondence of points between trajectories were automatically established. However, this assumption does not hold in activity analysis and semantic region modeling, and therefore they cannot be directly applied to solve our problem.

Trajectory clustering is also related to the problem of modeling semantic regions in the scene. It takes a lot of effort to manually input these structures, and they cannot be reliably detected based on the appearance of the scene.

There has been a lot work [51–54] on time dependent Dirichlet Process (DP) models published in recent years. Griffin et al. [52] proposed a framework, called Order-Based Dependent Dirichlet Processes (DDP), to model time series data. Caron et al. [53] introduced a class of time-varying DP mixture models using a generated polya urn scheme. These works modeled time dependency of DP mixtures without more complicated hierarchical structures (such as HDP). The work most relevant to us is [55] and [56]. Ren et al. [55] proposed a Dynamic Hierarchical Dirichlet Process model which was applied to music segmentation and analysis of gene expressions. Srebro et al. [56] integrated Ordered-Based DDP [52] into hierarchical topic models. Both [55] and [56] assumed that the data of different time slices all share the same set of topics, which are fixed over time and they only modeled the dynamic change of the mixture weights of topics. However, in our problem it is important to model the dynamic change of topics, which reveals the change of the spatial distribution of semantic regions over time. This will be shown by our experimental results. Thus although the models in [55] and [56] achieved success when modeling music and genes, they are not suitable for trajectory analysis and semantic region modeling. Our dynamic Dual-HDP model allows to dynamically update both the models of topics (semantic regions in our problem) and the mixture weights over topics, and it can better fit data at different time slices. Another importance difference is that learning the models in [55] and [56] require loading the data observed in all the time slices altogether and running in a batch mode, because they assumed that all the time slices share the same set of topics. However, our dynamic Dual-HDP is learned incrementally and runs in an online mode. The historical data is replaced by new data observed in the current time slice wihout being kept in the memory. Its complexity is much lower. Furthermore, dynamic Dual-HDP with two layers of hierarchical Dirichlet processes has a more complicated hierarchical structure than dynamic HDP in [55,56].

In the computer vision field, hierarchical Bayesian models [57,58,26,59,60] have been widely applied with success in recent years. They were used to solve the problems of scene categorization [61,28], object recognition [62,28,60], human action recognition [63,64] and video analysis [2]. Fox and Willsky et al. [32] used Dirichlet process to solve the problem of data association for multi-target tracking in the presence of an unknown number of targets. In this paper, we use a nonparametric Hierarchical Bayesian model for trajectory analysis and scene modeling.

Our framework differs from previous trajectory analysis and scene modeling approaches:

- Different from existing distance-based clustering approaches, it clusters trajectories using a generative model. There is a natural probabilistic explanation for the detection of abnormal trajectories.
- Previous approaches first clustered trajectories into activities and then segmented semantic regions. Our approach simultaneously learns activities and semantic regions, which are jointly modeled in *Dual-HDP*.
- Using Dirichlet Processes, the numbers of activity categories and semantic regions are automatically learnt from data instead of being manually set.
- It does not require trajectories to be temporally aligned.
- The space complexity of our algorithm is $O(N)$ instead of $O(N^2)$ in the number of trajectories.
- Using dynamic Dual-HDP, the models of activities and semantic regions can be dynamically updated, and clustering trajectories and detecting abnormal trajectories can be done in an online mode. Compared with Dual-HDP, trajectories are processed incrementally. Old trajectories are replaced over time. The space and time complexities are further reduced. It can process data over a very long period.
- Our approach clusters trajectory through modeling semantic regions. Different from distance-based methods, which cluster trajectories close in space, in our model two locations are in the same semantic region if they are connected by many trajectories. Considering the case when vehicles move on two side-by-side lanes in the same direction, some distance-based methods may group trajectories of these vehicles into one cluster while our approach will learn the two lanes as different semantic regions and separate trajectories into two clusters since locations on different lanes are rarely connected by trajectories.

3 Modeling Trajectories

We treat a trajectory as a document and the observations on the trajectory as words. The positions and moving directions of observations on a trajectory are computed as features which are quantized according to a codebook. The codebook uniformly quantizes the space of the scene into small cells and the velocity of objects into several directions. A trajectory is modeled as a bag of quantized observations without temporal order. In language processing, some topic models, such as LDA [58] and HDP [26], cluster co-occurring words into one topic. Each topic has a discrete distribution over the codebook. A document is modeled as a mixture of topics and documents share topics. If some words, such as "professor" and "university", often but not necessarily always occur in the same documents, a topic related to "education" will be learnt and its distribution has large weights on both "professor" and "university".

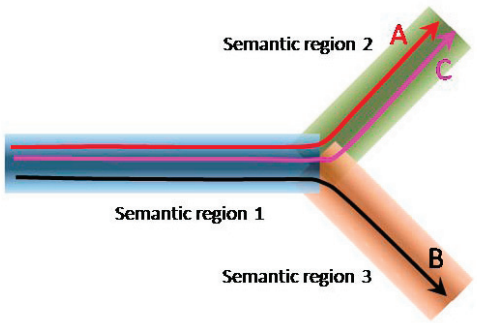


Fig. 2 An example to explain the modeling of semantic regions and activities. See details in text.

In the physical world, objects move along some paths. We refer to the subsets [2] of paths as semantic regions, i.e. two paths may share one semantic region as shown in Figure 2. When topic models are used to model trajectories, topics reveal semantic regions shared by trajectories, i.e. many trajectories pass through one semantic region with common directions of motion. A semantic region is modeled as a discrete distribution over the space of the scene and moving directions. If two trajectories pass through the same combination of semantic regions, they are on the same path and thus they belong to the same activity category. In our *Dual-HDP* model, each activity category corresponds to a path and has a prior distribution over semantic regions. It is learnt in an unsupervised way. All the trajectories clustered into the same activity category share the same prior distribution. Using Dirichlet Processes, Dual-HDP can learn the number of semantic regions and the number of activities from data.

In Figure 2, an example is shown to explain the modeling. There are three semantic regions (indicated by different colors) which form two paths. Both trajectories A and C pass through regions 1 and 2, so they are clustered into the same activity. Trajectory B passes through regions 1 and 3, so it is clustered into a different activity. To help readers better understand our approach, the concept mapping between surveillance and language processing is summarized in Table 1.

With the "bag-of-words" assumption, our approach does model the first order temporal information among observations since the codebook encodes the moving directions. If the locations of observations keep unchanged but their temporal orders are permuted, the observations will be assigned to different words because their moving directions are changed. It can distinguish some activities related to temporal features. For example, if objects visit several regions in opposite temporal order, they must pass through the same region in opposite directions. In our model, that region splits into two topics because of the velocity difference. So these two activities can be distinguished by our model, since they have different topics.

In Section 5 and 6, we will explain the HDP model proposed by Teh et al. [26] and Dual-HDP model [2], which is actually used for trajectory analysis. In Section 7 a new dynamic Dual-HDP model will be proposed. We will describe them as language models. However, remember that in our problem documents are trajectories, words are

[2] If a path is viewed as a set of quantized spatial locations and moving directions, semantic regions are subsets of paths and they can obtained through the operations of intersection and set difference between paths.

Trajectories	Documents
Observations on trajectories	words
Semantic regions	Topics
Activity (path) models	Combinations of topics

Table 1 Concept mapping between surveillance and language processing.

observations, and topics are semantic regions. Clusters of trajectories (activities) are explicitly modeled in Dual-HDP and dynamic HDP but not in HDP.

4 Dirichlet Process

Dirichlet Process (DP) [65] is used as a prior to sample probability measures in non-parametric Bayesian methods. It is defined by a concentration parameter α, which is a positive scalar, and a base probability measure H (for example H is a Dirichlet distribution in our case). A probability measure G randomly drawn from $DP(\alpha, H)$ is always a discrete distribution and it can be obtained from a stick-breaking reconstruction [66],

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}. \tag{1}$$

ϕ_k is a multinomial parameter vector sampled from Dirichlet distribution H, $\phi_k \sim H$. δ_{ϕ_k} is a Dirac delta function centered at ϕ_k. π_k is a non-negative scalar satisfying $\sum_{k=1}^{\infty} \pi_k = 1$, and it is constructed by $\pi_k = \pi'_k \prod_{l=1}^{k-1}(1-\pi'_l)$, $\pi'_k \sim Beta(1, \alpha)$. G often serves as a prior for infinite mixture models, which can be used to cluster data. Let $\{w_i\}$ be a set of observed data points. Under an infinite mixture model, w_i is sampled from a density function $p(\cdot|\theta_i)$, which is one of the ϕ_ks in Eq (1) and is sampled from G. Data points sharing the same parameter vector ϕ_k are clustered together under this mixture model. Given parameter vectors $\theta_1, \ldots, \theta_N$ of N data points $w_1, \ldots, w_N$, the parameter vector θ_{N+1} of a data point w_{N+1} can be sampled from a posterior by integrating out G,

$$\theta_{N+1}|\theta_1, \ldots, \theta_N, \alpha, H \sim \sum_{k=1}^{K} \frac{n_k}{N+\alpha}\delta_{\theta^*_k} + \frac{\alpha}{N+\alpha}H. \tag{2}$$

There are K distinct values $\{\phi_k\}_{k=1}^{K}$ (identifying K components) among the $\theta_1, \ldots, \theta_N$. n_k is the number of points whose parameter vectors are ϕ_k. θ_{N+1} can be assigned to one of the existing components $\{\phi_k\}_{k=1}^{K}$ (w_{N+1} is assigned to one of the existing clusters) or can sample a new component ϕ_{K+1} from H (a new cluster is created for w_{N+1}). The posterior of θ_{N+1} is

$$p(\theta_{N+1}|w_{N+1}, \theta_1, \ldots, \theta_N, \alpha, H) \approx p(w_{N+1}|\theta_{N+1})p(\theta_{N+1}|\theta_1, \ldots, \theta_N, \alpha, H). \tag{3}$$

It is likely for the infinite mixture model with DP prior to create a new component if existing components cannot well explain the data w_{N+1}. There is no limit to the number of components. These properties make DP ideal for modeling data clustering problems where the number of mixture components is not well-defined in advance. A more detailed description of DP can be found in [69].

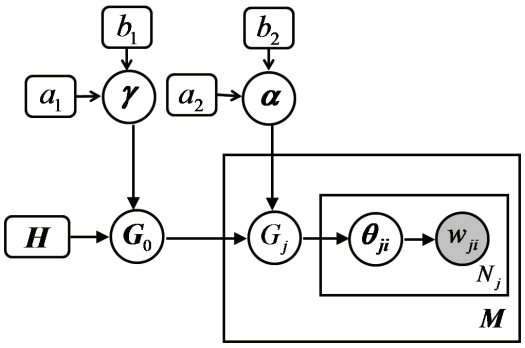


Fig. 3 The graphical model of *HDP*.

5 HDP

HDP proposed by Teh et al. [26] is a nonparametric hierarchical Bayesian model used to cluster co-occurring words in documents into topics (in our problem it clusters observations on trajectories into semantic regions).

The graphical model of HDP is shown in Figure 3. There are M trajectories. Each trajectory j has N_j quantized observations of positions and moving directions of objects. In HDP, a prior distribution G_0 over the whole data set is sampled from a Dirichlet process, $G_0 \sim DP(\gamma, H)$. $G_0 = \sum_{k=1}^{\infty} \pi_{0k} \delta_{\phi_k}$ is an infinite mixture in which $\{\phi_k\}_{k=1}^{\infty}$ (discrete distributions over quantized locations and moving directions) are the models of semantic regions and $\{\pi_{0k}\}_{k=1}^{\infty}$ are the mixture weights over semantic regions. Therefore, a scene is modeled as an infinite mixture of semantic regions. Observations on trajectories will be clustered into semantic regions. Each trajectory j samples a distribution G_j over semantic regions from Dirichlet process, $G_j \sim DP(\alpha, G_0)$. $G_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\phi_k}$ share the same set of semantic region models $\{\phi_k\}$ as G_0. However, they have different mixture weights $\{\pi_{jk}\}$ over semantic regions. For each observation i on trajectory j, a semantic region model θ_{ji}, which is one of the ϕ_k's, is sampled from G_j. The value of the observation w_{ji} is sampled from the semantic region, $w_{ji} \sim Discrete(\theta_{ji})$. The observations sampled from the same semantic region model ϕ_k are grouped into the same cluster k. The concentration parameters are sampled from some gamma priors, $\gamma \sim Gamma(a_1, b_1)$, $\alpha \sim Gamma(a_2, b_2)$, such that α and γ do not have to be manually tuned. As the hierarchical levels increase, hierarchical Bayesian models become more insensitive to the choice of hyperparameters [67]. That is the reason of introducing hyperparameters a_1, b_1, a_2 and a_2 on the top of α and γ.

In HDP, all the trajectories share semantic regions and the number of semantic regions, i.e. the number of non-zero elements of $\{\pi_{0k}\}$ is automatically learnt from data. HDP has high data likelihood if the distribution $\{\pi_{jk}\}$ of each trajectory concentrates on a few semantic regions instead of being uniform over all the semantic regions. Therefore, the quantized locations and moving directions which often co-occur on the same trajectories tend to be grouped into the same semantic regions in order to maximize the data likelihood. In Figure 3, $\{w_{ji}\}$ are observed. a_1, b_1, a_2, b_2 and H are hyperparameters to be set. All the others are hidden variables to be inferred.

6 Dual-HDP

Unfortunately, HDP does not cluster trajectories. We used a Dual-HDP model proposed in [2] to co-cluster both observations and trajectories. A trajectory is modeled as a distribution over semantic regions. Thus trajectories with similar distributions over semantic regions can be grouped into one cluster. Such a cluster corresponds to a path (a path may have more than one semantic regions) commonly taken by objects. There are two hierarchical Dirichlet processes modeling both semantic regions and paths. The graphical model of Dual-HDP and an exemplary illustrations are shown in Figure 4.

In Dual-HDP, each trajectory j is from one of the trajectory clusters. All the trajectory in cluster c pass through the same path and have the same prior distribution $\widetilde{G}_c$ over semantic regions. $\widetilde{G}_c = \sum_{k=1}^{\infty} \widetilde{\pi}_{ck}\delta_{\widetilde{\phi}_{ck}}$, the model of path c, is an infinite mixture of semantic regions. Since the number of trajectory clusters is unknown in advance, we model the clusters of trajectories as an infinite mixture,

$$Q = \sum_{c=1}^{\infty} \epsilon_c \delta_{\widetilde{G}_c}. \tag{4}$$

When a DP was first developed by Ferguson [65], the components (such as ϕ_k in Eq (1)) could only be scalars or vectors. MacEachern [68] generalized this to *Dependent Dirichlet Process* (DDP). In DDP, components could be stochastic processes. In our model, the parameters $\{(\widetilde{\pi}_{ck}, \widetilde{\phi}_{ck})\}_{k=1}^{\infty}$ of $\widetilde{G}_c$ can be treated as a stochastic process with index k. As shown in Figure 4, Q is generated from $DDP(\mu, \rho, G_0)$. In Eq (4), $\epsilon_c = \epsilon'_c \prod_{l=1}^{c-1}(1-\epsilon'_l)$, $\epsilon'_c \sim Beta(1, \mu)$, $\widetilde{G}_c \sim DP(\rho, G_0)$.

As explained in Section 5, $G_0 \sim DP(\gamma, H)$ is the prior distribution over the whole data set. $\{\widetilde{G}_c\}_{c=1}^{\infty}$, models of the paths, all share the same set of semantic regions as in G_0. i.e. $\widetilde{\phi}_{ck} = \phi_k$. However they have different mixture weights $\{\widetilde{\pi}_{ck}\}$ over semantic regions. Each trajectory j samples a pathway model $\widetilde{G}_{c_j}$ from Q as its prior. Different trajectories may choose the same pathway model $\widetilde{G}_c$, and thus they form one cluster c. Then trajectory j generates its own probability measure G_j from $G_j \sim DP(\alpha, \widetilde{G}_{c_j})$ where the base measure is provided by cluster c_j instead of the corpus prior G_0 (as HDP did). The following generative procedure is the same as HDP. Observation i in trajectory j samples a semantic region θ_{ji} from G_j and samples its observation value w_{ji} from $Discrete(\theta_{ji})$. The concentration parameters are also sampled from gamma priors.

Collapsed Gibbs sampling is used to do inference in three steps.

1. Given the path assignment $\{c_j\}$ of trajectories, sample the semantic region assignment $\{z_{ji}\}$ ($z_{ji} = k$ indicates $\theta_{ji} = \phi_k$) of observations, and semantic region mixtures $\{\pi_{0k}\}$ and $\{\widetilde{\pi}_{ck}\}$. Given $\{c_j\}$, Dual-HDP is simplified as HDP, and thus the sampling scheme proposed by Teh et al. [26] can be used. They showed that $\{\phi_k\}$ and $\{\pi_{jk}\}$ can be integrated out without being sampled.
2. Given $\{z_{ji}\}$, $\{\pi_{0k}\}$ and $\{\widetilde{\pi}_{ck}\}$, sample the path assignment c_j of trajectories. c_j can be assigned to one of the existing paths or to a new path. We use the Chinese restaurant franchise for sampling. See details in [2].
3. Given other variables, sample the concentration parameters using the sampling scheme proposed in [26].

In order to detect abnormal trajectories, we need to compute the likelihood of trajectory j given other trajectories, $p(\mathbf{w}_j|\mathbf{w}^{-j})$, where $\mathbf{w}_j = \{w_{ji}\}_{i=1}^{N_j}$ is the set

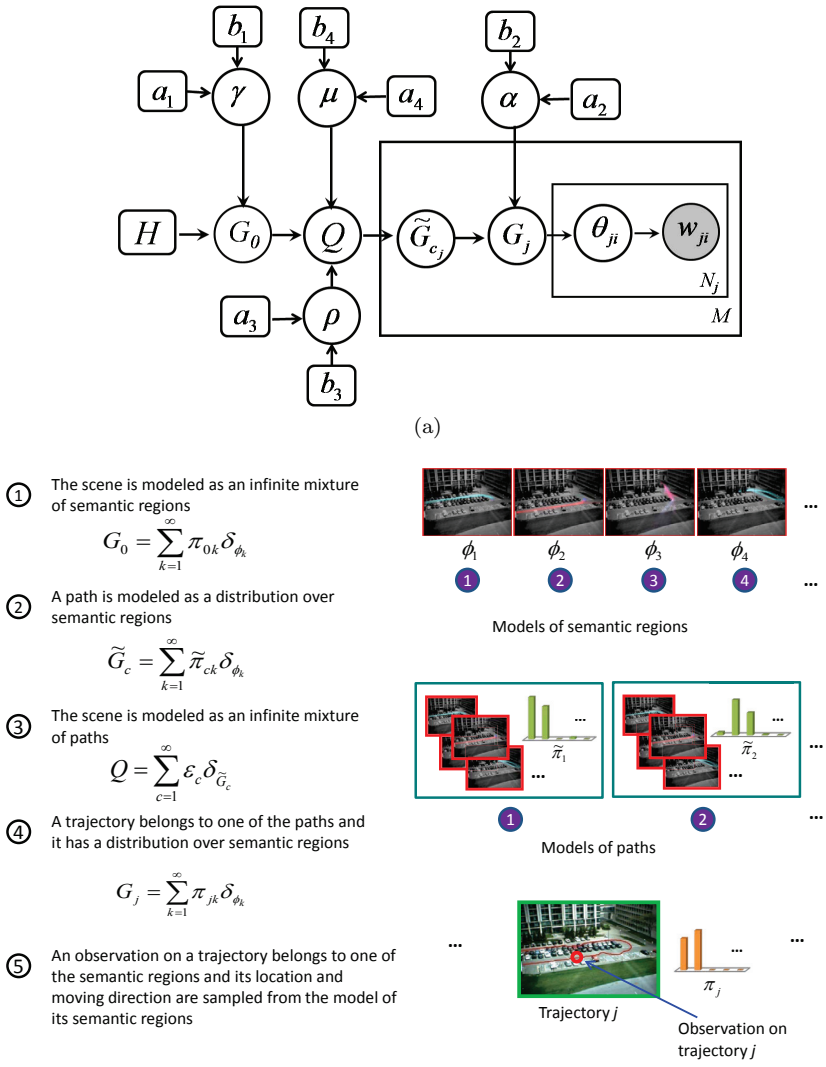


Fig. 4 Dual-HDP. (a) Graphical model. (b) Illustration of modeling semantic regions, paths, trajectories and observations. Semantic regions 1 and 2 form a path where objects enter the parking lot, make a u-turn and leave. Semantic regions 2 and 3 form another path where objects enter the parking lot, move upward and leave in a different direction. Semantic region 2 is the overlap region of the two paths. Path 1 has large distribution over the first two semantic regions. Trajectory j belongs to path 1. It samples its distribution G_j over semantic regions from the prior given by its path model. Observation i on trajectory j belongs to semantic region 1 and its value is sampled from the model of its semantic region.

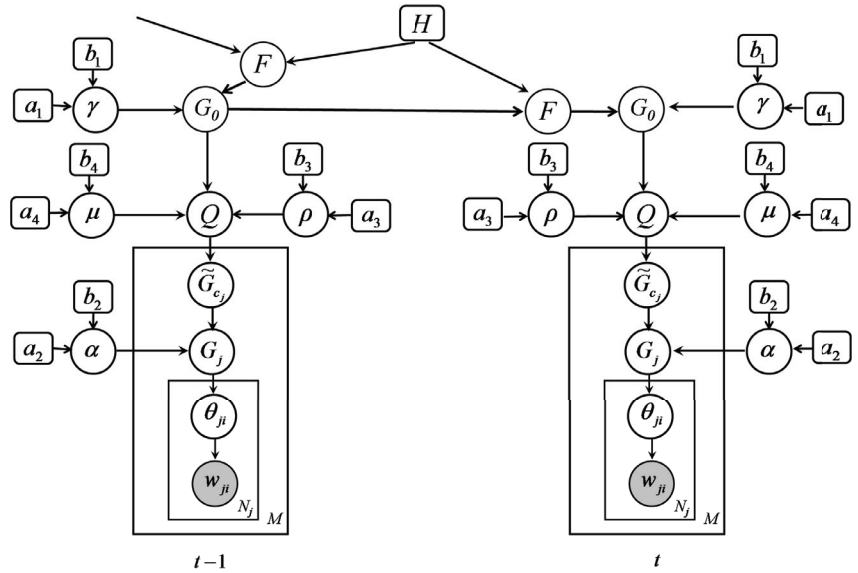
observations in trajectory j and $\mathbf{w}^{-j}$ represents the remaining trajectories excluding j. It can be approximated using the samples obtained during collapsed Gibbs sampling and a variational method. Computing $p(\mathbf{w}_j|\mathbf{w}^{-j})$ only needs some sufficient statistics of $\mathbf{w}^{-j}$ without comparing $\mathbf{w}_j$ with all other other trajectories. So the computation is efficient. See details in [2]. Besides the offline mode, activity analysis and abnormality detection can also be done in an online mode and run in realtime. Once Gibbs sampling on the training set converges, $\{\phi_k\}$, $\{\pi_{0k}\}$ and $\{\tilde{\pi}_{ck}\}$ can be esimated from the samples. A new trajectory outside the training set can be detected as abnormality by computing the likelihood $p(\mathbf{w}_j|\{\phi_k\},\{\pi_{0k}\},\{\tilde{\pi}_{ck}\})$. It also can be classified as one of the pre-learned clusters by computing the posteriors $p(c_j|\mathbf{w}_j,\{\phi_k\},\{\pi_{0k}\},\{\tilde{\pi}_{ck}\})$.

7 Dynamic Dual-HDP

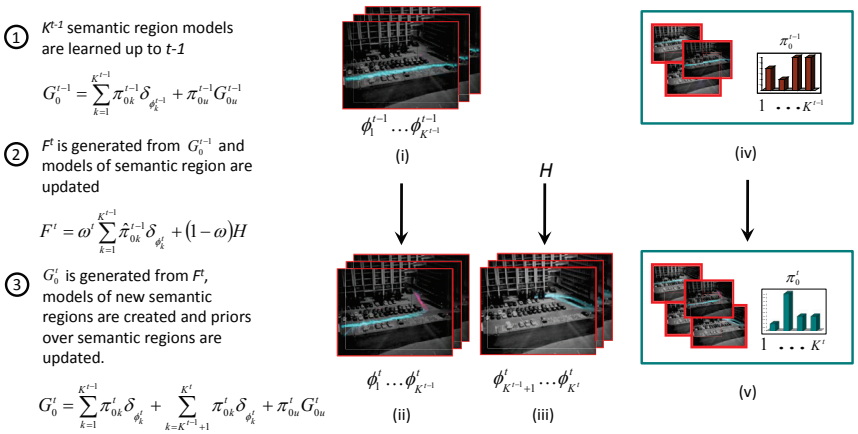
Under Dual-HDP, when the models of activities and semantic regions are learnt and fixed, classifying observed new trajectories into existing activity categories and detecting abnormal trajectories can be done in an online mode. However, there are still some reasons to extend the Dual-HDP model to a dynamic Dual-HDP model. First, people have interest in the dynamic change of models of activities over time. For example, exploring when a new mode of activity appears, when an old mode of activity disappears, and when a particular kind of activity becomes more dominant than other activities in the scene is of interest in surveillance applications. Abnormality detection may also change over time. An activity may be detected as an abnormality when it first appears in the scene. However, when more and more instances occur, it becomes typical. Similarly, a typical activity at an earlier time may become abnormal when it rarely happens later. Second, when a surveillance system monitors an area over months or even years, it is difficult to load all the huge amount of data once into memory and process it. Dynamic Dual-HDP learns the models of activities incrementally over time and does not have to keep old data.

In order to learn the models of activities dynamically, one option is to divide the entire data set into subsets according to the temporal order and learn the activity models of each subset independently using Dual-HDP. This method has two problems. First, the activity models learnt in different subsets are not aligned. Without manually permuting the activity models properly, people cannot observe how these models change over time. Second, since the models different subsets do not share data, if there is not enough data in a subset, the activity models cannot be well learnt from it. Blei et al. [29] proposed a model which allowed the topics to be dynamically updated. However, it assumed that the number of topics was fixed. Allowing the addition of new emerging activity models over time is of considerable interest in surveillance applications.

The graphical model of the dynamic Dual-HDP is shown in Figure 5. The data is divided into subsets according to the temporal order (e.g. a subset includes trajectories happening within one hour). Both models $\{\phi_k^t\}$ of semantic regions and the prior distribution $\{\pi_{0k}^t\}$ over semantic regions are dynamically updated. G_0^{t-1} is the mixture of semantic regions learnt up to time $t-1$, and is used as prior to predict G_0^t, which is the mixture of semantic regions learnt at the next time interval t. Assume that K^{t-1} semantic regions have been learned from the data up to $t-1$. Then G_0^{t-1} can be



(a)



(b)

Fig. 5 Dynamic Dual-HDP. (a) Graphical model. (b) Illustration of the dynamic update of the Dual-HDP model. (i) K^{t-1} models of semantic regions learned up to time $t-1$. (ii) The K^{t-1} old models of semantic regions are updated given new data observed at time t. (iii) The models of $K^t - K^{t-1}$ new semantic regions are created at time t. (iv) Prior distribution over semantic regions learned up to time t. (v) Prior distribution over semantic regions updated given new data observed at time t.

represented as

$$G_0^{t-1} = \sum_{k=1}^{K^{t-1}} \pi_{0k}^{t-1} \delta_{\phi_k^{t-1}} + \pi_{0u}^{t-1} G_{0u}^{t-1} \tag{5}$$

where the first K^{t-1} semantic regions have been assigned to the data observed up to time $t-1$, and $\pi_{0u}^{t-1} G_{0u}^{t-1} = \sum_{k=K^{t-1}+1}^{\infty} \pi_{0k}^{t-1} \phi_k^{t-1}$ is the remaining part of the infinite

mixture of semantic regions, none of which is assigned to any data observed up to time $t-1$ [26]. Both $\{\pi_{0k}^{t-1}\}_{k=1}^{K^{t-1}}$ and $\{\phi_k^{t-1}\}_{k=1}^{K^{t-1}}$ can be sampled from the data observed up to $t-1$. We assume that they are learnt and fixed when predicting G_0^t.

In order to use the mixture of semantic regions learnt up to $t-1$ as the prior of G_0^t, we first normalize $\{\pi_{0k}^{t-1}\}_{k=1}^{K^{t-1}}$ to $\{\hat{\pi}_{0k}^{t-1}\}_{k=1}^{K^{t-1}}$

$$\hat{\pi}_{0k}^{t-1} = \frac{\pi_{0k}^{t-1}}{\sum_{k'=1}^{K^{t-1}} \pi_{0k'}^{t-1}}.$$

Then a base probability measure F^t is generated from G_0^{t-1},

$$F^t = \omega^t \sum_{k=1}^{K^{t-1}} \hat{\pi}_{0k}^{t-1} \delta_{\phi_k^t} + (1-\omega^t)H, \tag{6}$$

where

$$\phi_k^t \sim Dir(\xi_k^t \cdot \phi_k^{t-1} + H), \tag{7}$$

H is the same Dirichlet distribution as in Section 6 without changing over time, ω^t is a scalar between 0 and 1, and ξ_k^t is a positive scalar. During the generation of F^t, the models $\{\phi_k^t\}$ of semantic regions are updated. $\{\phi_k\}$ are sampled from priors given by $\{\phi_k^{t-1}\}$ and are updated given the data observed at t, instead of being fixed over time as in [55] and [56]. ξ^t controls the temporal smoothness of the models of semantic regions. The larger is ξ^t, the more similar is ϕ_k^t to ϕ_k^{t-1}.

G_0^t is sampled from a Dirichlet process choosing F^t as the base measure and γ^t as the concentration parameter,

$$G_0^t \sim DP(\gamma^t, F^t). \tag{8}$$

By introducing this step, the models of new semantic regions not seen before can be created given new data observed at t and the prior distribution $\{\pi_{0k}^t\}$ over semantic regions is also updated. From Eq (6) and (8), we observe that the random measure G_0^t at time t includes the K^{t-1} semantic regions generated before t and also some new semantic regions. Some existing works such as [55] directly added dependency between G_0^{t-1} and G_0^t without introducing F_t, because they only needed to model the dynamic variation of mixture weights π_{0k}^t but not the topic models ϕ_k^t. Our dynamic Dual-HDP models the dynamic variations of both π_{0k}^t and ϕ_k^t by introducing F^t. When F^t is generated from G_0^{t-1}, ϕ_k^t is generated from ϕ_k^{t-1} and thus can be dynamic updated. In the following, we explain the inference by collapsed Gibbs sampling.

Suppose that at a sampling step there are K^t topics assigned to the data up to t (K^t changes during collapsed Gibbs sampling on the subset of t). Then an explicit construction for G_0^t is given as,

$$G_0^t = \sum_{k=1}^{K^{t-1}} \pi_{0k}^t \delta_{\phi_k^t} + \sum_{k=K^{t-1}+1}^{K^t} \pi_{0k}^t \delta_{\phi_k^t} + \pi_{0u}^t G_{0u}^t. \tag{9}$$

$\{\phi_k^t\}_{k=1}^{K^{t-1}}$ are the models of semantic regions existing before t. They are updated using the data observed at t. They are the same variables as in Eq (6). $\{\phi_k\}_{k=K^{t-1}+1}^{K^t}$ are the models of new semantic regions assigned to the data observed at t. $\pi_{0u}^t G_{0u}^t = \sum_{k=K^t+1}^{\infty} \pi_{0k}^t \phi_k^t$ is the remaining part of the infinite mixture of semantic regions, none

Algorithm 1 Inference under the dynamic Dual-HDP

Input trajectories collected from T time slices, $\{w_{ji}^t\}, t = 1, \ldots, T$.
Output models of activities and semantic regions and cluster labels of trajectories at different times.
Initialization $K^0 = 0$, $n_k^0 = 0$, $s^0 = 0$.
for $t = 1$ to T **do**
  repeat
    given other variables, sample the semantic region assignment $\{z_{ji}\}$ of the observations $\{w_{ji}^t\}$ obtained at time t, and sample the semantic region mixture weights $\{\tilde{\pi}_{ck}\}$ of trajectory clusters using the Chinese restaurant franchise sampling scheme proposed in [26].
    given other variables, sample the cluster labels c_j of trajectories observed at time t, and sample the mixture weights $\{\epsilon_c\}$ of trajectory clusters inEq (4) using the Chinese restaurant franchise sampling proposed in [2].
    given other variables, sample semantic region models $\{\phi_k^t\}$ and mixture weights $\{\pi_{0k}^t\}$ from the models $\{\phi_k^{t-1}\}$ and $\{\pi_{0k}^{t-1}\}$ learnt at time $t-1$ and the data observed at time t using Eq (23) and (24).
  until converge
  update n_k^t and s^t using Eq (21) and (22).
end for

of which is signed to any data up to time t. From Eq (6) and (8), $G_{0u}^t \sim DP(\gamma^t(1-\omega^t), H)$. $\boldsymbol{\pi}_0^t = (\pi_{01}^t, \ldots, \pi_{0K^t}^t, \pi_{0u}^t)$ and $\{\phi_k^t\}_{k=1}^{K^t}$ are the variables to be sampled. Given $\boldsymbol{\pi}_0^t$ and $\{\phi_k^t\}$, the sampling of other variables is the same as Dual-HDP. We use the Chinese Restaurant Franchise sampling scheme to sample $\boldsymbol{\pi}_0^t$ and $\{\phi_k^t\}$ given other variables. Details can in found in the appendix. The inference under dynamic Dual-HDP is summarized in Algorithm 1. The choice of parameters ω^t, ξ^t and γ^t controls how much the models learned from data observed before t affect the models to be learned at t. The two extreme cases are that the pre-learned models have on effect on clustering new data ($\omega^t = 0$, $\xi^t = 0$) and that the models learned at t are exactly the same as those learned at $t-1$ ($\omega^t = 1$, $\xi^t = \infty$, $\gamma = \infty$). However, we do not have to tune these parameters individually. As shown in the appendix, these parameters can be replaced by a single decreasing rate parameter r ($0 \leq r \leq 1$) during sampling. As data becomes older, its influence on the current model becomes weaker. r controls how fast the influence decreases.

In our problem, dynamic Dual-HDP is applied to online learning of activity models and online abnormality detection, where we assume that data in the future is unknown. Thus in Eq (15) and (16), π_0^t and ϕ_k^t are sampled from the posteriors given $\hat{\pi}_0^{t-1}$ and ϕ_k^{t-1} without knowing $\hat{\pi}_0^{t+1}$ and ϕ_k^{t+1}. If we assume that data both in the past and in the future is known, the posteriors become more complicated than Eq (15) and (16), and the collapsed Gibbs sampling inference may require keeping all data collected from the whole period in the memory. In the current sampling algorithm, we only need to keep the data observed at the current time slice for inference. All the old data can be replaced from memory, as its information has been included in the activity models $\{\pi_{0k}^{t-1}\}$ and ϕ_k^{t-1} and some sufficient statistics described in the appendix. When $t = 0$ or there is no data observed before t, the posteriors of $\boldsymbol{\pi}_0^t$ and ϕ_k^t are the same as in Dual-HDP.

8 Experimental Results

8.1 Trajectory Analysis without Dynamic Modeling

Our nonparametric hierarchical Bayesian models are evaluated on radar tracks collected from a maritime port, visual tracks collected from a parking lot, and a simulated data set. The results of the Dual-HDP model without dynamic modeling will be first presented. The results of the Dynamic Dual-HDP model will be reported in 8.2. Under the Dual-HDP model shown in Figure 4, hyperparameters a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4 and $H = (u_1, \ldots, u_W)$ are hyperparameters to be set. We choose $u_1 = \ldots = u_W = 0.001$ as a small number to avoid singularity during inference. The gamma prior parameters $a_1 \ldots a_4$ and $b_1 \ldots b_4$ are all set equal to one, the same as used in [26,2]. As the increase of hierarchical levels, hierarchical Bayesian models become less sensitive to hyperparameters [67]. This is one of the major advantages of employing hierarchical Bayesian models. Therefore we do not need to spend a lot of effort on tuning parameters. By adding gamma priors over concentration parameters α, γ, μ and ρ, these concentration parameters do not need to be manually specified. Our model is more robust to the choice of $a_1 \ldots a_4$ and $b_1 \ldots b_4$ than directly tuning α, γ, μ and ρ.

8.1.1 Results on Radar Tracks

In this section, experiments are done on a relatively small data set which has 577 radar tracks collected from a maritime port data set. They were acquired by multiple collaborating radars along the shore and recorded the locations of ships on the sea. Many existing approaches were evaluated on data sets with similar sizes as this one. In order to build the codebook, the spatial region is quantized into 75×42 cells and the moving directions are quantized into four. The choice of quantization parameters depends on the size of the data set and application requirements. A larger codebook can describe the models of scene structures at a higher resolution but it also requires a larger data set for the models of clusters to well learned. 23 semantic regions are discovered by our model. In Figure 6, we display the distributions of semantic regions (sorted by the number of observations assigned to semantic regions) over space and moving directions. As shown in Figure 6, the 1st, 4th, 6th, 8th and 15th semantic regions are five side by side shipping fairways, where ships move in two opposite directions. For comparison, we segment the five fairways using a threshold on the density, and overlay them in Figure 6 (c) in different colors, green (1st), red (4th), black (6th), yellow (8th), and blue (15th). Since they are so close in space, they may not be separated using some spatial distance based trajectory clustering approaches, such as Euclidean distance [37], as an example shown in Figure 7 [3]. In Figure 6 (d), we compare the 7th, 11th, and 13th semantic regions also by overlaying the segmented regions in red, green, and black colors. This explains the fact that ships first move along the 7th semantic region and then diverge along the 11th and 13th semantic regions [4].

[3] Here, we emphasize different behaviors of clustering algorithms. Our clustering algorithm does not depend on spatial distance between trajectories, but on the connectivity of spatial locations. Even if two pathways are very close to each other but there are no trajectories crossing between them, they will not be merged into one cluster. Whether clustering algorithms are successful or not on this data set also depends on the requirements of applications. In some applications, it may be expected to cluster trajectories of neighboring lanes together.

[4] This is interpreted by a maritime surveillance expert who is familiar with activities in this area. 11 and 13 are not the same semantic region with large variance.

Our approach groups trajectories into 16 clusters [5]. In Figure 7, we plot the eight largest clusters and some smaller clusters. Clusters 1, 4, 6 and 7 are close in space and their trajectories move along different shipping fairways. Trajectories in clusters 1 and 6 move along in one direction. Trajectories in clusters 4 and 7 move along in another direction. Clusters 3 and 5 occupy the same region, but ships in the two clusters moves in opposite directions. Clusters 2 and 5 partially overlap in space. As shown in Figure 6(d), ships first move along the same way and then diverge in different directions. Clusters 2 and 5 share the same semantic region. Only modeling semantic regions using HDP cannot separate these two clusters. According to a map [6] of shipping fairways in this area, most clusters, such as cluster 1, 2 and 3, have proper semantic meanings. Although some clusters, such as cluster 14 and 15, are not on the map, they do reveal some interesting activity patterns after interpreted by experts. For comparison, in the last two sub-figures of Figure 7 we also show two clusters of the result using Euclidean distance and spectral clustering [37] and setting the number of clusters as 16. In this approach a similarity matrix is computed by comparing the distance between each pair of trajectories. Then spectral clustering is used to compute an embedded space. Trajectories are projected to the embedded space and clustered by k-means. Some fine structures of shipping fairways are not separated using a spatial distance based clustering method. One of the advantages of our approach is that it learns the number of clusters from data. When spatial distance based clustering methods are evaluated on this data set, choosing an improper cluster number, say 8 or 25, causes the clustering performance to significantly deteriorate. In those cases, trajectories of different fairways are grouped into one cluster, or trajectories on the same fairway are split into many clusters.

In Figure 8, we display the top 20 abnormal trajectories based on their normalized log-likelihoods $log(p(\mathbf{w}^j|\mathbf{w}^{-j}))/N_j$. There are two possible reasons for the abnormality. (1) The trajectory does not fit any major semantic regions. Many examples can be found in Figure 8. (2) The trajectory fits more than one semantic region, but the combination of the semantic regions is uncommon. Remind that a semantic region is a part of a pathway. This means that although these semantic regions are commonly seen, the whole pathway formed by them is abnormal. The red trajectory in Figure 8 (a), and the red and green trajectories in Figure 8 (b) are such examples [7]. In Figure 7 it is observed that some clusters include abnormal trajectories. Under Dual-HDP, a cluster is created if there are a significant amount of trajectories with similar motion patterns. Based on our experience through experiments, if a single abnormal trajectory is dissimilar with any other trajectories, it is likely to be assigned to one of the closest existing clusters instead of creating a new cluster only containing itself. There is another possibility that many abnormal trajectories form one cluster like background noise in some sense. This can be observed in the parking lot data set (cluster 22 in Figure 10).

8.1.2 Results on tracks from a parking lot

There are $N = 40,453$ trajectories in the parking lot data set collected over one week and they are plotted in Figure 1. Because of the large number of samples, if some

[5] A cluster corresponds to a type of activities, which are characterized by the pathway where objects pass through with similar motion patterns. A pathway covers a few semantic regions.

[6] The map cannot be shown for the security reason.

[7] Some abnormal tracks may be caused by tracking errors. However, it is hard to tell since we only have radar tracks without image and video data.

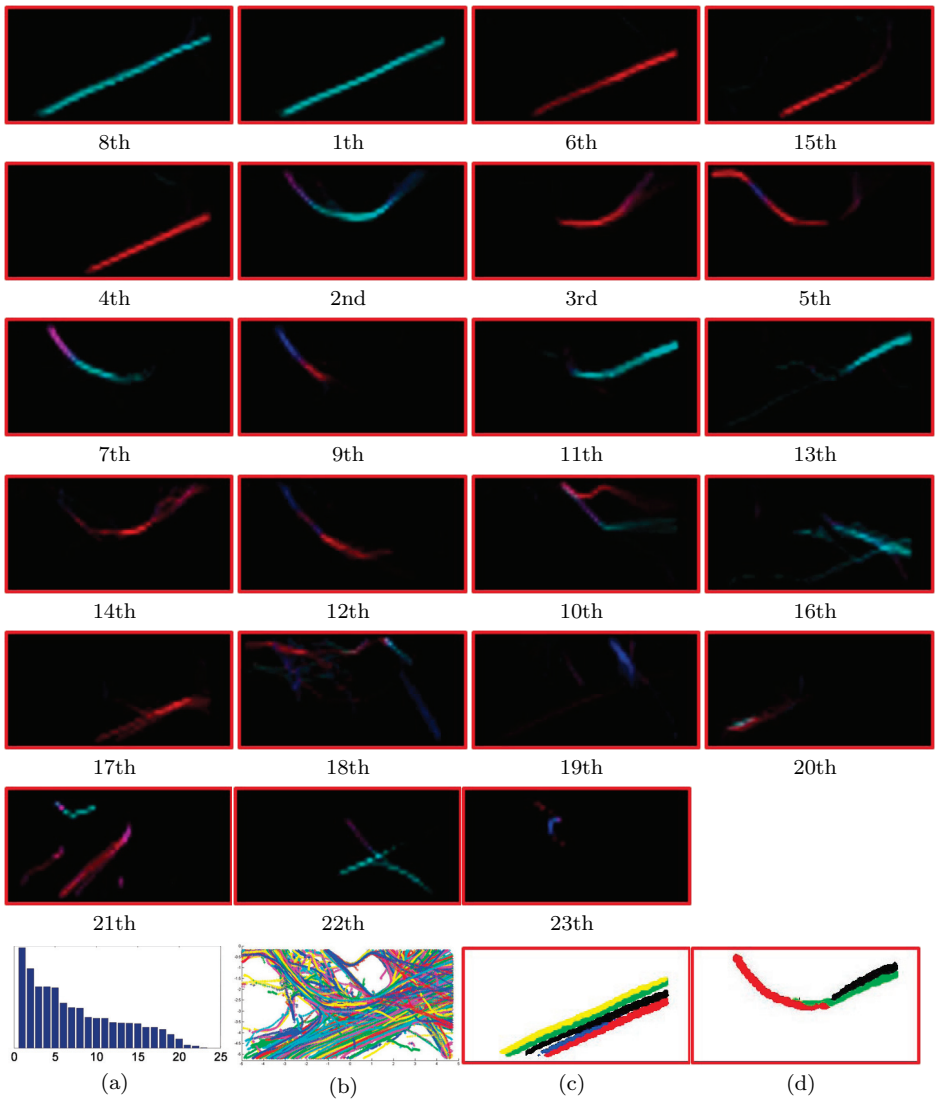


Fig. 6 Semantic regions at a maritime port learnt from the radar tracks. Distributions of semantic regions over space and moving directions are shown (for easier comparison, they are not shown in order). Colors represent different moving directions: $\rightarrow$ (red), $\leftarrow$ (cyan), $\uparrow$ (magenta), and $\downarrow$ (blue). (a) Histogram of observations assigned to different semantic regions. (b) All of the radar tracks. (c) Compare the 1st (green), 4th (red), 6th (blue), 8th (yellow), and 15th semantic regions. (d) Compare the 7th, 11th, and 13th semantic regions (see details in text).

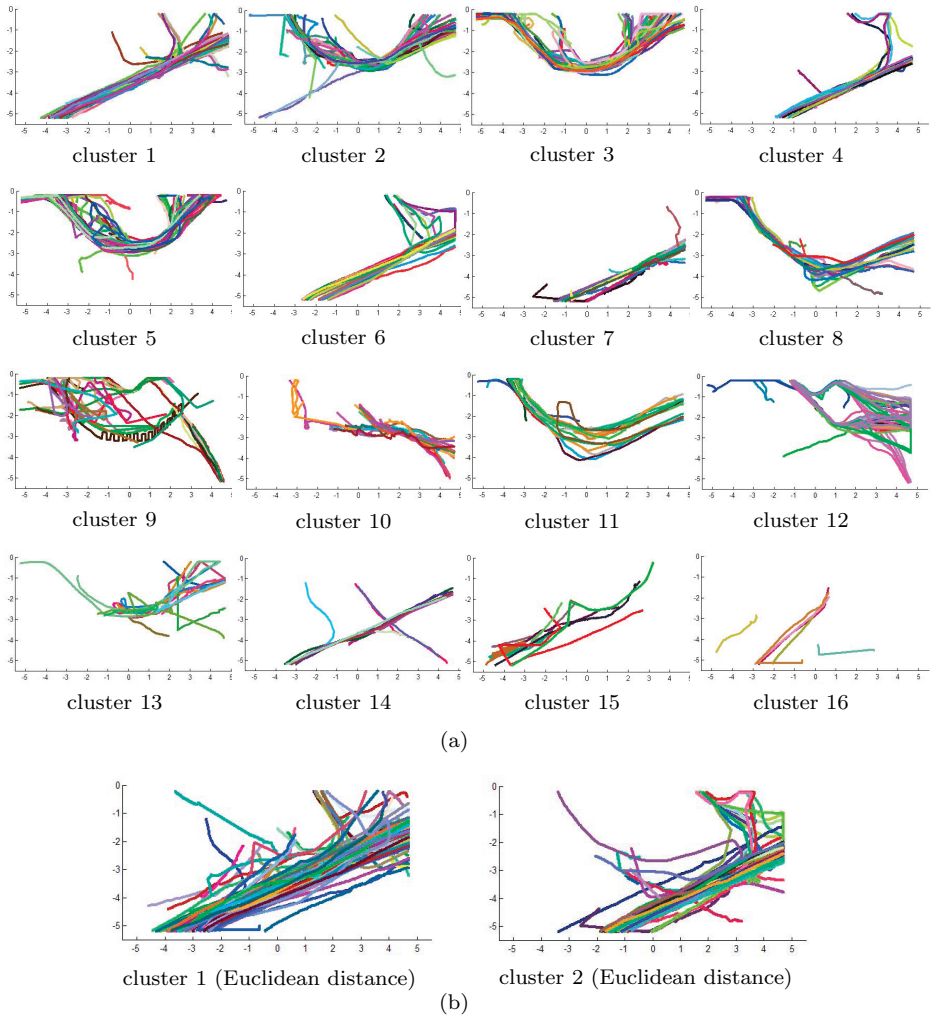


Fig. 7 Clusters of trajectories. Random colors are used to distinguish individual trajectories. For comparison the last two sub-figures show some trajectory clusters of the result using Euclidean distance and spectral clustering [37], which also chooses the cluster number of 16.

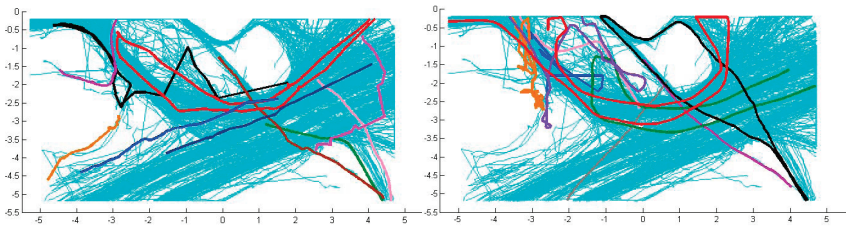


Fig. 8 Top 20 abnormal trajectories are plotted in different colors. Other trajectories are plotted in cyan color.

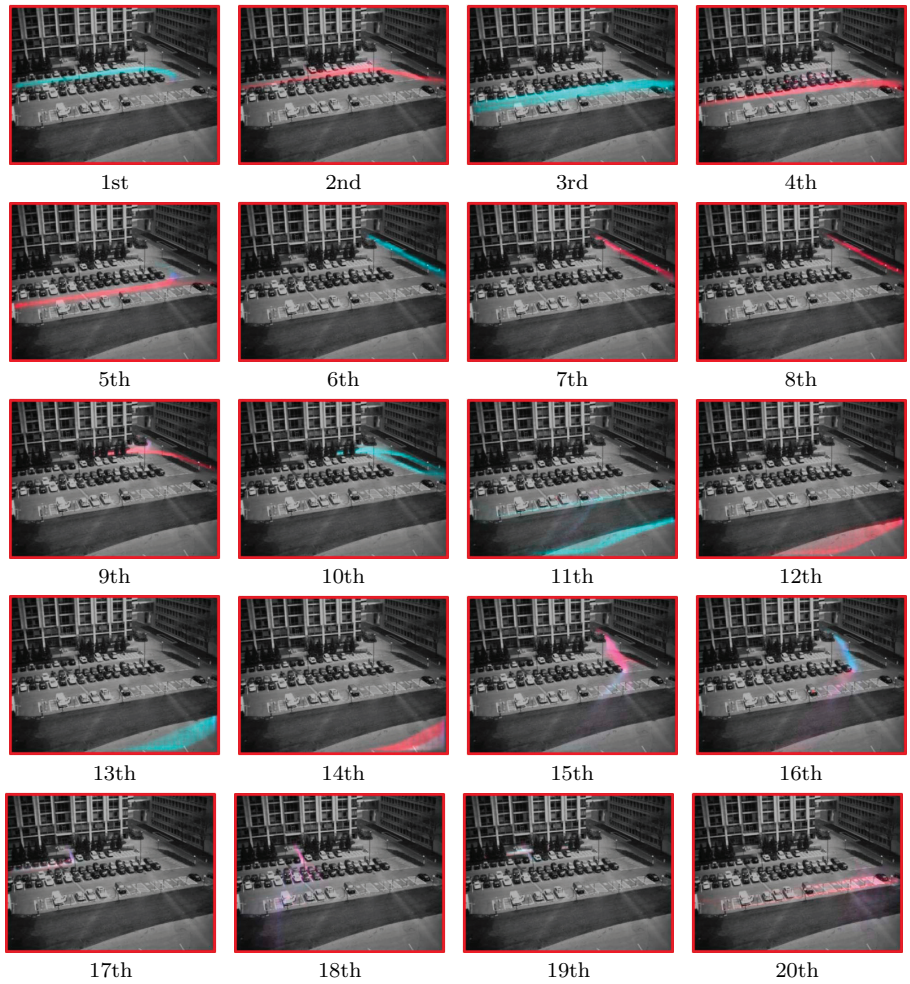


Fig. 9 Some semantic regions learnt from the parking lot data set. The meaning of colors is the same as Figure 6.

similarity based clustering methods require the similarity/distance matrix, it needs both large amounts of space (6GB) to store the $40,453 \times 40,453$ similarity matrix and high computational cost to compute the similarities of around $800,000,000$ pairs of trajectories. If spectral clustering is used, it is quite challenging to compute the eigenvectors of such a huge matrix without doing approximation by subsampling [43]. It is difficult for many existing approaches to work on this large data set. The space complexity of our nonparametric Bayesian approach is $O(N)$ instead of $O(N^2)$. The time complexity of each collapsed Gibbs sampling iteration is $O(N)$. It is difficult to provide theoretical analysis on the convergence of collapsed Gibbs sampling. However, we can gather empirical observations by plotting the likelihoods of data sets over Gibbs sampling iterations. On the smaller radar data set, the likelihood curve converges after $1,000$ iterations. This takes around 1.5 minutes running on a computer with 3GHz

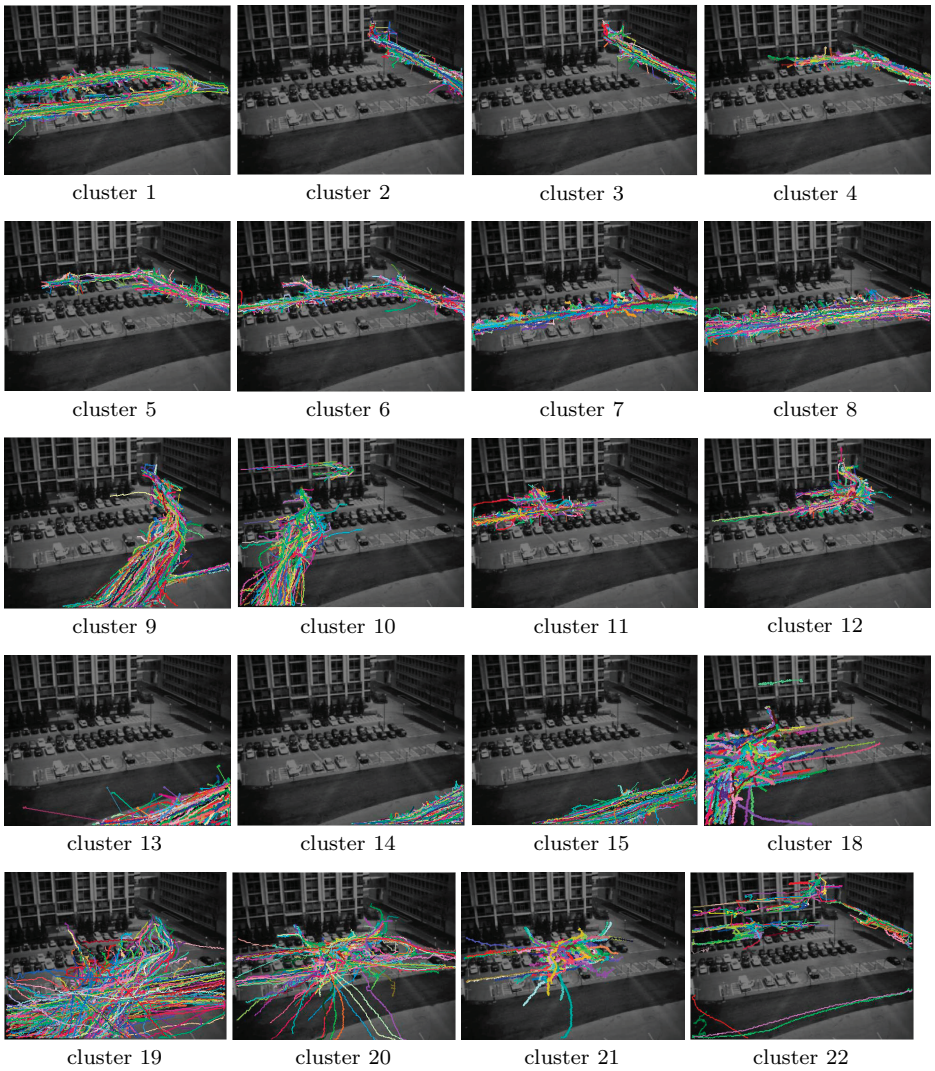


Fig. 10 Some clusters of trajectories from the parking lot data set.

CPU. On the parking lot data set, which is 70 times larger than the radar data set in the number of trajectories, the likelihood curve converges after $6,000$ iterations. It takes around 6 hours. According to our experiments, the time complexity of our approach is much smaller that $O(N^2)$.

The scene is in size of 360×480. In order to build the codebook, the spatial region is quantized into 72×96 cells and moving directions are quantized into four. 30 semantic regions and 22 clusters of trajectories are learnt from this data set. Some of them are shown in Figures 9 and 10. The first and fifth semantic regions explain vehicles entering and exiting the parking lot. Most other semantic regions are related to pedestrian activities. Because of opposite moving directions, some region splits into two semantic

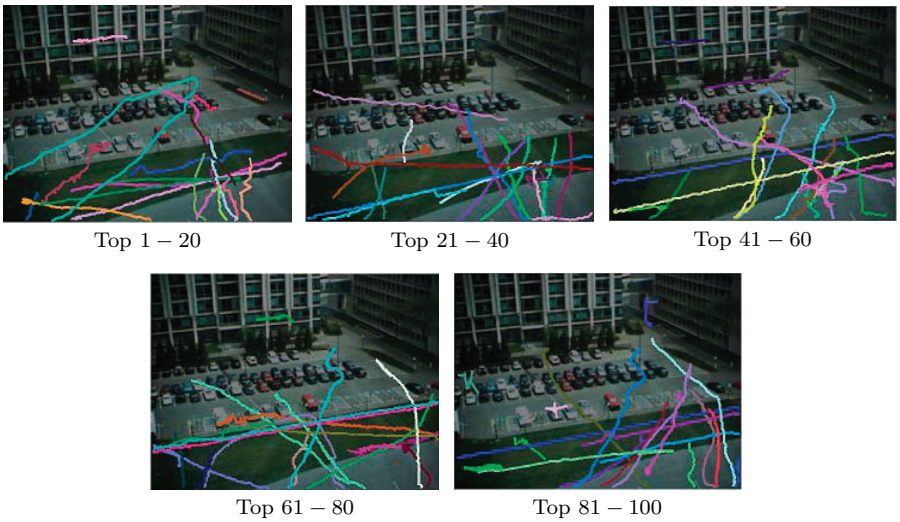


Fig. 11 Top 100 abnormal trajectories in the parking lot data set.

cluster number		5	10	15	20	22	25	30	35
Euclidean	$r_{complete}$	0.92	0.90	0.86	0.80	0.78	0.65	0.47	0.27
	$r_{correct}$	0.30	0.43	0.51	0.65	0.71	0.76	0.79	0.85
Modified Hausdorff	$r_{complete}$	0.94	0.93	0.88	0.82	0.80	0.75	0.53	0.41
	$r_{correct}$	0.40	0.46	0.55	0.78	0.79	0.83	0.89	0.91
LCSS	$r_{complete}$	0.95	0.93	0.90	0.86	0.84	0.78	0.56	0.38
	$r_{correct}$	0.37	0.51	0.59	0.75	0.78	0.85	0.93	0.94
Dual-HDP	$r_{complete}$	-	-	-	-	0.91	-	-	-
	$r_{correct}$	-	-	-	-	0.86	-	-	-

Table 2 Accuracies of completeness and correctness of different clustering methods: Euclidean distance [37], modified Hausdorff distance [19], LCSS [39] and Dual-HDP, on the parking lot data set.

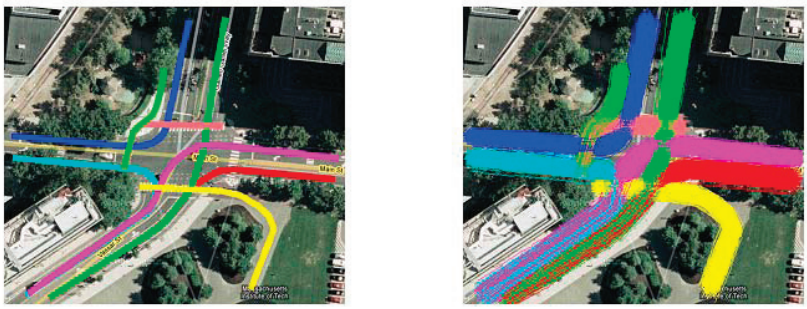
regions, such as semantic regions 3 and 4, 6 and 7, 13 and 14. Similarly objects on trajectories (see Figure 10) in clusters 2 and 3, 7 and 8 are moving in opposite directions. Cluster 19 is not as clean as some other clusters. It mainly includes horizontal trajectories along the grass field, trajectories crossing the grass field and some outlier trajectories. Many outlier trajectories are in small clusters, such as clusters 20, 21 and 22. The top 100 abnormal trajectories are shown in Figure 11. Some horizontal trajectories on the grass field are detected as abnormalities. They were caused by a worker shearing the grass, which happened only once.

We also quantitatively compare our method with three distance-based trajectory clustering methods which use Euclidean distance [37], modified Hausdorff distance [19] and long common subsequence (LCSS) [39] to compute distance between two trajectories. These three were among the top trajectory clustering methods according to a recent comparison study [42]. In our implementations, the modified Hausdorff distance and LCSS compare both spatial distance and velocity difference of observations on the trajectories. The velocity helps to resolve the ambiguity caused by spatial overlap between some clusters. The three distance-based methods adopt spectral clustering, and the Nyström method [43] is used for approximation in order to handle the large

scale data set. The results are shown in Table 2. In this experiment, the accuracies of correctness and completeness are used to evaluate cluster performance and it was also in used in other clustering works [70]. Correctness means that trajectories of different activity categories are not clustered together. Completeness means that trajectories of the same activity categories are clustered together. To measure correctness, we asked subjects who were familiar with this scene to label $1,000$ pairs of trajectories and each pair of trajectories are from different activity categories. $r_{correct}$ is calculated as the accuracy that they are also in different clusters based on the results of clustering algorithms. To measure completeness, $1,000$ pairs of trajectories by are labeled by subjects and each pair of trajectories are from the scame cativity category. $r_{complete}$ is calculated as the accuracy that they are also in the same clusters based on the results of clustering algorithms. The subjects only need to tell whether a pair of trajectories are from the same cluster or not and do not have to estimate the number of clusters during labeling (which is difficult when the data set is large and activities are complicated). In the meanwhile, correctness and completeness can be used to evaluate whether a proper cluster number has been chosen. Grouping all the trajectories in the same cluster results in 100% completeness and 0% correctness. Putting every trajectory into a singleton cluster results in 100% correctness and 0% completeness. In Table 2, the three distance-based methods are allowed to manually choose different numbers of clusters for spectral clustering, while our Dual-HDP automatically learns the cluster number (22) from data. It is observed that the correctness and completeness of the three distance-based methods in comparison are significantly affected by the chosen cluster number, which is difficult to know in advance. Overall our Dual-HDP has better accuracies than other methods. For example, in order to achieve a completeness higher than ours, LCSS has to choose a cluster number smaller than 15 and in those cases it correctness is 27% lower than ours. In order to achieve a correctness higher than ours, LCSS has to choose a cluster number larger than 25, and its completeness is 13% lower than ours.

8.1.3 Evaluation on Simulated Data

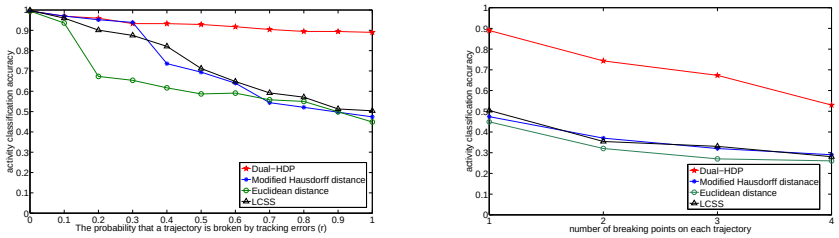
In this section, we simulate trajectories to evaluate how robust our model is to tracking errors. As shown in Figure 12 (a), eight paths are manually drawn on a scene. Some paths share the same semantic regions. A trajectory is randomly assigned to one of the eight predefined activities. A trajectory samples the location of its starting point from a Gaussian distribution centered at the starting point of its path with variance $\sigma_1 = 5$. It samples the remaining points sequentially following the direction specified by the path, with additive Gaussian noise of variance $\sigma_2 = 2$. The simulated trajectories are shown in Figure 12 (b). In reality, some trajectories are broken because of occlusions and scene clutter during tracking. In our simulation, we decide whether a trajectory is broken in a random way with probability r ($0 \leq r \leq 1$). If a trajectory is broken, the breaking point is uniformly sampled along the trajectory. A larger r simulates the case when there are more tracking errors. There are other types of tracking errors, such as wrong associations, not simulated in this experiment. However, breaking is one of the most common tracking errors, since some other tracking errors can be transferred to breaking errors by simply stopping tracking when the tracker is confused or there is not enough evidence to support the hypothesis. After the trajectories are clustered by our algorithm, we manually specify each of the clusters as an activity category, so each trajectory is assigned an activity label by our algorithm. By comparing with



(a) (b)

Fig. 12 Simulate trajectories of different activities. (a) The central lines of eight paths manually drawn in the scene. They are distinguished by different colors. (b) Trajectories simulated from the eight paths. They are also displayed in the eight colors.

the ground truth, the accuracy of activity classification is computed. Figure 13(a) plots the activity classification accuracies with different r. In this experiment, our algorithm always successfully converges to the right number of clusters according the ground truth. It is observed that the performance does not significantly drop when tracking errors increase. This shows that our algorithm is robust to tracking errors to some extent. We also compare our algorithm with Euclidean distance [37], modified Hausdorff distance [19] and long common subsequence (LCSS) [39]. The method using Euclidean distance requires that trajectories are temporally aligned. If two broken trajectories are on the same path but they only partially overlap or even have no overlap because of breaking errors, then their Euclidean distance is large. The modified Hausdorff distance encounters the similar problem, although working better than the Euclidean distance. The performance of these three distance-based methods, especially when using Euclidean distance, drops significantly when the trajectory data set has tracking errors. Our approach can well cluster trajectories with the existence of such tracking errors because it models the global pathways and cluster trajectories based on path models. Even if a broken trajectory only partially passes through a pathway, it still can be properly classified by the path model. Our learning process does not require trajectories are complete. Suppose broken trajectories have partial overlap. Some trajectories connect locations L_1 and L_2, some connect locations L_2 and L_3. Even if L_1 and L_3 are not directly connected, they are connected through L_2 and still can grouped into the same path model. In Figure 13 (a), each trajectory has at most one breaking point. Figure 13 (b) shows more challenging cases, where each trajectory is certainly broken and has more than one breaking points. The performance of our method drops as the number of breaking points increases. When there are more than one breaking points per trajectory, our methods does not exactly find the right number of clusters. However, its performance is still better than the other distance-based methods in comparison.



(a) (b)

Fig. 13 Activity classification accuracies of Dual-HDP and three distance-based methods (Euclidean distance [37], modified Hausdorff distance [19] and LCSS [39]) when the simulated trajectories are broken. (a) A trajectory is broken with different probabilities from 0 to 1. (b) A trajectory is certainly broken and has multiple breaking points. In our implementations, the modified Hausdorff distance and LCSS compare both spatial distance and velocity difference of observations on the trajectories.

8.2 Trajectory Analysis with Dynamic Modeling

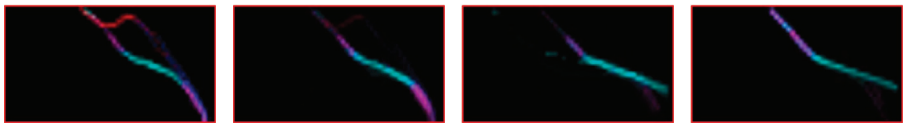
In this section, the results of the dynamic Dual-HDP model will be presented. The hyperparameters $a_1 \dots a_4$, $b_1 \dots b_4$ and H under dynamic Dual-HDP are set the same as those in Dual-HDP. The decreasing rate r controls how fast the influence of old data decreases and it is equal to 0.7 in our experimental settings. It is adjustable in practical applications. If r is large, the temporal variation of the learned cluster models is small. If $r = 0$, the models of clusters for different time slices are learned independently.

8.2.1 Results on Radar Tracks

In this section we conduct experiments on a much larger data set than that used in Section 8.1.1. It includes 8,478 radar tracks collected from 304 hours. The trajectories are divided into $T = 304$ slices by hours. In Figure 14, 15, 16, 17, and 18, we show some semantic regions learnt at different time slices. The first subfigure shows when this semantic region first appears as a new mode under the dynamic Dual-HDP model. As shown in Figure 14, semantic region 1 first appears at the 35th hour and its shape changes over time. As shown in Figure 15, semantic region 2 first appears at the 47th hour. However, it appears noisy in the first few time slices. Its shape forms after the 112nd hour. This mode gradually disappears after 244 hours [8]. Semantic region 3 and 4 are first coupled in the same topic at the early stage (see the first two subfigures in Figure 17). They are well separated when more data is observed later on.

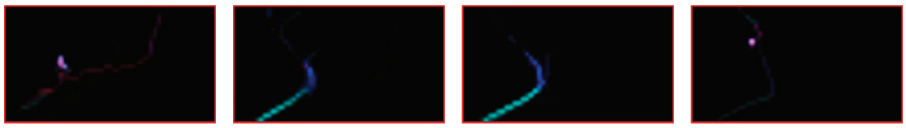
Figure 19 shows the abnormal radar tracks detected at different time. Since the activity models and semantic regions change over time, the detected abnormal trajectories are also different depending on the temporal context. Trajectories detected as abnormal at some time slices may become normal when they appear at other time slices. For example, as shown in Figure 19, some abnormal trajectories detected at the

[8] If no observations belonging to a certain semantic region are obtained in a time slice, the model of the semantic region will keep the same as the model, which is used as a prior, learned from previous time slices. But the prior will become weaker and weaker if no more data of this semantic region is observed. Eventually, it may be replaced by another new semantic region. Figure 15 shows that after no data is observed for a long time the model of that semantic region is degraded like a null region.



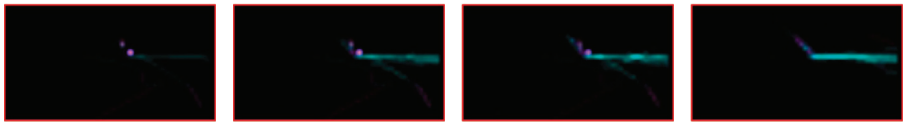
the 35th hour the 121st hour the 171st hour the 305th hour

Fig. 14 The dynamic change of semantic region 1 over time learnt from the radar tracks. In the first subfigure, we show when the semantic region first appears, i.e. semantic region 1 first appears as a new mode learnt by the Dual-HDP model at the 35th hour. Figure 15, 16, 17, 18 follow the same convention.



the 47th hour the 113rd hour the 201st hour the 245th hour

Fig. 15 The dynamic change of semantic region 2 over time learnt from the radar tracks.

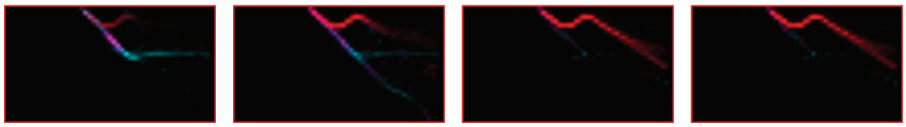


the 25th hour the 91st hour the 135th hour the 179th hour

Fig. 16 The dynamic change of semantic region 3 over time learnt from the radar tracks.

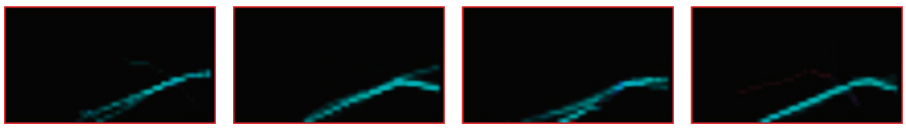
7th hour and the 9th hour actually pass through semantic regions 2, 3, 4, and 5. However, these modes are learnt later. In the first few hours, only a few trajectories passing through these semantic regions are observed. So they are detected as abnormal. When more trajectories of the same activities are observed and the activity models are well learnt in the later time slices, they will not be detected as abnormal any more.

As a quantitative evaluation, in Figure 20, we compare the data likelihoods of three methods: (1) dynamic Dual-HDP; (2) learning a different Dual-HDP from the data subset within each time slice independently, and (3) learning a single Dual-HDP model from the data in all the time slices. Half of the data is used to train the models, which are used compute the log likelihood of the remaining testing data. The average log likelihoods per trajectory within each time slice are shown in Figure 20. A higher data likelihood indicates that the models learned can better explain the data. It is observed that the dynamic Dual-HDP has a much better performance than the other two approaches. When learning different Dual-HDP models for different time slices independently, the data likelihood fluctuates dramatically because the data set within each time slice is small and it is easy for the learned models to overfit. When learning a single Dual-HDP model for all the time slices, the data likelihood is stable but lower than dynamic Dual-HDP, since a single model cannot well explain the distribution of data which change dynamically.



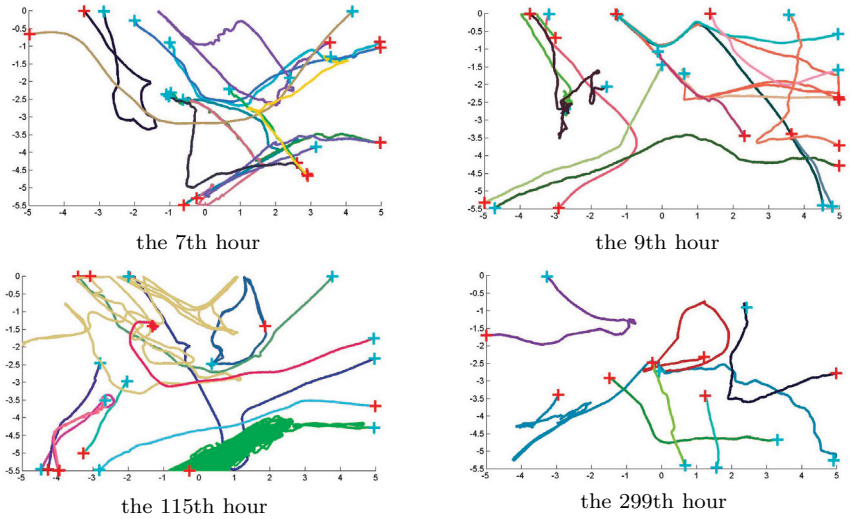
the 25th hour | the 157th hour | the 267th hour | the 305th hour

Fig. 17 The dynamic change of semantic region 4 over time learnt from the radar tracks.



the 15th hour | the 93rd hour | the 221st hour | the 305th hour

Fig. 18 The dynamic change of semantic region 5 over time learnt from the radar tracks.



the 7th hour | the 9th hour

the 115th hour | the 299th hour

Fig. 19 Abnormal radar tracks detected at different time slices. The same threshold of data likelihood is used for all the time slices.

8.2.2 Results on Tracks from a Park Lot

The 40, 453 trajectories of the parking lot data set are collected from one week. We divide them into time slices by hours. Figure 22, 21, and 23 show the dynamic change of some semantic regions over time. We can observe some cyclic change of the distributions of semantic regions. There are fewer activities happening around midnight and early in the morning. The distributions of semantic regions are sparser compared with those in the afternoon and in the evening. The second column of Figure 21 shows pathways crossing parking spaces. The reason is that early in the morning (before 8am), many parking spaces were empty and some pedestrians took a short cut crossing empty

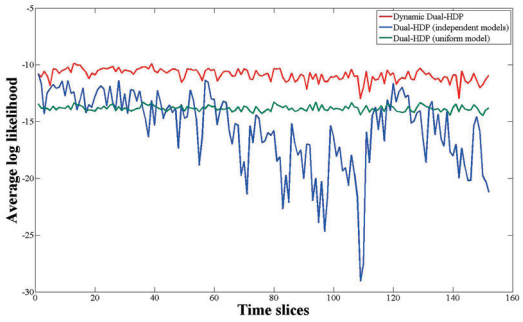


Fig. 20 Data log likelihood of using three different approaches to learn activity models on the radar data set. Dual-HDP (independent models): learn a Dual-HDP model for the data in each time slice independently; Dual-HDP (uniform model): learn a single Dual-HDP model for the data in all the time slices; Dynamic Dual-HDP: learn a dynamic Dual-HDP model. Half of the data in each time slice is selected to learn the activity models as training data and compute the average log-likelihood per trajectory (normalized by the lengths of trajectories) of the remaining data as testing data.

parking spaces [9]. As comparison, the third and the fourth columns of Figure 21 show that the pathways are restricted by parking cars in other time slots. In Figure 21, the shape of the semantic region at time slice between 13 o'clock and 14 o'clock on May 16 changes, because there are more people from the top entering the parking lot and exiting from left at the particular time interval. In Figure 23, the shape of semantic region 3 also changes over time. People may exit the parking lot from the left of the scene or from a gate in the middle area of the scene (somewhere between two rows of trees).

Figure 24 shows the abnormal trajectories detected at different time. Between 8 o'clock and 9 o'clock in the morning everyday, many trajectories passing through the bottom right corner of the scene are detected as abnormal. As shown in Figure 22, there are not many trajectories of this activity happening in the morning. When they start to appear in a large number between 8 o'clock and 9 o'clock, they are detected as abnormal since the algorithm have not seen this type of activities for a long time. In contrast, these kinds of trajectories are not detected as abnormal between 14 o'clock and 15 o'clock in the afternoon. This phenomenon can be understood as a delay of learning in some sense. Many trajectories detected as abnormal are those passing through the grass field. An interesting example occurred between 13 o'clock and 15 o'clock on May 16th. A worker was mowing the grass around this time. Many trajectories moving back and forth horizontally on the grass field are detected as an abnormality between 13 o'clock and 14 o'clock. However, after the model of this activity has been well learnt at this time slice, similar trajectories [10] are not detected as abnormal in the next hour.

In Figure 25, we compare the data log likelihood of dynamic Dual-HDP compared with the other two methods in the same way as described in Section 8.2.1. Dynamic

[9] We used the same background image, which was taken during the day time, in all the subfigures of Figure 21.

[10] After checking the data set, it shows that there are trajectories of this type between 14 o'clock and 15 o'clock.

Dual-HDP has a better performance, which is consistent with what we has observed on the radar set in Figure 20.

We also use the accuracies of completeness and correctness to compare three different methods: dynamic Dual-HDP, learning different Dual-HDP models for different time slices independently, and learning a single Dual-HDP model for all the time slices. The results are shown in Table 3. For each of the three selected time slices, 100 pairs of trajectories of the same activity categories and 100 pairs of trajectories of different activity categories are randomly sampled and labeled by subjects as ground truth. On the subset collected between 13 o'clock and 14 o'clock dynamic Dual-HDP has similar performance as Dual-HDP. However on the other two subsets, Dual-HDP is not as good as dynamic Dual-HDP because the subsets are smaller in size and the dynamic change of activity models.

8.3 Discussion

Our experiments on both large (the parking lot data set) and small (the radar data set) scale data sets have achieved good results. However, the Dual-HDP model may encounter problems if the number of trajectories is too small because there may not be enough trajectories connecting different locations of the same semantic region in order to learn the model of the semantic region effectively. The minimum number of trajectories sufficient for learning depends on the distributions of the trajectories, the complexity of activities, and also the size of cells after quantization. This is an open question left for the future work. Usually, if the size of cells is larger, a smaller number of trajectories are needed. The dynamic Dual-HDP model works well even if there are only a very small number of trajectories observed in a time slice, because the models learned from previous time slices serve as prior and they effectively overcome the overfitting problem. Even if there is only one trajectory in a time slice, this trajectory will be classified into one of the clusters learned from previous time slices, and the models will not be updated much by the data observed in this time slice. As shown in Figure 20 and 25, if trajectories within each time slice are clustered independently, the clustering performance drops significantly because the data set is not large enough.

9 Conclusion and Future Work

We propose a nonparametric hierarchical Bayesian framework, which uses an existing Dual-HDP model and a new dynamic Dual-HDP model to cluster trajectories, learn the models of semantic regions, and detect trajectories related to abnormal activities. Different from many existing spatial distance based trajectory clustering approaches with ad hoc nature, we formulate these problems in a transparent probabilistic way. The number of semantic regions and clusters of trajectories are learnt through the hierarchical Dirichlet processes. The proposed dynamic Dual-HDP model uses the models learned from historical data as priors to update the models of activities over time. It can better explain activities at different time. It clusters trajectories incrementally and does not have to keep old data in the memory. So it has much lower space and time complexity than Dual-HDP. This feature is very important if we need to cluster data collected over months or even years in surveillance applications.

	Dynamic Dual-HDP		Dual-HDP (independent models)		Dual-HDP (uniform model)	
Time slice	$r_{complete}$	$r_{correct}$	$r_{complete}$	$r_{correct}$	$r_{complete}$	$r_{correct}$
07-08, May 19	0.89	0.84	0.84	0.79	0.83	0.77
13-14, May 19	0.93	0.90	0.92	0.90	0.95	0.89
19-20, May 19	0.90	0.86	0.87	0.82	0.86	0.85

Table 3 Accuracies of completeness and correctness on the data subsets of three different time slices using dynamic Dual-HDP, Dual-HPD (independent models), and Dual-HDP (uniform model). The description of these models is the same as in Figure 20.

The collapsed Gibbs sampling algorithms used in this work are still computationally expensive, especially for Dual-HDP. In the future work, we will explore more efficient inference algorithms, such as variational inference. With the fast development of GPU and multi-core processors, parallel computing has become an important way to enhance the speed. Some parallel sampling algorithms have been developed for HDP [71]. We will also explore parallel sampling algorithms for our models.

In our dynamic Dual-HDP model, a single point estimation of model parameters is taken at a time slice and it is used as prior for the inference of the next time slice. This model is greedy and information about the uncertainty of parameters is ignored. One alternative is to use particle filtering to propagate uncertainty and to avoid local minimum. This is another direction to be explored in the future work.

Our dynamic Dual-HDP models the temporal smoothness between two successive time slices. There are other types of temporal correlations not modeled by dynamic Dual-HDP. For example, the activity models for 8am-9am across different days should share more similarity and some smoothness constraints could be added among them, since they are all for rush hours. Similarly, activity models for Sundays may be more correlated. In order to model these temporal correlations, more priors could be added into the hierarchical Bayesian models at different levels.

A big challenge for research on trajectory clustering for activity analysis in video surveillance is that it is difficult to obtain the ground truth, especially for large scale data sets and complicated scenes. Although in this work we required subjects to label a pair of trajectories as belonging to the same activity categories or not to measure completeness and correctness, it does not reveal all the aspects of the desired ground truth information. One possible solution is to use the simulated data with ground truth. However, it has many open problems to be answered, such as how to simulate different types of tracking errors happening in real surveillance scenarios, how to simulate outlier trajectories and how to simulate the dynamic variations of activity models over time. A good simulator in the future will help the evaluation of our algorithms and also the development of new algorithms.

10 Appendix: Collapsed Gibbs Sampling for Dynamic Dual-HDP

Under dynamic Dual-HDP, $\boldsymbol{\pi}_0^t = (\pi_{01}^t, \ldots, \pi_{0K^t}^t, \pi_{0u}^t)$ and $\{\phi_k^t\}_{k=1}^{K^t}$ are the variables to be sampled. Given $\boldsymbol{\pi}_0^t$ and $\{\phi_k^t\}$, sampling other variables is the same as Dual-HDP. We focus on sampling $\boldsymbol{\pi}_0^t$ and $\{\phi_k^t\}$ given other variables and suppose the semantic region assignments to observations on trajectories at t are given. We use the Chinese

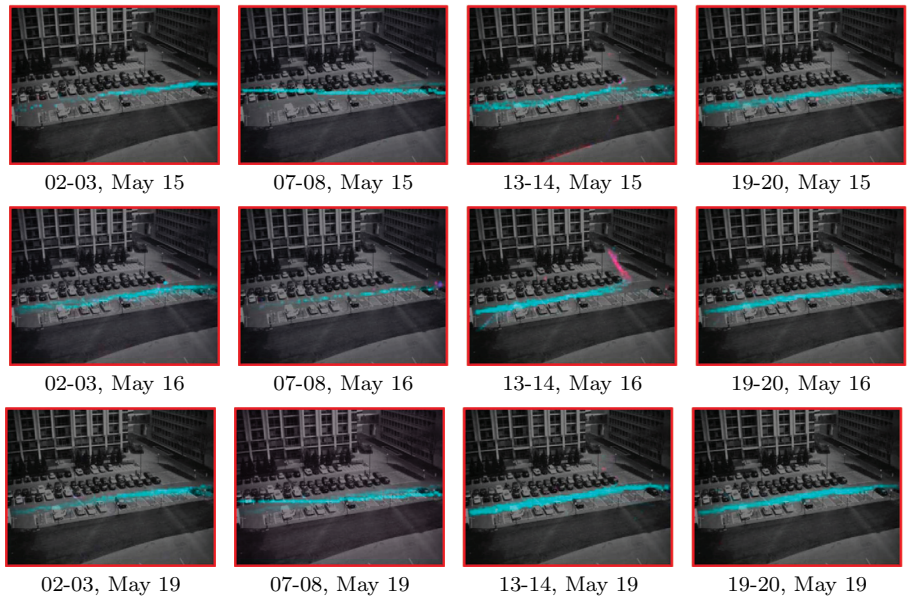


Fig. 21 The dynamic change of semantic region 1 over time learnt from the trajectories collected from a parking lot.

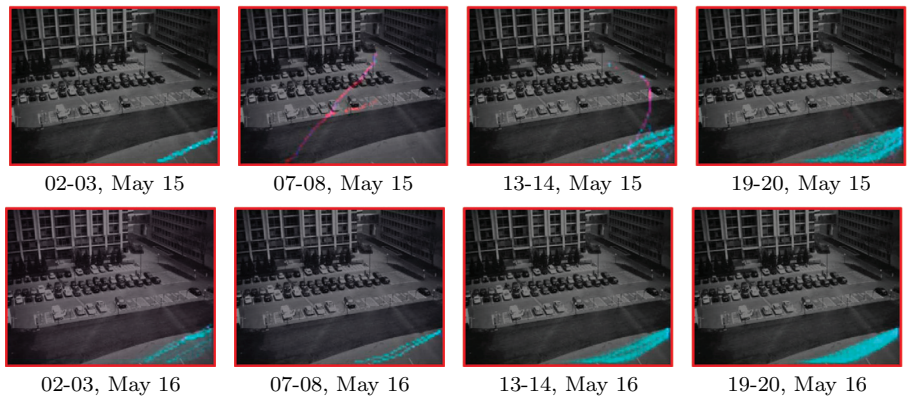


Fig. 22 The dynamic change of semantic region 2 over time learnt from the trajectories collected from a parking lot.

Restaurant Franchise sampling [11], which was also used by HDP [26] and Dual-HDP [2]. Under the Chinese Restaurant Franchise samping scheme, let n_{kw} be the number of observations with value w assigned to semantic region k, n_k be the total number of observations assigned to semantic region k, s_j be the number of big tables serving dish (semantic region) k, and s be the total number of big tables. n_{kw}, n_k, s_k, s are all

[11] When we describe our sampling algorithm, some terminologies such as "table", "big tables", "dish" (dishes correspond to semantic regions) in Chinese Restaurant Franchise are used. Because of space limit, their meanings cannot be well explain in this paper. Find their details from [26] and [2].

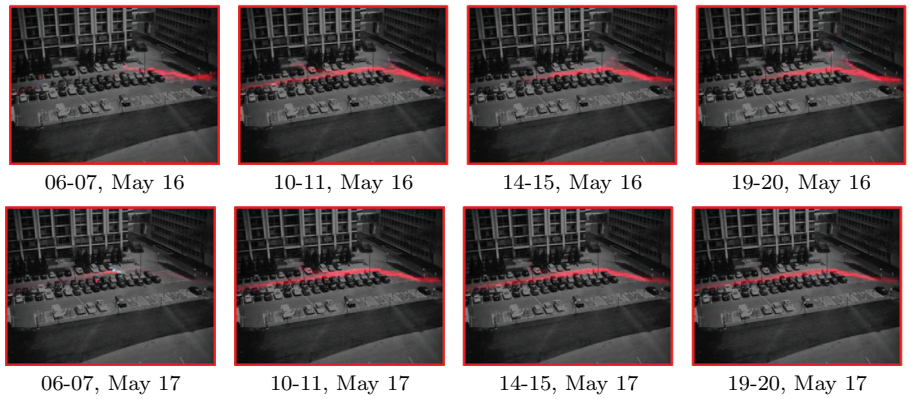


Fig. 23 The dynamic change of semantic region 3 over time learnt from the trajectories collected from a parking lot.

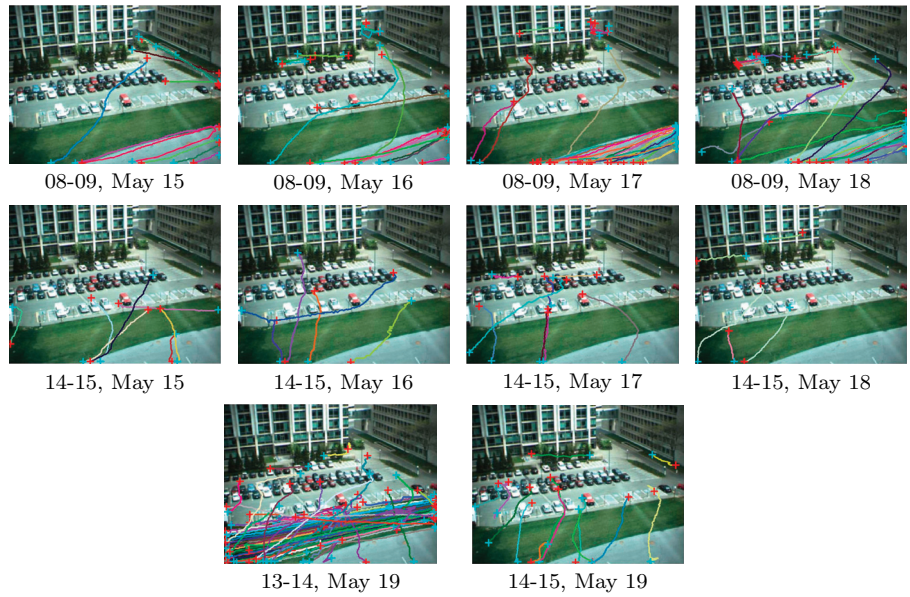


Fig. 24 Abnormal trajectories of the parking lot data set detected at different time slices. The same threshold of data likelihood is used for all the slices.

statistics from the data subset of time t. Since G_0^{t-1} provides prior of G_0^t as shown in Eq (6) (7) (8),

$$p(\boldsymbol{\pi}_0^t|\{\hat{\pi}_{0k}^{t-1}\}_{k=1}^{K^{t-1}}) = Dir(\gamma^t\omega^t\hat{\pi}_{01}^{t-1},\ldots,\gamma^t\omega^t\hat{\pi}_{0K^{t-1}}^{t-1},0,\ldots,0,\gamma^t(1-\omega^t)). \quad (10)$$

When $1 \leq k \leq K^{t-1}$,

$$p(\phi_k^t|\phi_k^{t-1}) = Dir(\xi_k^t \cdot \phi_k^{t-1} + H), \quad (11)$$

and when $K^{t-1} < k \leq K^t$,

$$p(\phi_k^t) = Dir(H). \quad (12)$$

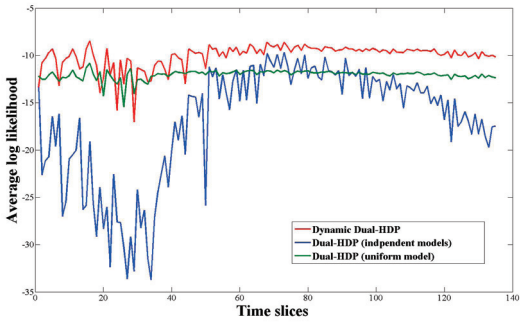


Fig. 25 Data log likelihood of using three different approaches to learn activity models on the parking lot data set. Dual-HDP (independent models): learn a Dual-HDP model for the data in each time slice independently; Dual-HDP (uniform model): learn a single Dual-HDP model for the data in all the time slices; Dynamic Dual-HDP: learn a dynamic Dual-HDP model. Half of the data in each time slice is selected to learn the activity models as training data and compute the average log-likelihood per trajectory of the remaining data as testing data.

The data likelihoods are

$$p(n_{k1},\ldots,n_{kW}|n_k,\phi_k^t) = Multinomial(n_k,\phi_k^t), \tag{13}$$

where W is the size of the codebook, and

$$p(s_1,\ldots,s_{K^t}|s,\boldsymbol{\pi}_0^t) = Multinomial(s,\boldsymbol{\pi}_0^t). \tag{14}$$

So $\boldsymbol{\pi}_0^t$ and ϕ_k^t can be sampled from posteriors,

$$\begin{aligned}&\boldsymbol{\pi}_0^t|\{s_k\}_{k=1}^{K^t},\{\hat{\pi}_{0k}^{t-1}\}_{k=1}^{K^{t-1}}\\ &\sim Dir(s_1+\gamma^t\omega^t\hat{\pi}_{01}^{t-1},\ldots,s_{K^{t-1}}+\gamma^t\omega^t\hat{\pi}_{0K^{t-1}}^{t-1},s_{K^{t-1}+1},\ldots,s_{K^t},\gamma^t(1-\omega^t)),\end{aligned} \tag{15}$$

when $1\le k\le K^{t-1}$,

$$\phi_k^t|\{n_{kw}\}_{w=1}^W,\phi_k^{t-1} \sim Dir(n_{k1}+\xi_k^t\cdot\phi_{k1}^{t-1}+u_1,\ldots,n_{kW}+\xi_k^t\cdot\phi_{kW}^{t-1}+u_W), \tag{16}$$

where $H=(u_1,\ldots,u_W)$. When $K^{t-1}<k\le K^t$,

$$\phi_k^t \sim Dir(n_{k1}+u_1,\ldots,n_{kW}+u_W). \tag{17}$$

Properly choosing ω^t, γ^t and ξ_k^t, we can control how much the old data up to $t-1$ influences the inference of models of the current time t. In this work, we choose

$$\omega^t = \frac{r\cdot s^{t-1}}{r\cdot s^{t-1}+\gamma}, \tag{18}$$

$$\gamma^t = r\cdot s^{t-1}, \tag{19}$$

$$\xi_k^t = r\cdot n_k^{t-1}. \tag{20}$$

r is a scalar between 0 and 1 controlling how fast the influence of old data decrease. n_k^t and s^t are the accumulated effective numbers of words assigned to topic k and big tables. They are updated over time,

$$n_k^t = r\cdot n_k^{t-1}+n_k, \tag{21}$$

$$s^t = r \cdot s^{t-1} + s. \tag{22}$$

Remind that n_k and s are the statistics obtained from the subset of t. At initialization $n_k^0 = 0$ and $s^0 = 0$. Then Eq 15 and 16 become

$$\begin{aligned}&\boldsymbol{\pi}_0^t | \{s_k\}_{k=1}^{K^t}, \{\hat{\pi}_{0k}^{t-1}\}_{k=1}^{K^{t-1}} \\ &\sim Dir(s_1 + rs^{t-1}\hat{\pi}_{01}^{t-1}, \ldots, s_{K^{t-1}} + rs^{t-1}\hat{\pi}_{0K^{t-1}}^{t-1}, s_{K^{t-1}+1}, \ldots, s_{K^t}, \gamma),\end{aligned} \tag{23}$$

$$\phi_k^t | \{n_{kw}\}_{w=1}^{W}, \phi_k^{t-1} \sim Dir(n_{k1} + rn_k^{t-1}\phi_{k1}^{t-1} + u_1, \ldots, n_{kW} + rn_k^{t-1}\phi_{kW}^{t-1} + u_W). \tag{24}$$

When the data becomes older, its influence on the current models is weaker. The decreasing rate is r. The inference under dynamic Dual-HDP is summarized in Algorithm 1.

References

1. X. Wang, K. T. Ma, G. Ng, and E. Grimson. Trajectory analysis and semantic region modeling using a nonparameteric hierarchical bayesian model. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2008.
2. X. Wang, X. Ma, and W. E. L. Grimson. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31:539–555, 2009.
3. L. Zelnik-Manor and M. Irani. Event-based analysis of video. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2001.
4. H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2004.
5. T. Xiang and S. Gong. Video behaviour profiling and abnormality detection without manual labelling. In *Proc. of IEEE Int'l Conf. Computer Vision*, 2005.
6. P. Smith, N. Lobo, and M. Shah. Temporalboost for event recognition. In *Proc. of IEEE Int'l Conf. Computer Vision*, 2005.
7. Y. Wang, T. Jiang, M. S. Drew, Z. Li, and G. Mori. Unsupervised discovery of action classes. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2006.
8. T. T. Truyen, D. Q. Phung, H. H. Bui, and S. Venkatesh. Adaboost.mrf: Boosted markov random forests and application to multilevel activity recognition. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2006.
9. X. Wang, X. Ma, and E. Grimson. Unsupervised activity perception by hierarchical bayesian models. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2007.
10. N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition. In *Proc. of British Machine Vision Conference*, 1995.
11. J. Fernyhough, A. Cohn, and D. Hogg. Generation of semantic regions from image sequences. In *Proc. of European Conf. Computer Vision*, 1996.
12. C. Stauffer and E. Grimson. Learning patterns of activity using real-time tracking. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000.
13. I. Haritaoglu, D. Harwood, and L. S. Davis. W4: Real-time surveillance of people and their activities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22:809–830, 2000.
14. M. Brand and V. Kettnaker. Discovery and segmentation of activities in video. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22:844–851, 2000.
15. S. Honggeng and R. Nevatia. Multi-agent event recognition. In *Proc. of IEEE Int'l Conf. Computer Vision*, 2001.
16. G. Medioni, I. Cohen, F. BreAmond, S. Hongeng, and R. Nevatia. Event detection and analysis from video streams. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23:873–889, 2001.
17. D. Makris and T. Ellis. Automatic learning of an activity-based semantic scene model. In *Proc. of AVSBS*, 2003.

18. T. Xiang and S. Gong. Beyond tracking: Modelling activity and understanding behaviour. *International Journal of Computer Vision*, 67:21–51, 2006.
19. X. Wang, K. Tieu, and E. Grimson. Learning semantic scene models by trajectory analysis. In *Proc. of European Conf. Computer Vision*, 2006.
20. W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank. A system for learning statistical motion patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28:1450–1464, 2006.
21. Z. Zhang, K. Huang, T. Tan, and L. Wang. Trajectory series analysis based event rule induction for visual surveillance. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2007.
22. D. Makris and T. Ellis. Path detection in video surveillance. *Image and Vision Computing*, 20:859–903, 2002.
23. I. Junejo, O. Javed, and M. Shah. Multi feature path modeling for video surveillance. In *Proc. of IEEE Int'l Conf. Pattern Recognition*, 2004.
24. I. Junejo and H. Foroosh. Trajectory rectification and path modeling for video surveillance. In *Proc. of IEEE Int'l Conf. Computer Vision*, 2007.
25. R. Kaucic, A. Perera, G. Brooksby, J. Kaufhold, and A. Hoogs. A unified framework for tracking through occlusions and across sensor gaps. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2005.
26. Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical dirichlet process. *Journal of the American Statistical Association*, 2006.
27. A. Rodriguez, D. B. Dunson, and A. E. Gelfand. The nested dirichlet process. Technical report, Working Paper 2006-19, Duke Institute of Statistics and Decision Sciences., 2006.
28. E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky. Describing visual scenes using transformed objects and parts. *International Journal of Computer Vision*, 77:291–330, 2007.
29. D. M. Blei and J. D. Lafferty. Dynamic topic models. In *Proc. of International Conference on Machine Learning*, 2006.
30. S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems.* Artech House, 1999.
31. N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22:831–843, 2000.
32. E. B. Fox, D. S. Choi, and A. S. Willsky. Nonparameteric bayesian methods for large scale multi-target tracking. In *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, 2006.
33. G. Gennari and G. D. Hager. Probabilistic data association methods in visual tracking of groups. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 04.
34. P. Nillius, J. Sullivan, and S. Carlsson. Multi-target tracking - linking identities using bayesian network inference. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2006.
35. S. K. Pang, J. Li, and S. J. Godsill. Models and algorithms for detection and tracking of coordinated groups. In *Proceedings of Aerospace Conference*, 2008.
36. Z. Fu, W. Hu, and T. Tan. Similarity based vehicle trajectory clustering and anomaly detection. In *Proc. of IEEE Int'l Conf. Image Processing*, 2005.
37. W. Hu, D. Xie, Z. Fu, W. Zeng and S. Mayband Semantic-Based Surveillance Video Retrieval In *IEEE Trans. on Image Processing*, 16:1168-1181, 2007.
38. E. Keogh and M. Pazzani. Scaling up dynamic time scaling up dynamic time. In *Proc. of ACM SIGKDD*, 2000.
39. M. Vlachos, G. Kollios, and D. Gunopulos. Discovering Similar Multidimensional Trajectories In *Proc. IEEE Conf. Data Engineering*, 2002.
40. A. Y. Ng, M. I. Jordan, and Y. Weiss On Spectral Clustering: Analysis and an Algorithm In *Proc. Neural Information Processing Systems Conf.*, 2002.
41. X. Li, W. Hu, and W. Hu A Coarse-to-Fine Strategy for Vehicle Motion Trajectory Clustering. In *Proc. Int'l Conf. Pattern Recognition*, 2006.
42. B. Morris and M Trivedi Learning Trajectory Patterns by Clustering: Experimental Studies and Comparative Evaluation In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2009.
43. C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the Nystrom method. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26:214-225, 2004.

44. N. Anjum and A. Cavallaro. Multifeature Object Trajectory Clustering for Video Analysis In *IEEE Trans. on Circuits and Systems for Video Technology*, 18:1555-1564, 2008.
45. T. Zhang, H. Lu, and S. Z. Li. Learning Semantic Scene Models by Object Classification and Trajectory Clustering In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2009.
46. I. Saleemi, K. Shafique, and M. Shah Probabilistic Modeling of Scene Dynamics for Applications in Visual Surveillance In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31:1472-1485, 2009.
47. R. Vidal and R. Hartley. Motion Segmentation with Missing Data Using Powerfactorization and GPCA. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2004.
48. S. Rao, R. Tron, R. Vidal, and Y. Ma. Motion Segmentation in the Presence of Outlying, Incomplete, or Corrupted Trajectories. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32:1832-1845, 2010.
49. J. Rittscher, P. Tu, and N. Krahnstoever. Simultaneous Estimation of Segmentation and Shape. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2005.
50. G. Brostow, and R. Cippola. Unsupervised Bayesian Detection of Independent Motion in Crowds. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2006.
51. D. B. Dunson, N. Pillai, and J. H. Park. Bayesian density regression. *Journal of the Royal Statistical Society: Series B*, 69:163–183, 2006.
52. J. E. Griffin and M. F. J. Steel. Order-based dependent dirichlet processes. *Journal of the American Statistical Association*, 101:179–194, 2006.
53. F. Caron, M. Davy, and A. Doucet. Generalized polya urn for time-varying dirichlet process mixtures. In *Proc. of Uncertainty in Artificial Intelligence*, 2007.
54. X. Zhu, Z. Ghahramani, and J. Lafferty. Time-sensitive dirichlet process mixture model. Technical report, Technical Report, School of Computer Science, Carnegie Mellon University, 2005.
55. L. Ren, D. B. Dunson, and L. Carin. The dynamic hierarchical dirichlet process. In *Proc. of International Conference on Machine Learning*, 2008.
56. N. Srebro and Roweis. Time-varying topic models using dependent dirichlet process. Technical report, Technical Report, Department of Computer Science, University of Toronto, 2005.
57. T. Hofmann. Probabilistic latent semantic analysis. In *Proc. of Uncertainty in Artificial Intelligence*, 1999.
58. D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
59. E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky. Describing visual scenes using transformed dirichlet processes. In *Proc. of Neural Information Processing Systems Conference*, 2005.
60. X. Wang and E. Grimson. Spatial latent dirichlet allocation. In *Proc. of Neural Information Processing Systems Conference*, 2007.
61. L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2005.
62. J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman. Discovering object categories in image collections. In *Proc. of IEEE Int'l Conf. Computer Vision*, 2005.
63. J. C. Niebles, H. Wang, and L. Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. In *Proc. of British Machine Vision Conference*, 2006.
64. J. C. Niebles and L. Fei-Fei. A hierarchical model of shape and appearance for human action classification. In *Proc. of IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2007.
65. T. S. Ferguson. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230, 1973.
66. J. Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.
67. A. Gelman, H. S. Stern, and H. S. Rubin. Bayesian Data Analysis. *CRC Press*, 2004
68. S. MacEachern, A. Kottas, and A. Gelfand. Spatial nonparametric bayesian models. Technical report, Institute of Statistics and Decision Sciences, Duke University, 2001.
69. Y. W. Teh Dirichlet Processes. *Encyclopedia of Machine Learning*, 2010.
70. B. Moberts, A. Vilanova, and J. W. Jake. Evaluation of Fiber Clustering Methods for Diffusion Tensor Imaging. *Proc. of IEEE Visualization*, 2005.
71. A. Asuncion, P. Smyth, and M. Welling. Asynchronous Distributed Learning of Topic Models. *Proc. of Neural Information Processing Systems Conference*, 2008.