Trajectory Analysis and Semantic Region Modeling Using A Nonparametric Bayesian Model

Xiaogang Wang[1] Keng Teck Ma[2] Gee-Wah Ng[2] W. Eric L. Grimson[1]

[1]CS and AI Lab, MIT,
77 Massachusetts Ave., Cambridge, MA, 02139, USA
{xgwang, welg}@csail.mit.edu

[2]DSO National Laboratories,
20 Science Park Drive Singapore 118230
{mkengtec, ngeewah}@dso.org.sg

Abstract

We propose a novel nonparametric Bayesian model, Dual Hierarchical Dirichlet Processes (Dual-HDP), for trajectory analysis and semantic region modeling in surveillance settings, in an unsupervised way. 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 intersections of paths commonly taken by objects, related to activities in the scene are also modeled. Dual-HDP advances the existing Hierarchical Dirichlet Processes (HDP) language model. HDP only clusters co-occurring words from documents into topics and automatically decides the number of topics. Dual-HDP co-clusters both words and documents. It learns both the numbers of word topics and document clusters from data. Under our problem settings, HDP only clusters observations of objects, while Dual-HDP clusters both observations and trajectories. Experiments are evaluated on two data sets, radar tracks collected from a maritime port and visual tracks collected from a parking lot.

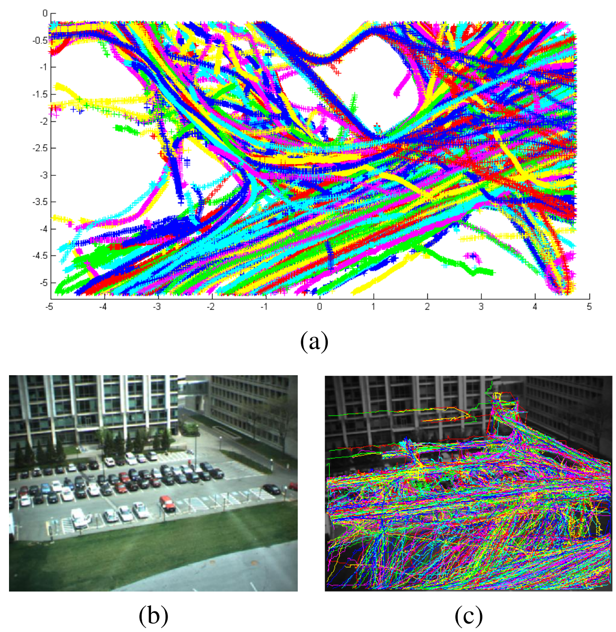

Figure 1. Trajectories in the our two data sets. (a) Radar tracks collected from a port. (b) The background image of a parking lot. (c) Tracks collected from a parking lot scene (only $4,404$ out of $40,453$ tracks are shown here).

1. Introduction

Activity analysis has always been one of the foci of research in surveillance. Over the past decade significant work has been reported on this topic. Although some approaches [18, 10, 14] modeled activities by directly extracting motion and appearance features from the videos without relying on tracking, most approaches [11, 2, 8, 16, 15, 17] 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. Through tracking, an activity executed by a single object can be separated from other co-occurring activities, and features related to the activity can be integrated as a track. In many far-field surveillance settings, 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, such as gestures, local motions, or appearance of objects within the tracks. Usually only positions of objects are recorded along the tracks, which are called trajectories. Although quite simple, the information encoded by trajectories can distinguish many different activity patterns, especially in far-field settings. The goal of

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

this work is to model activities by trajectory analysis: clustering trajectories into different activities, detecting abnormal trajectories, and modeling semantic regions.

We propose a framework using a nonparametric Bayesian model, Dual Hierarchical Dirichlet Processes (*Dual-HDP*), for trajectory analysis. *Dual-HDP* advances the existing Hierarchical Dirichlet Processes (*HDP*) [12] language model. *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. Wang et al. [14], proposed an *HDP* mixture model to co-cluster both words and documents. However, it required one to manually specify the number of document clusters. Our *Dual-HDP* also co-clusters words and documents, but it automatically decides the numbers of both word topics and document clusters. 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 intersections of paths commonly taken by objects, in the scene, and trajectories are clustered into different activities.

We evaluate our approach on two data sets (see Figure 1): 577 radar tracks collected from a port in 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 [11] to obtain tracks in this data set.

2. Related Work

Most 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 [3], Hausdorff distance and its variations [4, 15], and Dynamic Time Warping (*DTW*) [6]. These similarity-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 larger distance to other trajectories. This abnormality detection lacks a probabilistic explanation. Second, they do not provide a solution to the number of clusters. They require that the cluster number is known in advance. Third, they measure the spatial distance between observations on two trajectories. However, spatial distance does not reflect the statistical nature of activities. For example, vehicles moving on two side by side lanes may be close in space, but their trajectories represent different activities. Spatial distance is also sensitive to projective distortion. 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.

Trajectory clustering is also related to the problem of modeling semantic regions in the scene. The knowledge of the structures of the scene (e.g. roads, paths, entry and exit points) can help not only the high-level description of activities [15], but also low-level tracking and classification [5]. It takes a lot of effort to manually input these structures. They cannot be reliably detected based on the appearance of the scene either. In some cases, e.g. detecting shipping fairways on the sea, there is no appearance cue available at all. It is of interest to detect these structures by trajectory analysis. Usually paths are detected by modeling the spatial extents of trajectory clusters [2, 8, 15]. Semantic regions are detected as intersections of paths [8]. Entry and exit points are detected at the ends of paths [15].

Our framework differs from previous approaches:

- Different from prior similarity-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 number of activity categories and semantic regions are automatically learnt from data instead of requiring manual definition.
- Instead of using a spatial distance measure uniformly over the scene, it models the spatial distributions of activities. It separates activity-related structures close in space. It is more robust to projective distortion.
- The space complexity of our algorithm is $O(N)$ instead of $O(N^2)$ in the number of 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 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 *HDP*, cluster co-occurring words into one topic.

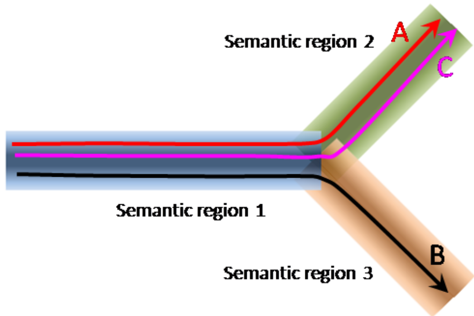

Figure 2. An example to explain the modeling of semantic regions and activities. See details in text.

Each topic has a multinomial distribution over the codebook. A document is modeled as a mixture of topics and documents share topics. If some words, such as "professor" and "education", often but not necessarily always occur in the same documents, a topic related to "education" will be learnt and its multinomial distribution has large weights on these words. When these 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. Semantic regions are intersections of paths. Two paths may partially share one semantic region. A semantic region is modeled as a multinomial distribution over the space of the scene and moving directions. If two trajectories pass through the same set of semantic regions, they belong to the same activity. In our *Dual-HDP* model, each activity cluster has a prior distribution over topics (semantic regions). It is learnt in an unsupervised way. All the trajectories clustered into the same activity 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.

With the "bag-of-words" assumption, our approach does model the first order temporal information among observations since the codebook encodes the moving directions. 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. [12] and our *Dual-HDP* model, which is actually used for trajectory analysis. We will describe them as language models. However, remember that in our problem documents are trajectories, words are observations, and topics are semantic regions. Clusters of trajectories (activities) are explicitly modeled in *Dual-HDP* but not in *HDP*.

4. Dirichlet Process

A Dirichlet Process (*DP*) [1] is a nonparametric distribution whose domain is a set of probability distributions. A *DP* 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 [9],

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

where δ_{ϕ_k} is a Dirac delta function centered at ϕ_k, ϕ_k is a multinomial parameter vector sampled from Dirichlet distribution H, $\phi_k \sim H$, and π_k is a non-negative scalar satisfying $\sum_{k=1}^{\infty} \pi_k = 1$, $\pi_k = \pi'_k \prod_{l=1}^{k-1}(1-\pi'_l)$, $\pi'_k \sim Beta(1, \alpha)$. G is often used as a prior for infinite mixture models. When data points are sampled from G, there is no limit to the number of distinct components which may be generated. Given a set of data points $\theta_1, \ldots, \theta_N$ sampled from G, it turns out that the posterior of sampling a new data point can be obtained 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 $\{\theta_k^*\}_{k=1}^{K}$ (identifying K components) among the N data points. n_k is the number of points with value θ_k^*. The new data point θ_{N+1} can be assigned to one of the existing components or can sample a new component from H. These properties make *DP* ideal for modeling data clustering problems where the number of mixture components is not well-defined in advance.

5. *HDP*

HDP proposed by Teh et al. [12] is a nonparametric hierarchical Bayesian model used to cluster co-occurring words in documents into topics (in our problem it clusters observations on the trajectories into semantic regions). The graphical model of *HDP* is shown in Figure 3. There are M documents (trajectories) in the corpus. Each document j has N_j words (quantized observations of positions and moving directions of objects). In *HDP*, a prior distribution G_0 over the whole corpus is sampled from a Dirichlet process, $G_0 \sim DP(\gamma, H)$. $G_0 = \sum_{k=1}^{\infty} \pi_{0k} \delta_{\phi_k}$. ϕ_k is the parameter of a topic, which is modeled as a multinomial distribution over the codebook. ϕ_k is sampled from Dirichlet

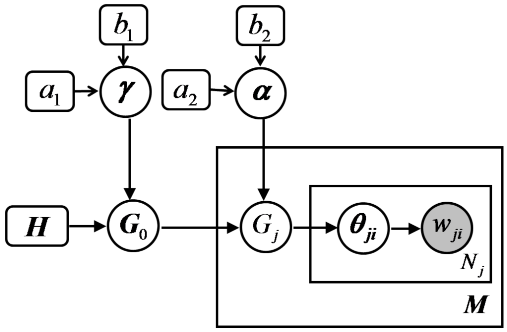

Figure 3. The graphical model of *HDP*

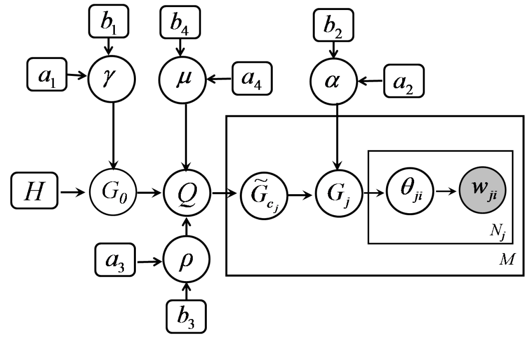

Figure 4. The graphical model of *Dual-HDP*

prior H. All the words in the corpus will be sampled from some topics $\{\phi_k\}$. For each document j, a prior distribution G_j over all the words in that document is sampled from Dirichlet process, $G_j \sim DP(\alpha, G_0)$. $G_j = \sum_{k=1}^{\infty} \pi_{jk}\delta_{\phi_k}$ share the same components ϕ_k as G_0, i.e. all the documents share the same set of topics. For each word i in document j, a topic θ_{ji}, which is one of the ϕ_k's, is sampled from G_j. The word value w_{ji} is sampled from the topic θ_{ji}, $w_{ji} \sim Discrete(\theta_{ji})$. The concentration parameters are sampled from some gamma priors, $\gamma \sim Gamma(a_1, b_1)$, $\alpha \sim Gamma(a_2, b_2)$. In *HDP*, all the documents share topics and the number of topics, i.e. the number of non-zero elements of $\{\pi_k\}$ is learnt from data.

6. *Dual-HDP*

Unfortunately, *HDP* does not cluster documents (trajectories in our problem). We propose a *Dual-HDP* model to co-cluster both words and documents. A document is modeled as a distribution over topics. Thus documents with similar distributions over topics can be grouped into one cluster. There are two hierarchical Dirichlet processes modeling topics of words and clusters of documents. The graphical model of *Dual-HDP* is shown in Figure 4.

In *Dual-HDP*, each document j is from one of the document clusters. All the documents in cluster c have the same prior distribution $\widetilde{G}_c$. $\widetilde{G}_c = \sum_{k=1}^{\infty} \widetilde{\pi}_{ck}\delta_{\widetilde{\phi}_{ck}}$ is an infinite mixture of topics. Since the number of document clusters is not known in advance, we model the clusters of documents as an infinite mixture,

$$Q = \sum_{c=1}^{\infty} \epsilon_c \delta_{\widetilde{G}_c} \tag{3}$$

When a *DP* was first developed by Ferguson [1], the components (such as ϕ_k in Eq 1) could only be scalars or vectors. MacEachern [7] 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 3, $\epsilon_c = \epsilon'_c \prod_{l=1}^{c-1}(1-\epsilon'_l)$, $\epsilon'_c \sim Beta(1, \mu)$, $\widetilde{G}_c \sim DP(\rho, G_0)$. As explained in Section 5, $G_0 \sim DP(\gamma, H)$ is the prior distribution over the whole corpus. $\{\widetilde{G}_c\}_{c=1}^{\infty}$ all have the same topics in G_0. i.e. $\widetilde{\phi}_{ck} = \phi_k$. However they have different mixtures $\{\widetilde{\pi}_{ck}\}$ over these topics. Each document j samples a probability measure $\widetilde{G}_{c_j}$ from Q as its prior. Different documents may choose the same prior $\widetilde{G}_c$, thus they form one cluster c. Then document 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*. Word i in document j samples its topic θ_{ji} from G_j and samples its word value w_{ji} from $Discrete(\theta_{ji})$. The concentration parameters are also sampled from gamma priors.

Gibbs sampling is used to do inference in three steps.

1. Given the cluster assignment $\{c_j\}$ of documents, sample the word topic assignment $\{z_{ji}\}$ ($z_{ji} = k$ indicates $\theta_{ji} = \phi_k$), topic 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. [12] 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 cluster assignment c_j of documents. c_j can be assigned to one of the existing document clusters or to a new cluster. We use the Chinese restaurant franchise for sampling. See details in [13].

3. Given other variables, sample the concentration parameters using the sampling scheme proposed in [12].

In order to detect abnormal documents (trajectories), we need to compute the likelihood of document j given other documents, $p(\mathbf{w}_j|\mathbf{w}^{-j})$, where $\mathbf{w}_j = \{w_{ji}\}_{i=1}^{N_j}$ is the set words in document j and $\mathbf{w}^{-j}$ represents the remaining documents excluding j. It can be approximated using the samples obtained during Gibbs sampling and a variational method. See details in [13].

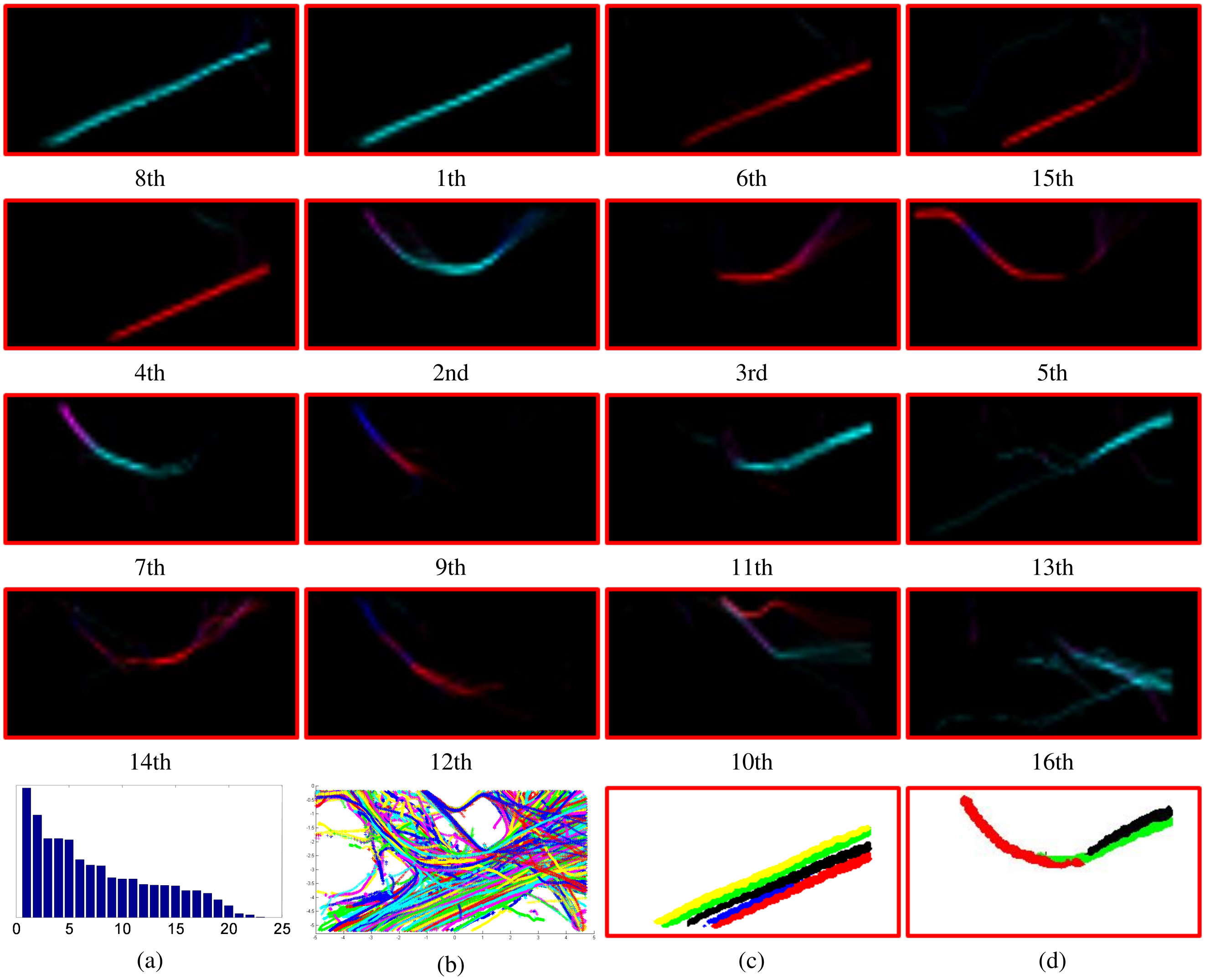

Figure 5. Semantic regions at a maritime port learnt from the radar tracks. Distributions of the first 16 semantic regions over space and moving directions are shown (for easier comparison, they are not shown in order). Colors represent different moving directions: $\rightarrow$ (red), $\leftarrow$ (cyan), $\uparrow$ (magenta), and $\downarrow$ (blue). (a) Histogram of observations assigned to different semantic regions. (b) All of the radar tracks. (c) Compare the 1st, 4th, 6th, 8th, and 15th semantic regions. (d) Compare the 7th, 11th, and 13th semantic regions (see details in text).

7. Results on radar tracks

There are 577 radar tracks in our maritime port data set. They were acquired by multiple collaborating radars along the shore and recorded the locations of ships on the sea. 23 semantic regions are discovered by our model. In Figure 5, we display the distributions of the first 16 semantic regions (sorted by the number of observations assigned to semantic regions) over space and moving directions. As shown in Figure 5, 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 5 (c) in different colors, green (1st), red (4th), black (6th), yellow (8th), and blue (15th). Since they are so close in space, they cannot be separated using spatial distance based trajectory clustering approaches. In Figure 5 (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.

Our approach groups trajectories into 16 clusters. In Figure 6, we plot the eight largest clusters and some smaller clusters. Clusters 1, 4, 6 and 7 are close in space but occupy different regions. 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 5(d), ships first move along the same way and then

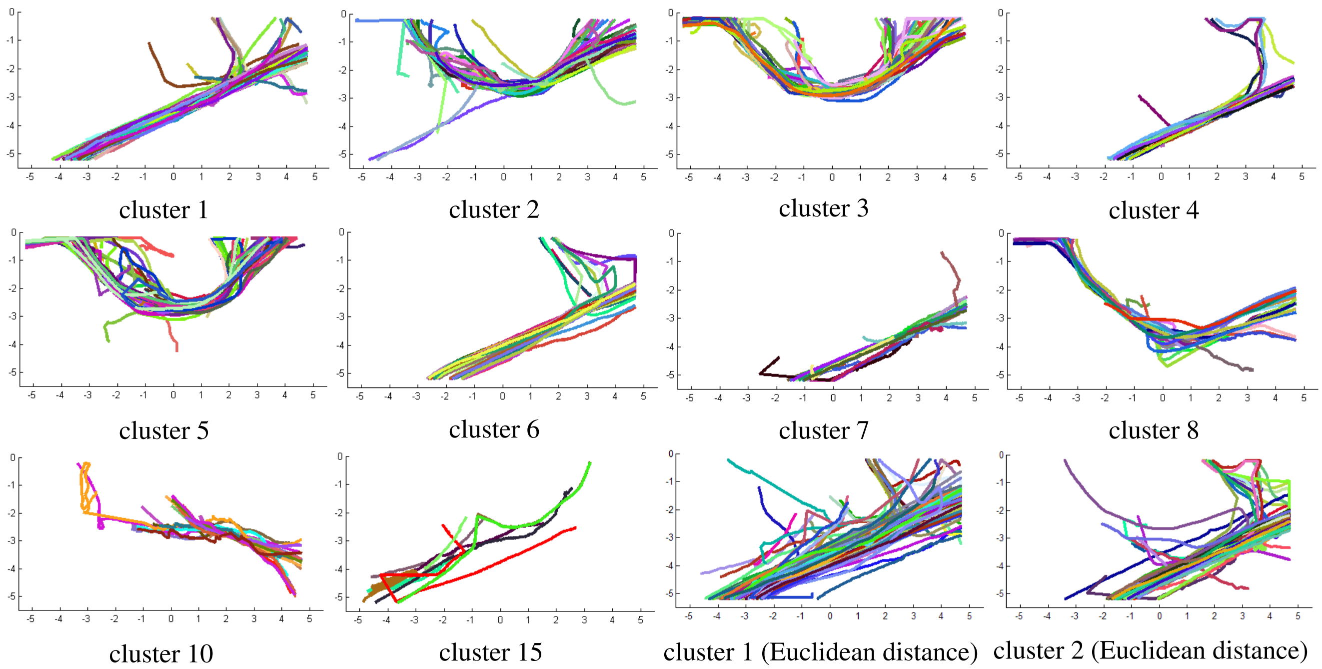

Figure 6. Clusters of trajectories. Random colors are used to distinguish individual trajectories. For comparison the last two sub-figures show some trajectory clusters of the result using Euclidean distance and spectral clustering [3].

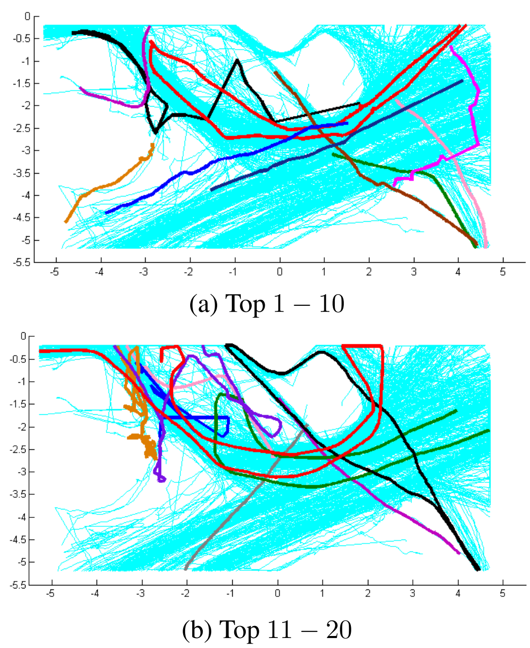

Figure 7. Top 20 abnormal trajectories are plotted in different colors. Other trajectories are plotted in cyan color.

diverge in different directions. For comparison, in the last two sub-figures of Figure 6 we also show two clusters of the result using Euclidean distance and spectral clustering [3] and setting the number of clusters as 16. Some fine structures of shipping fairways cannot be 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, the clustering performance significantly deteriorates.

In Figure 7, 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 7. (2) The trajectory fits more than one semantic region, but the combination of the semantic regions is uncommon. The red trajectory in Figure 7 (a), and the red and green trajectories in Figure 7 (b) are such examples.

8. Results on tracks from a parking lot

There are $N = 40,453$ trajectories in the parking lot data set collected over one week. Figure 1 plots $4,404$ trajectories from one day. Because of the large number of samples, similarity based clustering methods require 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. It is difficult for them to work on this large data set. The space complexity

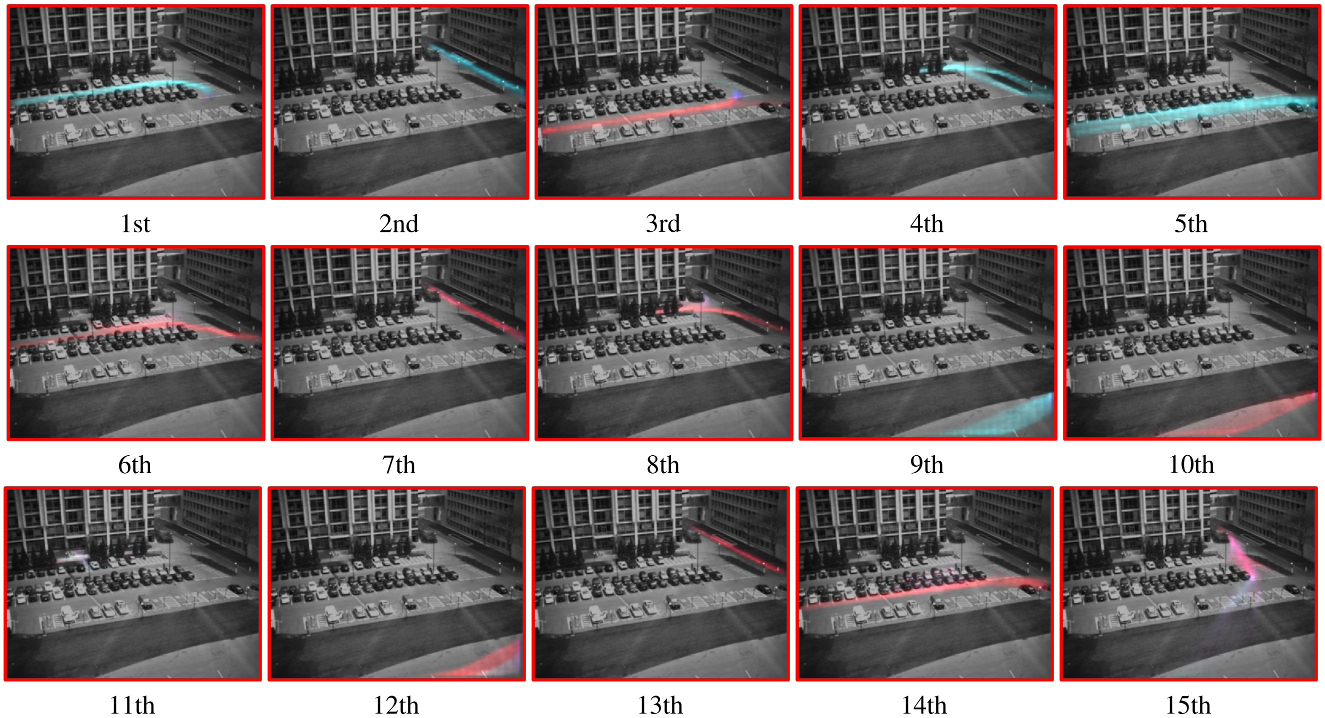

Figure 8. Some semantic regions learnt from the parking lot data set. The meaning of colors is the same as Figure 5.

of our nonparametric Bayesian approach is $O(N)$ instead of $O(N^2)$. The time complexity of each Gibbs sampling iteration is $O(N)$. It is difficult to provide theoretical analysis on the convergence of Gibbs sampling. However, there is some 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 CPU. On the parking lot data set, which is 70 times large than the radar data set in the number of trajectories, the likelihood curve converges after $6,000$ iterations. It takes around 6 hours. In our experiments, the time complexity of our approach is much smaller that $O(N^2)$

30 semantic regions and 22 clusters of trajectories are learnt from this data set. Some of them are shown in Figure 8 and 9. The first and third 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 regions, such as semantic regions 2 and 7, 9 and 12, 5 and 14. Similarly objects on trajectories (see Figure 9) in clusters 2 and 3, 5 and 11 are moving in opposite directions. Many outlier trajectories are in small clusters, such as clusters 20, 21 and 22. The top 100 abnormal trajectories are shown in Figure 10. Some horizontal trajectories on the grass field are detected as abnormalities. They were caused by a worker shearing the grass, which happened only once.

9. Conclusion

We propose a nonparametric Bayesian framework to cluster trajectories, learn the models of semantic regions, and detect trajectories related to abnormal activities. Different from most of the existing spatial distance based trajectory clustering approaches with ad hoc nature, we formulate these problem in a transparent probabilistic way. The number of semantic regions and clusters of trajectories are learnt through the hierarchical Dirichlet processes. The space complexity of our algorithm is $O(N)$.

10. Acknowledgment

The authors wish to acknowledge DSO National Laboratories (Singapore) for partially supporting this research.

References

[1] T. S. Ferguson. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230, 1973.

[2] J. Fernyhough, A. Cohn, and D. Hogg. Generation of semantic regions from image sequences. In *Proc. of ECCV*, 1996.

[3] Z. Fu, W. Hu, and T. Tan. Similarity based vehicle trajectory clustering and anomaly detection. In *Proc. of ICIP*, 2005.

[4] I. Junejo, O. Javed, and M. Shah. Multi feature path modeling for video surveillance. In *Proc. of ICPR*, 2004.

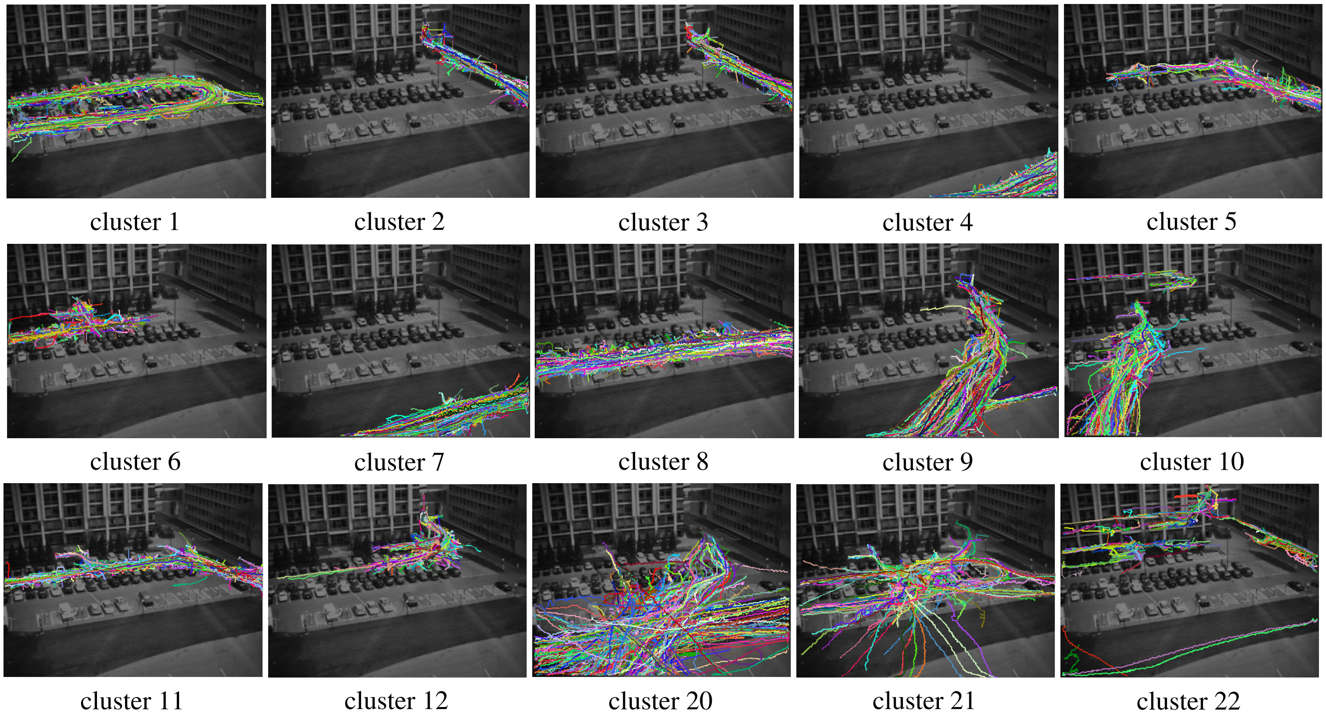

Figure 9. Some clusters of trajectories from the parking lot data set.

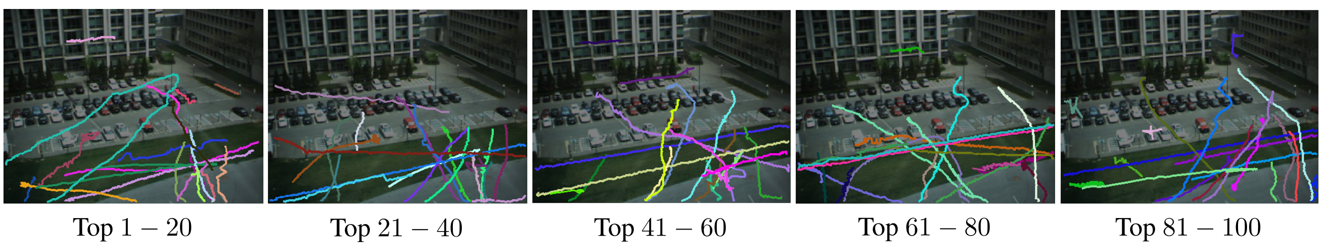

Figure 10. Top 100 abnormal trajectories in the parking lot data set.

[5] 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 CVPR*, 2005.

[6] E. Keogh and M. Pazzani. Scaling up dynamic time scaling up dynamic time. In *Proc. of ACM SIGKDD*, 2000.

[7] S. MacEachern, A. Kottas, and A. Gelfand. Spatial nonparametric bayesian models. Technical report, Institute of Statistics and Decision Sciences, Duke University, 2001.

[8] D. Makris and T. Ellis. Automatic learning of an activity-based semantic scene model. In *Proc. of AVSBS*, 2003.

[9] J. Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

[10] P. Smith, N. Lobo, and M. Shah. Temporalboost for event recognition. In *Proc. of ICCV*, 2005.

[11] C. Stauffer and E. Grimson. Learning patterns of activity using real-time tracking. In *IEEE Trans. on PAMI*, 2000.

[12] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical dirichlet process. *Journal of the American Statistical Association*, 2006.

[13] X. Wang, K. T. Ma, G. Ng, and E. Grimson. Trajectory analysis and semantic region modeling using a nonparametric bayesian model (http://hdl.handle.net/1721.1/40808). Technical report, Computer Science and Artificial Intelligence Lab, MIT, 2008.

[14] X. Wang, X. Ma, and E. Grimson. Unsupervised activity perception by hierarchical bayesian models. In *Proc. of CVPR*, 2007.

[15] X. Wang, K. Tieu, and E. Grimson. Learning semantic scene models by trajectory analysis. In *Proc. of ECCV*, 2006.

[16] T. Xiang and S. Gong. Beyond tracking: Modelling activity and understanding behaviour. *International Journal of Computer Vision*, 67:21–51, 2006.

[17] Z. Zhang, K. Huang, T. Tan, and L. Wang. Trajectory series analysis based event rule induction for visual surveillance. In *Proc. of CVPR*, 2007.

[18] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity. In *Proc. of CVPR*, 2004.