DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-381

Unsupervised tracking with a low computational cost using the doubly stochastic Dirichlet process mixture model

Xing Sun[1]*, Nelson H.C. Yung*[2]*, Edmund Y. Lam*[1]*, and Hayden K.-H. So*[1]

[1] *Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong*

[2] *Yung & Partners Limited, Hong Kong*

Abstract

This paper presents an unsupervised tracking algorithm with a low computational cost using the Temporal Doubly Stochastic Dirichlet Process (TDSDP) mixture model, and we demonstrate it in tracking fish in low quality videos for water quality assurance. The object is captured in the temporal domain with a global dependency prior instead of the Markov assumption, making it particularly suitable for long-term tracking. Furthermore, the TDSDP mixture model can calculate the number of object trajectories automatically. We describe how to construct this mixture model from thinning multiple Dirichlet Process Mixtures (DPMs) with conjugate priors, followed by details of the algorithm for object tracking. Experiments on a fish dataset illustrate that the TDSDP can track multiple fish, and performs well even when they are overlapping in the view. Further experiments also suggest that TDSDP can be applied to other tracking problems.

Introduction

Object detection and tracking play critical roles in intelligent surveillance systems. Specifically, in environmental protection and marine biology, the tracking, detection, and counting of fish are effective methods for monitoring the water-quality and studying fish ecology [1]. Erratic fish behavior, for instance, would indicate possible heavy metal poisoning.

The surveillance video for fish tracking is often of low quality, and this creates two major problems. First, the lack of sufficient features makes it challenging to distinguish multiple fish when they are overlapping. Second, manual preprocessing of fish tracking is expensive, including tasks such as labeling the fish position and setting the constraint of the tracking trajectory. For example, the number of fish in each frame cannot be assumed as known or fixed. Similar problems also exist in human and vehicle tracking applications. To overcome these issues, we turn to the powerful nonparametric Bayesian methods.

Nonparametric Bayesian models based on Dirichlet Process (DP) prior have been widely used to model topics over space and time [2, 3, 4, 5, 6, 7]. The term “topic” is employed in the machine learning area, which represents the characteristics of each cluster. For instance, the topic of fish in a single frame can be considered as the position center of that fish or other feature parameters (e.g. average color). These nonparametric Bayesian models relax the assumption that the topic number is fixed or known. Numerous approaches have been proposed to analyze topic trajectories via modeling dependency in topic space. While DDP models based on GP [7, 8] are too complex to be implemented to track fish, some other models [2, 6] have been proposed to bypass the GP priors. However, these models achieve dependency modeling under Markov assumption with a stricter assumption [2, 9], or discard modeling the location variation and model dependency via overlapping regions [6]. Several approaches, which model topics dependency under the non-Markov framework [3, 4, 5], are unable to obtain the position variation of the fish.

In our research, we seek to present an unsupervised algorithm, which can model temporal variations of the fish as well as location variations. Moreover, it makes sense to estimate the varying number of fish along all frames automatically. Furthermore, this algorithm needs to be compact and general, with a low computational cost and can be extended to various machine vision applications.

To achieve this, first, we extract interest points from the video clips based on the Harris detector [10], and the positions of these points constitute the observation data. Second, a TDSDP mixture model is proposed with a Sigmoid Gaussian Process (SGP) prior intensity. The SGP is employed to model the varying intensity of the trajectories along frame stamps. Subsequently, the mixture model of the TDSDP is constructed elegantly with a thinning procedure applied to multiple DPMMs. It allows us to handle massive transportation video data with a lower computational cost. Finally, the TDSDP mixture model enables us to track varying fish positions along different video frames with one global SGP intensity prior. Moreover, it allows us to capture location variations, birth and disappearance of objects based on dependency modeling in all topic domains. It uses the continuous probability measure sampled from the Global GP prior to fit the distribution of the variation of objects and the global GP prior enables us to achieve a robust result.

Overall, the robustness and low computational cost enable the proposed TDSDP mixture model to be used for the low-quality fish dataset. The global intensity modeled by the TDSDP mixture model robustly tracks fishes with only location feature. Low computational inference enables TDSDP to process many frames of video sequence efficiently.

Derivation of DSDP

Construction of DSDP mixture model

Dirichlet process mixture model (DPMM) [11] is a powerful Bayesian model based on Dirichlet Process prior. Here, we provide an elegant way to derive the mixture model of DSDP. A DSDP mixture model thinned from a single DPMM can be con-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-381

structed in [12]

$$\begin{aligned} Y &\sim GP(m,\kappa) \\ D|Y &\sim DP(\sigma(Y)\alpha^* H_0) \\ \theta_i|D,Y &\overset{iid}{\sim} D|Y \\ x_i|\theta_i &\overset{iid}{\sim} f(\theta_i), \end{aligned} \tag{1}$$

where Y is an intensity function sampled from sigmoid GP prior with parameter m and κ, thinned function $\sigma(Y)$ is the sigmoid function $\sigma(Y) = (1+e^{-Y})^{-1}$, α^* and H_0 denote the upper bound concentration parameter and the conjugate prior. Topic parameter θ_i is independently sampled from the conditional DSDP $D|Y$. Observation data x_i is sampled from distribution $f(\theta_i)$ parameterized by θ_i.

DSDP mixture model applied on multiple DPs

A DSDP mixture model thinned from multiple DPMMs has multiple conditional DSDP priors $D_t|Y_t$, their corresponding GP functions Y_t are sampled from global GP prior $GP(\cdot)$ at regions $t = 1,\ldots,T$ [13].

$$\begin{aligned} Y_t &\sim GP(m_t,\kappa_t) \\ D_t|Y_t &\sim DP(\sigma(Y_t)\alpha^* H_{0t}) \\ \theta_{ti}|D_t,Y_t &\overset{iid}{\sim} D_t|Y_t \\ x_{ti}|\theta_{ti} &\overset{iid}{\sim} f(\theta_{ti}), \end{aligned} \tag{2}$$

where topic parameter θ_{ti} at region t is independently sampled from its corresponding conditional DSDP $D_t|Y_t$. The conjugate prior H_{0t} is updated based on data $\{x_{ti}\}_i$ at this region, therefore it can vary along the regions. Globally sampled GP functions Y_t make the varying local parameters (α_t and H_t) retain the global dependency. They follow

$$H_t(\theta) = \frac{H_t^Y(\theta)}{H_t^Y(\Theta)}, \quad \text{and} \quad \alpha_t = \alpha^* H_t^Y(\Theta), \tag{3}$$

where $H_t^Y(\theta) \triangleq H_{0t}(\theta)\sigma(Y_t(\theta))$.

Fish detection algorithm

Based on the Temporal DSDP provided above, a novel fish detection approach is presented in this section. The fish topic inference algorithm follows.

- We first cluster the extracted potential fish interest points (feature) into fish topics frame by frame, this process is shown in top panels in the Fig. 2.
- Then the clustered fish topics are updated based on the clustering assignment and their Metropolis-Hastings acceptance ratio.
- Last, based on the newly updated topics, the hyper-parameters are inferred

Therefore, the fish detection algorithm mainly iterates three steps: the fish topic assignments sampling, the fish topic updating and the hyper-parameter sampling. The detailed information of this algorithm is shown in the Algorithm 1.

Fish topic assignment sampling

Sampling inference procedures for the Sigmoidal GP DSDP mixture model are illustrated as follows: First, make sure that all K clusters are non-empty. Then draw a new $z_i^{(t)}$ at t^{th} iteration according to the criteria in Eq.4. This sampling is developed from algorithm 2 in [14] for non-conjugate prior DPMM sampling.

$$P\left(z_i^{(t)} = k\right) = \begin{cases} \frac{\sigma(Y_k) n_{-i,k}^{(t-1)}}{n-1+\alpha^{(t-1)}} f_k(x_i), k \leqslant K \\ \frac{\sigma(Y_k)\alpha^{(t-1)}}{n-1+\alpha^{(t-1)}} f_k(x_i), k > K \end{cases}, \tag{4}$$

where $n_{-i,k}^{(t-1)}$ is the number of data clustered into k^{th} topic except x_i. The likelihood $f_k(x_i)$ is hard to be calculated directly with integral computation. Instead, $f_k(x_i)$ can be approximated in the large data limit as:

$$f_k(x_i) = \begin{cases} \int p\left(x_i, \theta_k | \vec{x}_{-i,k}\right) H_0\left(d\theta_k\right), k \leqslant K \\ \int p\left(x_i, \theta_k | \vec{x}_{-i,1:K}\right) H_0\left(d\theta_k\right), k > K \end{cases}, \tag{5}$$

where $x_{-i,k} \triangleq \left\{x_j : z_j = k, j \neq i\right\}$ and $\vec{x}_{-i,1:K} \triangleq \left\{x_{-i,k}\right\}_{k=1}^K$. In Eq. 4, $\alpha^{(t)} = \alpha^{*(t)} \int_\Omega \sigma\left(Y\left(d\theta\right)\right)$ is concentrate parameter at t^{th} iteration, where upper bound $\alpha^{*(t)} \sim p\left(\alpha^* | K^{(t)} + M^{(t)}, n, a, b\right)$ [15] and initial parameter follows $\alpha^* \sim Gamma(a,b)$. $K^{(t)}$ and $M^{(t)}$ are the number of topic variables and latent variables at t^{th} iteration, respectively. n is number of observation data, and normally a is set as $n/20$ and $b=1$. Detailed algorithm is shown in Algorithm 1.

Fish topics updating

The Metropolis-Hastings acceptance ratio for updating topic parameter θ_k to newly sampled variable θ^* follows:

$$a(\theta^*) = \frac{G_k\left(\theta^*\right)\left(1+\exp\left(Y(\theta_k)\right)\right)}{G_k\left(\theta_k\right)\left(1+\exp\left(Y(\theta^*)\right)\right)}, \tag{6}$$

where $G_k\left(\theta\right)$ is posterior distribution $p\left(\theta | x_{t,-i,k}\right)$ given data partition $x_{t,-i,k}$.

The Metropolis-Hastings acceptance ratio for updating latent variables $\theta_{m,i}^*$ to newly sampled variable θ^* follows:

$$a(\theta^*) = \frac{H_0\left(\theta^*\right)\left(1+\exp\left(-Y(\theta_k)\right)\right)}{H_0\left(\theta_k\right)\left(1+\exp\left(-Y(\theta^*)\right)\right)}, \tag{7}$$

where $H_0(\cdot)$ is the conjugate prior .

If the Metropolis-Hastings acceptance ratio $a(\theta^*)$ is bigger than 0.5, this newly sampled variable for both of topic and latent variable is accepted.

Sampling upper bound concentration parameter

The upper bound concentration parameter α^* is defined based on [15]:

$$\alpha^* \sim p\left(\alpha^* | K^{(t)} + M, n, a, b\right). \tag{8}$$

Its initial parameter follows $\alpha^* \sim Gamma(a,b)$. K and M are the number of topic variables and latent variables, respectively. n is the number of observation data, and normally a is set as $n/20$ and $b=1$. To make sure that the upper bound α^* is large enough, we safely set $a = n/2$.

DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-381

Algorithm 1 DSDP MCMC Sampling inference

Input: Observation data $x_{1:n}$
Output: Inferred topics $\theta_{1:K}$ and GP intensity $Y_{1:K+M}$
Sample conjugate prior H_0 and concentrate parameter α^* given data $x_{1:n}$
Sample a DP posterior $\theta_{1:K}^{(0)}$ based on initial upper bound intensity $(\alpha^* H_0)$ with initial assignments $z_{1:n}^{(0)}$
for $t = 0 : Time$ **do**
  Sample the latent variables $\theta^{(t)}{}_{K+1:K+M}$ and its amount $M^{(t)}$
  Sample the GP functions $Y_{1:K+M}^{(t)}$ and co-variance parameter $\eta^{(t)}$ via Hamiltonian Monte Carlo
  for $i \in \tau(1), ..., \tau(n)$ **do**
    for k=1:K+M **do**
      Sample a new topic θ_{ki}^*
      Sample new GP function $Y(\theta_{k,i}^*)$ at θ_{ki}^* based on current GP functions $Y_{1:K+M}^{(t)}$
      Calculate acceptance rate a_k via Eq. 6 and Eq. 7
      if $a_k > 1$ **then**
        $\theta_{k,i} = \theta_{k,i}^*$ and $Y(\theta_{k,i}) = Y(\theta_{k,i}^*)$
      else
        $\theta_{k,i} = \theta_k$ and $Y(\theta_{k,i}) = Y(\theta_k)$
      end if
      Calculate likelihood $\ell(x_i|\theta_{k,i})$ and data amount $n_{-i,k}$
    end for
    Sample assignment $\vec{z}_{1:n}^{(t+1)}$ via Eq. 4
  end for
  Sample new topics $\theta_{1:K}^{(t+1)}$ via Eq. 6 given new clusters $x_{1:K}^{(t+1)}$
  Sample upper bound concentration parameter α^* via Eq. 8
end for

Global fish trajectories modeling

In this section, we present an approach to track multiple fish trajectories without initial localization or detection. This approach based on the detected fish topics enables us to categorize fish flows, which follow the global SGP prior.

GP intensity clustering

This approach clusters the inferred SGP intensity with fish topics $\vec{\theta}_{1:K}$ into L fish trajectories. Furthermore, the number of fish trajectories L can be learned by following expression.

$$p(c_i = l) = \frac{g_{-i,l} \sum_{c_j = l} f(d_{ij})}{\gamma + \sum_{l=1}^{L} g_{-i,l} \sum_{c_j = l} f(d_{ij})}, \tag{9}$$

where i denotes topic index, L denotes the current number of non-empty topic trends, d_{ij} is the distance between topics θ_i and θ_j, c_i is the assignment for topic θ_i. And there is a probability that generates a new fish trajectory

$$p(c_i = L+1) = \frac{\gamma}{\gamma + \sum_{l=1}^{L} g_{-i,l} \sum_{c_j = l} f(d_{ij})}. \tag{10}$$

According to the distance-based Chinese Restaurant Process proposed in [16], $f(d_{ij})$ is window decay:

$$f(d_{ij}) = \frac{\exp(-d_{ij} + a)}{1 + \exp(-d_{ij} + a)} \delta(0 < d_{ij} < a). \tag{11}$$

Delay parameter a and delta function $\delta(\cdot)$ are defined in [16]. Eq. 11 indicate that only these customers with the distance less than a are taken into account. In Eq. 9, we use a kernel classifier, which was proposed in [17], to make L GP gating network independent with $g_{-i,l} \triangleq g\left(\{\theta_j\}_{c_j = l, j \neq i}\right)$ defined as below:

$$g_{-i,l} = \frac{\sum_{j \neq i} \kappa(\theta_i, \theta_j; \eta)\, \delta(c_j, l)}{\sum_{j \neq i} \kappa(\theta_i, \theta_j; \eta)}, \tag{12}$$

where $\kappa(\theta_i, \theta_j; \eta)$ is kernel function parameterized by η, and is defined based on [18]:

$$\kappa_d(\theta_d, \theta'_d) \triangleq \sigma_{d,f}^2 exp\left(-\frac{(\theta_d - \theta'_d)^2}{2h_d^2}\right) + \sigma_{d,n}^2 \delta_{ii'}, \tag{13}$$

where $\eta_d \triangleq \left\{\sigma_{d,f}^2, \sigma_{d,n}^2, h_d^2\right\}$ is d^{th} component of variance parameter η and $\delta_{ii'} \triangleq \delta(\theta, \theta')$. The model expressed in Eq. 11 indicates that the probability of topic θ_i assigned to the l^{th} cluster is proportional to two factors: the l^{th} cluster topic density in the range of a from topic θ_i, and kernel function calculated by topics $\{\theta_i\}_{c_i = l}$ in same cluster. Readers interested in object tracking based on GP clustering can refer to [19, 20].

Finally, the fish trajectories is obtained as shown in the last panel in Fig. 2. The global fish assignment v_i for the i^{th} fish is obtained by local topic assignment z_i and global assignment for topic c_k, the global fish assignment v_i follows $v_i = c_{z_i}$.

An example: an overlapping case in low-quality fish dataset

In this section, we use a video sequence from the low-quality fish dataset as an example to illustrate how the proposed TDSDP mixture model works. This video includes trajectory overlapping

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-381

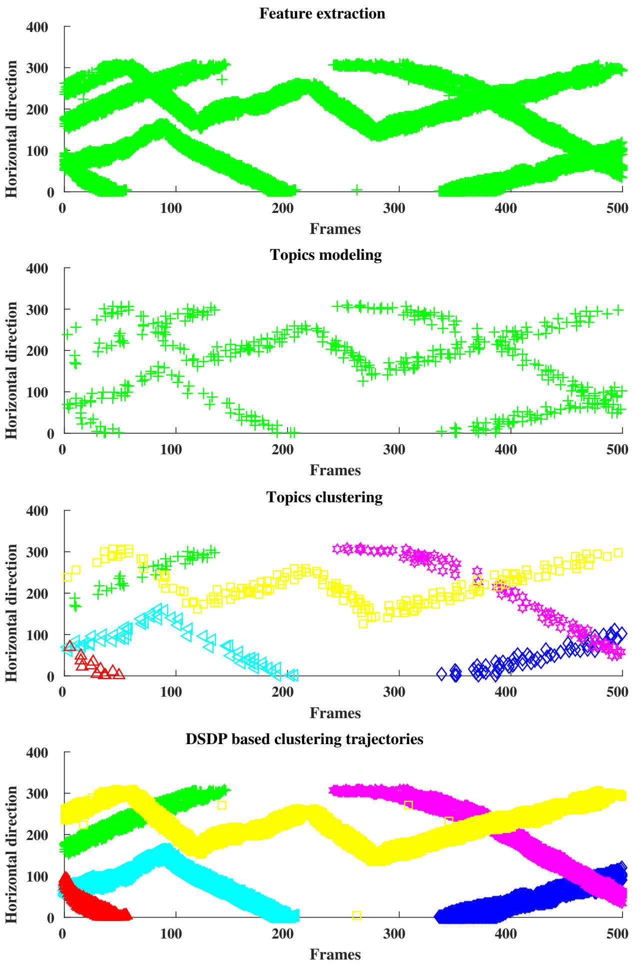

Figure 1. Fish tracking diagram

of fishes and fluctuate number of fishes, the example is shown in Fig. 2. Modeling the overlapping topic trajectories shows that the TDSDP is able to infer the SGP prior. Estimating the components in each mini-path or frames shows that the TDSDP is able to automatically determine the inhomogeneous concentration parameters in the temporal domain.

To begin with, interest points are extracted from the video based on the Harris detector proposed in [10] and positions of these points forms the observation data. Initial DPs can be applied on the observation data in frames or in mini patches which are obtained by clustering frames. In order to reduce the global GP computational cost $\left(\mathrm{O}\left(c^3k^3\right)\right)$ in the following experiments on video sequence, we cluster f frames into c mini-patches, in which k is the topics number extracted in each mini-patch. This clustering is achieved by DPMM, which is applied on time stamp of observation data. For this low-quality fish, $f/c \approx 15$. 500 frames are clustered into 50 patches in this example, and 1000 frames clustered into 100 patches for the whole video.

Experiment steps are illustrated with flow diagram in Fig. 1: 1) mini-patch decomposition at the observation level: observation points are presented in the top panels as observed mixture trajectories. 2) Then topic trajectories can be inferred by TDSDP is shown in the second panel. There are mainly two important characters sampled in this step: (a) Concentration parameter is thinned in the frame or temporal patch t: $\alpha(t) = \alpha_0(t)\int_{\Omega_t}\sigma(Y(\theta))d\theta$; and (b) Distribution of topic variant is sampled from the global SGP prior $H_t(\theta) = \left(\alpha_0(t)/\alpha(t)\right)\sigma(Y(\theta))H_{0t}$ at t can be modeled. Furthermore DP sampling in each patch with underlying SGP prior and GP prior sampling given whole global topics. 3) Topics trajectories clustering: the SGP prior clustering algorithm is used to cluster overlapping topic trajectories presented in the third panel. 4) Then the clustered overlapped trajectories are shown in the bottom panel in Fig. 1.

Fish tracking result comparison

	GPUDDP	TDSDP	TDSDP	TDSDP
α_0	/	$0.1n$	$0.5n$	n
SFDA	95.7%	97.3%	97.5%	97.1%
ATA	66.2%	94.4%	95.4%	95.2%

Experiments

Low-quality fish dataset

Figure 3 shows some video clips of the fish dataset, different color signs represent the detected clusters. This dataset is more complex than the video example show in the previous context in Fig. 1 and Fig. 2, there are more overlapping in this proposed dataset, such as topic trajectories of fishes are provided in the top panel of Fig. 4. Bottom panel shows the clustered fish trajectories. The total clustering accuracy for the DPMM is about 37.12%, and the proposed TDSDP mixture model is 97.93%. There are totally seven fish occurred in this video, some of them came in and came out, some of them wander over to the edge of the camera viewing angle. The accuracy of the number of fish in each frame is 75% for the DPMM and 94% for the TDSDP mixture model. Clustering accuracies of the TDSDP and the DPMM in each frame are quite similar (such as 98.47% and 96.5%). The table shows the different results of the proposed TDSDP and the Generalized Polya Urn based DDP tracking [21] on the overlapping low-quality fish dataset. Sequence Frame Detection Accuracy (SFDA) and Average Tracking Accuracy (ATA) [22] are utilized to quantify the experiment results. SFDA measures the detection performance in each frame, and ATA quantifies the performance of a tracking algorithm for detecting objects across frames. The result of the GPUDDP applied on this low-quality fish dataset is shown in the Table 1, which is obtained by tuning parameter values for some samples from this dataset. The consequences of the TDSDP with various initial hyper-parameters α_0, where n is the average data amount in each patch. The table indicates that at the frame level, TDSDP and DDP methods also have similar detection accuracy (such as 97.5% and 95.7% with the best results for SFDA). However, TDSDP has a far superior tracking performance compared to the GPUDDP in terms of ATA. This result further illustrates that the robust TDSDP with the global SGP prior fits the varying topics better than the Markov assumption based DDP models.

Text trends topic modeling

This sub-section presents another supporting experiment, which focus on text topic modeling based on the temporal DSDP

DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-381

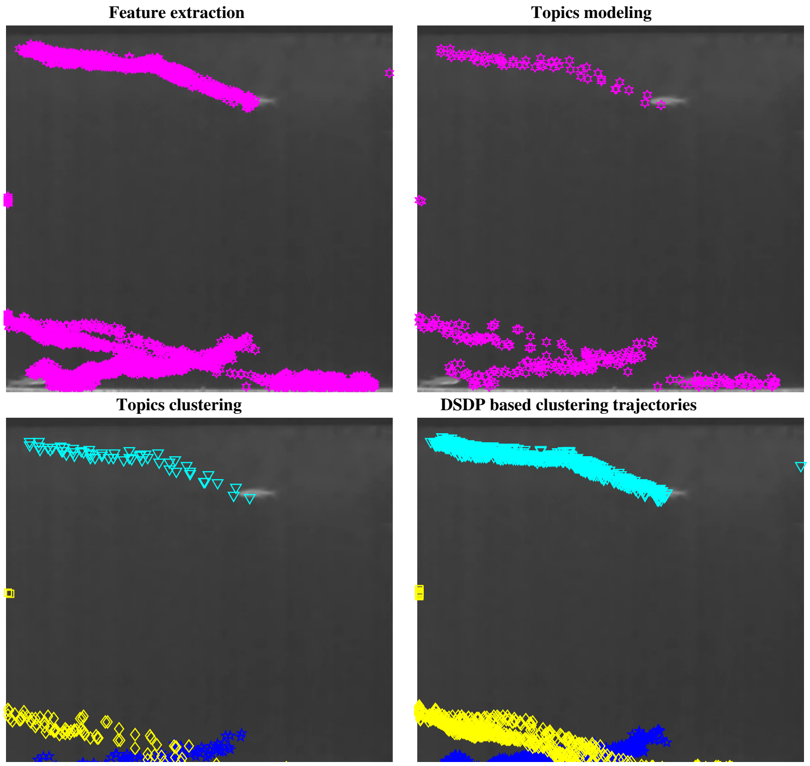

Figure 2. *Fish tracking example corresponding to diagram*

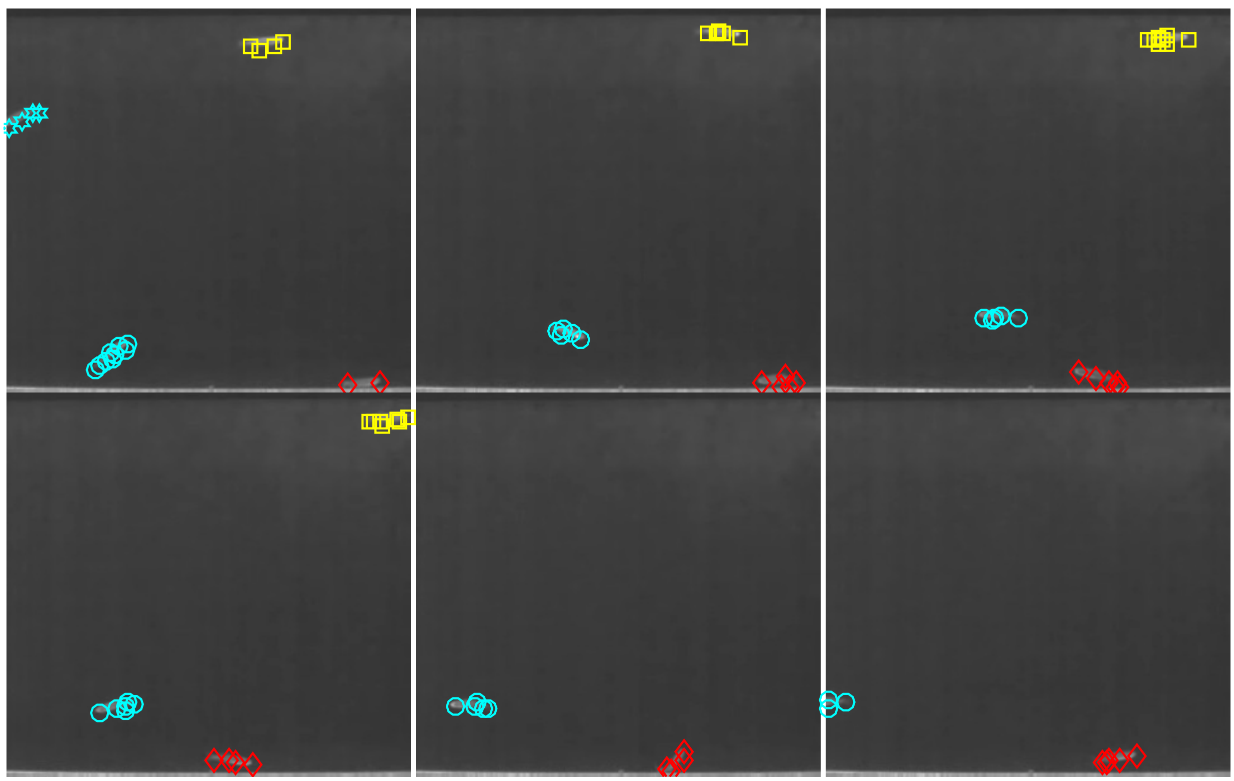

Figure 3. *Fish tracking examples*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-381

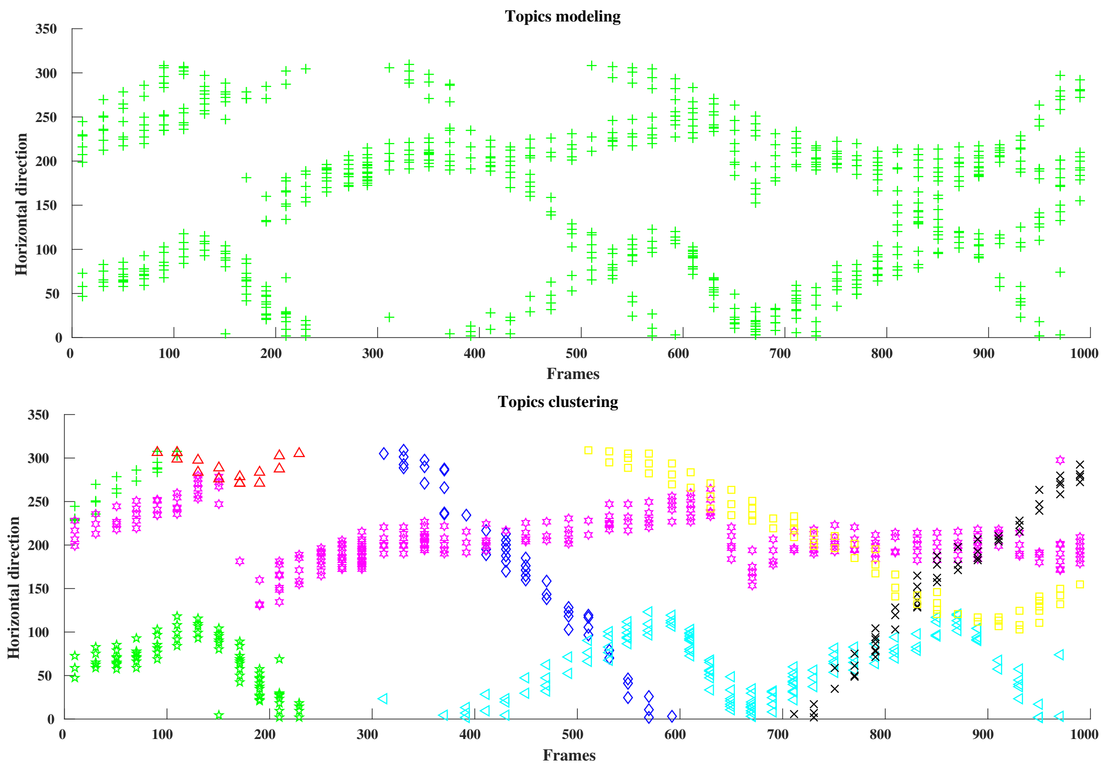

Figure 4. Topics clustering for fish dataset

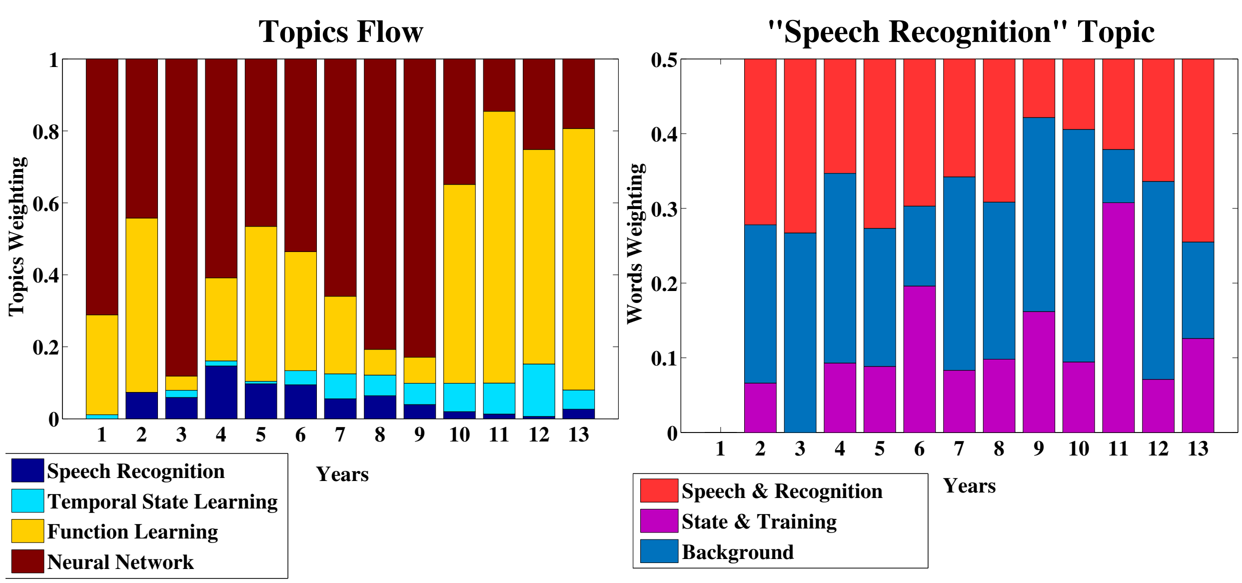

Figure 5. Topic flow for NIPS 0-12 dataset and Topic variation over years for "speech recognition" topic

DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-381

mixture model. The DSDP mixture model is applied on the text data with insufficient features, which is a 4-dimension feature vector obtained by Unsupervised Kernel Spectral Regression [23] over a number of years. NIPS 0-12 dataset [1] over 13 years shown in the left panel in Fig. 5 is applied. Four main paper topics vary along the years: "Speech Recognition", "Temporal State Learning", "Function Learning" and "Neural Network". Similar to the result shown in [3], "Neural Network" topic plays an important role in the first 9 years. Then it co-occurs with "Function Learning" topic in the middle 90's. Furthermore, "Speech Recognition" topic persists and then it gradually transfers into a more general topic: "Temporal State Learning". The right panel in Fig. 5 illustrates the topic parameters (words weighting) for "Speech Recognition" topic variations are varying over years, which enable better fitting of topic parameters over time compared with the Markov assumption based methods. Not surprisingly, "Speech Recognition" topic includes "Speech" & "Recognition", "State" & "Training" and background words.

Conclusions

A Temporal Doubly Stochastic Dirichlet Process mixture model is proposed for the unsupervised tracking algorithm. Global intensity prior of the proposed TDSDP model provides the ability to handle the fish tracking in the low quality video, which includes occlusion of the fish. Moreover, the TDSDP enables to model the trajectories of fish, whose number and positions are varying along temporal frames. The thinning procedure enables the TDSDP to reduce the computational cost significantly. Lastly, this algorithm can be extended to other surveillance applications, for instance the human and vehicle tracking.

Acknowledgments

This work was supported in part by the University Research Committee of the University of Hong Kong under projects 102009399 and 104003465.

References

[1] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. B. Fisher, "Detecting, tracking and counting fish in low quality unconstrained underwater videos.," *VISAPP (2)*, vol. 2008, pp. 514–519, 2008.
[2] D. Lin, E. Grimson, and J. W. Fisher, "Construction of Dependent Dirichlet processes based on Poisson processes," in *Advances in Neural Information Processing Systems*, pp. 1396–1404, 2010.
[3] X. Wang and A. McCallum, "Topics over time: a non-Markov continuous-time model of topical trends," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 424–433, ACM, 2006.
[4] D. D. Walker, K. Seppi, and E. K. Ringger, "Topics over nonparametric time: A supervised topic model using Bayesian nonparametric density estimation," in *UAI Applications Workshop*, 2012.
[5] A. Dubey, A. Hefny, S. Williamson, and E. P. Xing, "A nonparametric mixture model for topic modeling over time," in *SIAM International Conference on Data Mining*, pp. 530–538, SIAM, 2013.
[6] V. Rao and Y. W. Teh, "Spatial normalized Gamma processes," in *Advances in Neural Information Processing Systems*, pp. 1554–1562, 2009.
[7] A. E. Gelfand, A. Kottas, and S. N. MacEachern, "Bayesian nonparametric spatial modeling with Dirichlet process mixing," *Journal of the American Statistical Association*, vol. 100, no. 471, pp. 1021–1035, 2005.
[8] S. N. MacEachern, "Dependent Dirichlet processes," *technical report, Department of Statistics, The Ohio State University*, 2000.
[9] C. Guo, S. Mita, and D. McAllester, "Robust road detection and tracking in challenging scenarios based on markov random fields with unsupervised learning," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 3, pp. 1338–1354, 2012.
[10] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 65–72, IEEE, 2005.
[11] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, pp. 209–230, 1973.
[12] X. Sun, N. H. C. Yung, E. Y. Lam, and H. K. So, "Computationally efficient hyperspectral data learning based on the doubly stochastic dirichlet process," *Submitted*.
[13] X. Sun, N. H. C. Yung, and E. Y. Lam, "A novel unsupervised tracking method with the doubly stochastic dirichlet process mixture model," *Submitted*.
[14] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 249–265, 2000.
[15] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 577–588, 1995.
[16] D. M. Blei and P. I. Frazier, "Distance dependent chinese restaurant processes," *The Journal of Machine Learning Research*, vol. 12, pp. 2461–2488, 2011.
[17] C. E. Rasmussen and Z. Ghahramani, "Infinite mixtures of Gaussian process experts," *Advances in Neural Information Processing Systems*, vol. 2, pp. 881–888, 2002.
[18] C. E. Rasmussen and C. Williams, "Gaussian processes for machine learning. 2006," *The MIT Press, Cambridge, MA, USA*, vol. 38, pp. 715–719, 2006.
[19] J. Ross and J. Dy, "Nonparametric mixture of Gaussian processes with constraints," in *Proceedings of the 30th International Conference on Machine Learning*, pp. 1346–1354, 2013.
[20] M. Lázaro-Gredilla, S. Van Vaerenbergh, and N. D. Lawrence, "Overlapping mixtures of Gaussian processes for the data association problem," *Pattern Recognition*, vol. 45, no. 4, pp. 1386–1395, 2012.
[21] W. Neiswanger, F. Wood, and E. Xing, "The dependent Dirichlet process mixture of objects for detection-free tracking and object modeling," in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, pp. 660–668, 2014.
[22] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and

[1] http://www.stats.ox.ac.uk/ teh/data.html.

DOI: 10.2352/ISSN.2470-1173.2016.14.IPMVA-381

J. Zhang, “Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 319–336, 2009.

[23] D. Cai, X. He, W. V. Zhang, and J. Han, “Regularized locality preserving indexing via spectral regression,” in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pp. 741–750, ACM, 2007.