Globally Continuous and Non-Markovian Crowd Activity Analysis from Videos

He Wang & Carol O'Sullivan

h.e.wang@leeds.ac.uk & Carol.OSullivan@scss.tcd.ie

Disney Research LA, US & University of Leeds, UK & Trinity College Dublin, Ireland

Abstract

Automatically recognizing activities in video is a classic problem in vision and helps to understand behaviours, describe scenes and detect anomalies. We propose an unsupervised method for such purposes. Given video data, we discover recurring activity patterns that appear, peak, wane and disappear over time. By using non-parametric Bayesian methods, we learn coupled spatial and temporal patterns with minimum prior knowledge. To model the temporal changes of patterns, previous works compute Markovian progressions or locally continuous motifs whereas we model time in a globally continuous and non-Markovian way. Visually, the patterns depict flows of major activities. Temporally, each pattern has its own unique appearance-disappearance cycles. To compute compact pattern representations, we also propose a hybrid sampling method. By combining these patterns with detailed environment information, we interpret the semantics of activities and report anomalies. Also, our method fits data better and detects anomalies that were difficult to detect previously.

Introduction

Understanding crowd activities from videos has been a goal in many areas. The main problem is essentially mining recurrent patterns over time from video data, shown in Fig. 1.

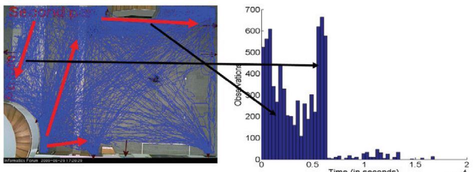

Figure 1: Left: trajectories. Right: Obervation number over time. Crowd activities should be captured by flows and their temporal profiles.

Previous methods only consider either space or with time but as **Markovian progressions** or **local motifs**. We propose a Spatio-temporal Hierarchical Dirichlet Process (STHDP) model (Fig. 2).

1. We present an **unsupervised** method (a **topic model**) for activity analysis that requires no prior knowledge about the crowd dynamics, user labeling or predefined pattern numbers.
2. Compared to static HDP variants, we explicitly model the **time-varying presence** of activity patterns.
3. Complementary to other dynamic HDP variants, we model time in a **globally continuous** and **non-Markovian** way, which provides a new perspective for temporal analysis of activities.
4. We also propose a non-trivial split-merge strategy combined with Gibbs sampling to make the patterns more compact.

Methodology

We compute activities as flows consists of bundles of similar trajectories. To feed trajectories into our topic model, we first convert video data into text data. By discretizing the space and velocity domain into grids, We feed position-velocity pairs (w_{d_n}) and time stamps (t_{d_n}) into STHDP (Fig. 2 Left)

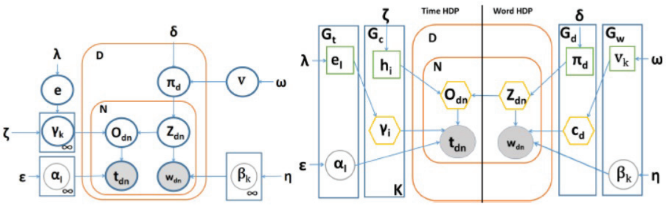

Figure 2: Left: STHDP Model. Right: Stick-breaking representation used for sampling.

We compute the posterior for which we proposed a hybrid sampling scheme combining Gibbs sampling and Metropolis-Hasting sampling (Fig. 2 Right).

Experimental Results

We tested our method on three widely used datasets: Forum [2], Carpark [3] and TrainStation [4]:

Activity Computation

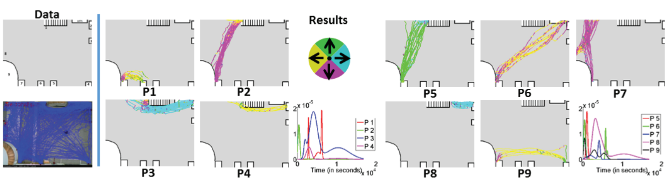

Figure 3: Top Left: Environment of Forum. Bottom Left: Trajectories overlayed on the environment. Right: Some activities shown by representative trajectories and their respective time activities. Colors indicate orientations described by the legend in the middle.

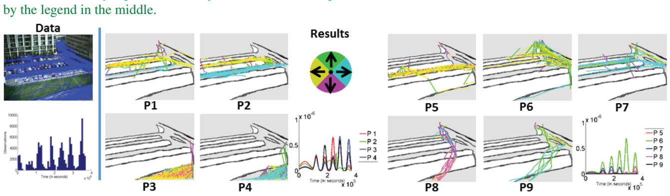

Figure 4: Top Left: Environment of the car park. Bottom Left: Observation numbers over time.Right: Some activities shown by representative trajectories and their respective time activities. Colors indicate orientations described by the legend in the middle.

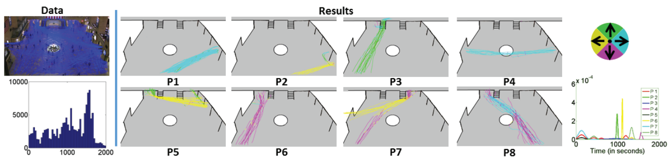

Figure 5: Top Left: Environment of the New York Central Terminal. Bottom Left: Observation numbers over time. Right: Some activities shown by representative trajectories and their respective time activities. Colors indicate orientations described by the legend on the right.

Anomaly Detection

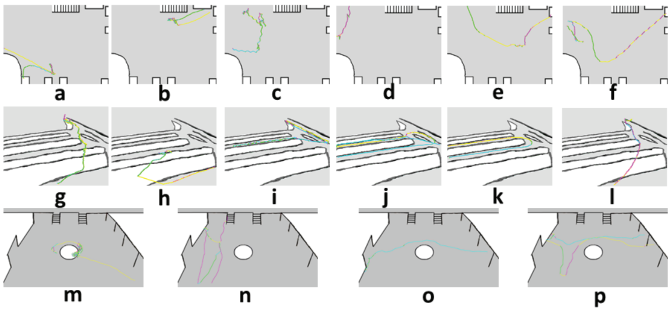

Figure 6: Top anomalies from three datasets in three rows respectively.

Comparison

By comparison, our method provides better likelihoods (Tab. 1 Left) and more in alignment with human judgements (Tab.1 Right). $r_{correct}$ and $r_{complete}$ are computed against human labelling.

	STHDP	DHDP[3]	MOTIF[1]
Forum	**-3.84**	-9.8	-54.38
Carpark	**-2.8**	-7.75	-62.13
TrainStation	**-3.5**	-4.9	-62.2

$r_{correct}$/$r_{complete}$	STHDP	DHDP[3]	MOTIF[1]
Forum	0.92/**0.88**	0.95/0.63	0.87/0.78
Carpark	0.83/**0.9**	0.89/0.31	0.85/0.42
TrainStation	**0.84/0.75**	0.72/0.55	0.69/0.58

Table 1: Left: Best per-word log likelihoods. Right: $r_{correct}$ and $r_{complete}$ accuracies from 0-1, 1 is the best. $r_{correct}$ is the clustering accuracy and $r_{complete}$ is the clustering completeness.

Anomaly Not Detected By Existing Methods

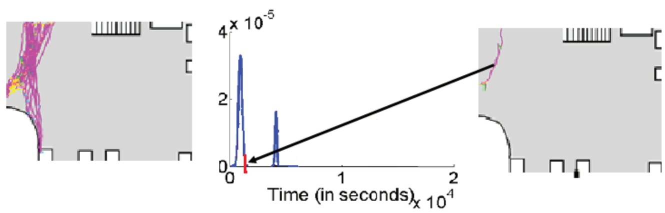

Figure 7: Left: A pattern detected. Middle: The temporal profile of the pattern. Right:An anomaly detected. Note spatially it is not an anomaly but temporally it is, which is difficult to detect for existing methods.

References

[1] R. Emonet, J. Varadarajan, and J. Odobez. Extracting and locating temporal motifs in video scenes using a hierarchical non parametric Bayesian model. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3233–3240, June 2011.

[2] B. Majecka. *Statistical models of pedestrian behaviour in the Forum*. MSc Dissertation, School of Informatics, University of Edinburgh, Edinburgh, 2009.

[3] Xiaogang Wang, Keng Teck Ma, Gee-Wah Ng, and W. Eric L. Grimson. Trajectory Analysis and Semantic Region Modeling Using Nonparametric Hierarchical Bayesian Models. *Int J Comput Vis*, 95(3):287–312, May 2011.

[4] Shuai Yi, Hongsheng Li, and Xiaogang Wang. Understanding pedestrian behaviors from stationary crowd groups. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3488–3496, June 2015.