Unsupervised Detection and Tracking of Multiple Objects with Dependent Dirichlet Process Mixtures

Willie Neiswanger,[1] Frank Wood[2]
[1]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY
[2]Department of Statistics, Columbia University, New York, NY

Introduction

This work proposes a multiple object detection and tracking method intended to reduce the need for explicit target identification and serve as a general method to track targets with arbitrary characteristics moving over diverse backgrounds. The technique uses a dependent Dirichlet process (DDP) mixture known as the Generalized Polya Urn (GPUDDP) [1] to model spatial and color data that can be extracted from videos containing multiple moving targets via frame differencing.

Data

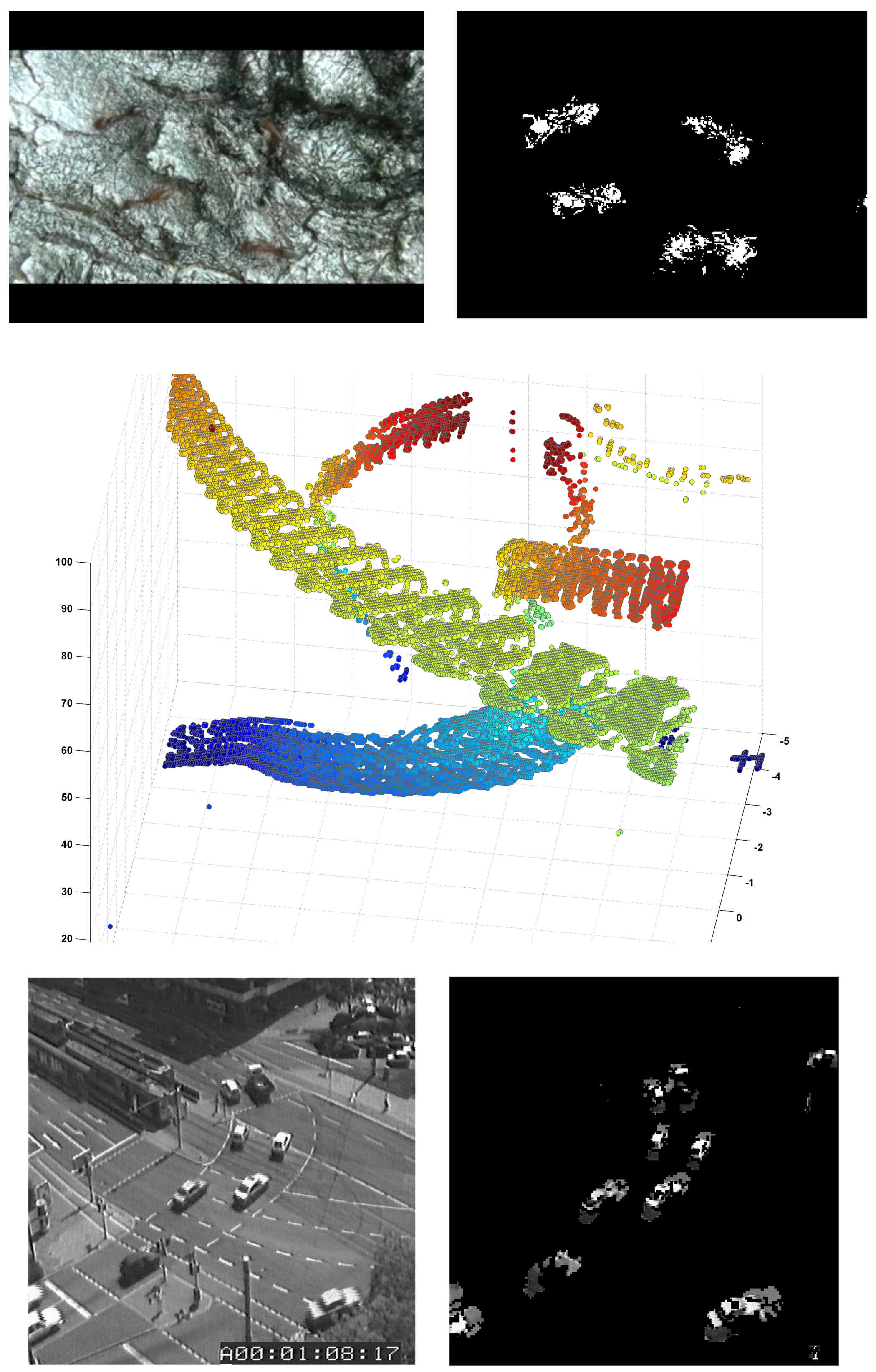


Figure 1. Video frames and corresponding frame differences showing motion pixels.

[1] F. Caron, M. Davy, and A. Doucet. Generalized polya urn for time-varying Dirichlet process mixtures. In *23rd Conference on Uncertainty in Artificial Intelligence (UAI'2007), Vancouver, Canada, July 2007*, 2007.

[2] J. Gasthaus, F. Wood, D. Görür, and Y. W. Teh. Dependent Dirichlet process spike sorting. In *Advances in Neural Informations Processing Systems 22*, 2008.

[3] Jan Gasthaus. Spike sorting using time-varying Dirichlet process mixture models, 2008.

Model

Observations: $\mathbf{x} = (\mathbf{x}^s, \mathbf{x}^c, t) = (x^{s_1}, x^{s_2}, x^{c_1}, \ldots, x^{c_V}, t)$

Likelihood: $P(\mathbf{x}|\theta) = \mathcal{N}(\mathbf{x}^s|\boldsymbol{\mu}, \Sigma)\mathcal{M}n(\mathbf{x}^c|\mathbf{p})$

Prior: $\mathbb{G}_0(\theta) = \mathcal{N}i\mathcal{W}(\boldsymbol{\mu}, \Sigma|\boldsymbol{\mu}_0, k_0, v_0, \Lambda_0)\mathcal{D}ir(\mathbf{p}|\mathbf{q}_0)$

where $\mathbf{x}^s$ is the location within a frame, $\mathbf{x}^c$ is a vector of color counts, and t is the time index, for a given pixel extracted via frame differencing.

Dependent Dirichlet Process Mixture

The GPUDDP can be written generatively as

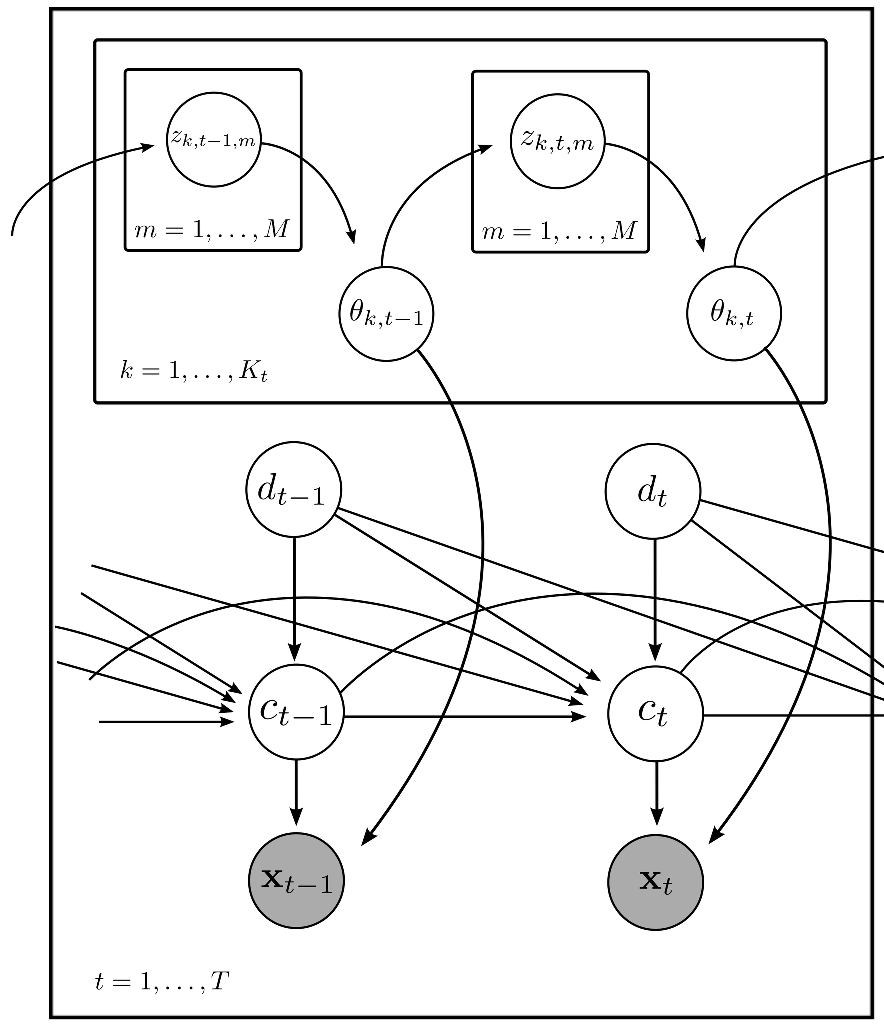
$$\mathbf{x}_i|c_i, \theta_{c_i} \sim \mathcal{N}(\boldsymbol{\mu}_{c_i}, \Sigma_{c_i})\mathcal{M}n(\mathbf{p}_{c_i})$$

$$c_i|c_1, \ldots, c_{i-1}, d_1, \ldots, d_i, \alpha \sim \mathrm{CRP}(c_1, \ldots, c_{i-1}, d_1, \ldots, d_i, \alpha)$$

$$d_i - i - 1|\rho \sim \mathrm{Geo}(\rho)$$

$$\theta_{k,t}|\theta_{k,t-1}, \mu_0, \kappa_0, \Lambda_0, \nu_0 \sim \begin{cases} \mathrm{TK} & \text{if } k \in \{1, \ldots, K_{t-1}\} \\ \mathbb{G}_0(\mu_0, \kappa_0, \Lambda_0, \nu_0) & \text{if } k = K_{t-1} + 1 \end{cases}$$

where TK represents a transition kernel that dicatates the dependence of cluster parameters on cluster parameters in adjacent time steps. It is chosen such that the model is marginally a Dirichlet process at each time step.



where $\theta_{k,t}$ denotes the parameters of mixture component k at time step t, $z_{k,t,m}$ denotes one of m auxiliary variables which plays a part in the transition kernel, c_t denotes the assignments of the observations $\mathbf{x}_t$ at time t, and d_t denotes the deletion time at which observations are removed from their assignments. There are an infinite number of mixture components, of which only a finite number are associated with data at any time step t.

Figure 2. A graphical model of the GPUDDP.

Results

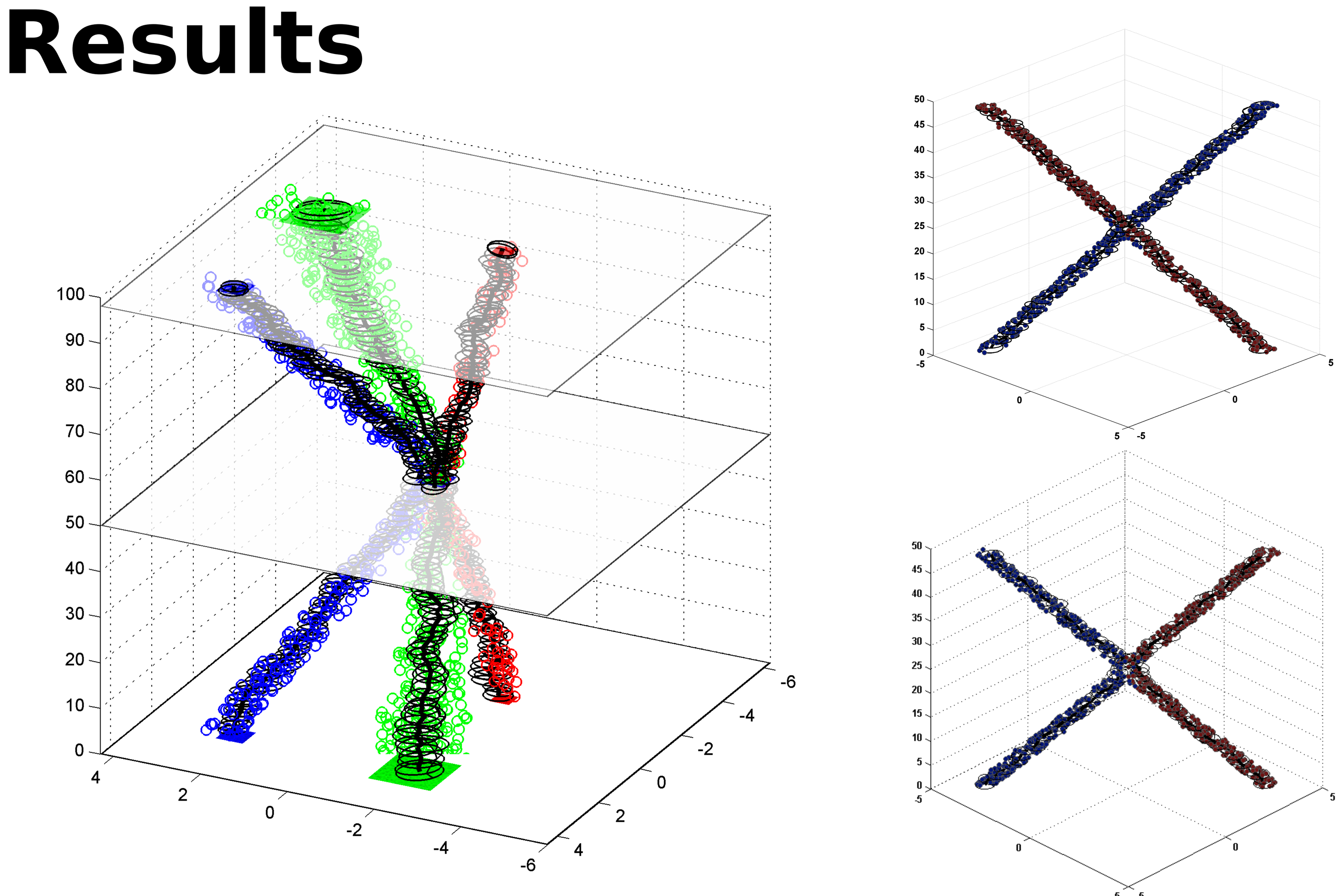


Figure 3. Results from synthetic videos of moving colored squares. Results demonstrate successful tracking through occlusions.

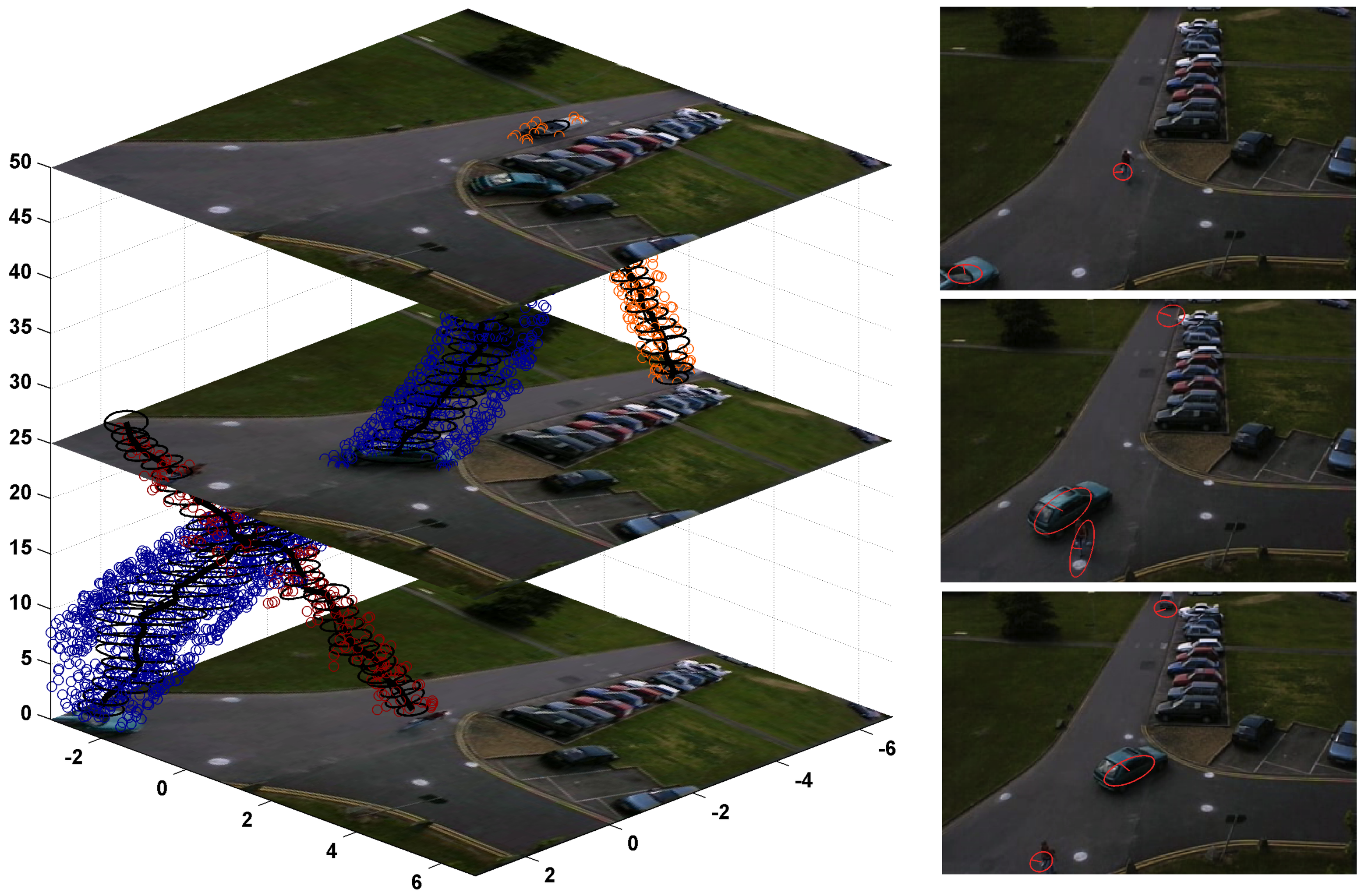


Figure 4. Results on a benchmark video data from the PETS2001 workshop.

Conclusion

We have demonstrated detection and tracking of multiple objects through occlusions, appearances / disappearances of objects, and over diverse backgrounds. Future work will involve modifying the prior distribution for more sophisticated appearance modeling and object detection.