

# Video Behaviour Mining Using a Dynamic Topic Model

Timothy Hospedales · Shaogang Gong · Tao Xiang

Received: 21 August 2009 / Accepted: 22 November 2011 / Published online: 8 December 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** This paper addresses the problem of fully automated mining of public space video data, a highly desirable capability under contemporary commercial and security considerations. This task is especially challenging due to the complexity of the object behaviors to be profiled, the difficulty of analysis under the visual occlusions and ambiguities common in public space video, and the computational challenge of doing so in real-time. We address these issues by introducing a new dynamic topic model, termed a Markov Clustering Topic Model (MCTM). The MCTM builds on existing dynamic Bayesian network models and Bayesian topic models, and overcomes their drawbacks on sensitivity, robustness and efficiency. Specifically, our model profiles complex dynamic scenes by robustly clustering visual events into activities and these activities into global behaviours with temporal dynamics. A Gibbs sampler is derived for offline learning with unlabeled training data and a new approximation to online Bayesian inference is formulated to enable dynamic scene understanding and behaviour mining in new video data online in real-time. The strength of this model is demonstrated by unsupervised learning of dynamic scene models for four complex and crowded public scenes, and successful mining of behaviors and detection of salient events in each.

**Keywords** Behaviour profiling · Video behaviour mining · Topic models · Learning for vision · Bayesian methods · Probabilistic modelling

## 1 Introduction

The proliferation of cameras in modern society is producing an ever increasing volume of video data which is thus far only weakly and inefficiently exploited. Ideally, users would like the ability to mine large volumes of recorded data to extract useful information about behaviour patterns of individuals and groups in the area under surveillance; and to monitor the scene for saliency in real-time in order to provide the potential for immediate response. Learning spatio-temporal behaviour patterns from videos of a public space is frequently of intrinsic commercial or security interest for users to gain more knowledge about activity patterns in public spaces which they are responsible for. For instance, retailers may be interested in shoppers browsing habits, while managers of public infrastructure sites might be interested in understanding typical behaviors. The learned behavior patterns may then be exploited for online analysis of the site state and for detection of salient activity pattern requiring further investigation. The alternative is the expensive and laborious manual analysis of the data and customization of detection software—which is prohibitive for many installations.

In practice, large volumes of recorded video data are frequently only stored passively for record purposes because of the challenges in developing such automatic and robust analysis methods. There are in general three challenges: dealing with the variety of potentially interesting behaviors, achieving sufficient robustness for practical use and real-time online operation.

---

T. Hospedales (✉) · S. Gong · T. Xiang  
School of Electronic Engineering and Computer Science, Queen  
Mary University of London, London E1 4NS, UK  
e-mail: [tmh@eecs.qmul.ac.uk](mailto:tmh@eecs.qmul.ac.uk)

S. Gong  
e-mail: [sgg@eecs.qmul.ac.uk](mailto:sgg@eecs.qmul.ac.uk)

T. Xiang  
e-mail: [txiang@eecs.qmul.ac.uk](mailto:txiang@eecs.qmul.ac.uk)