Modelling Multi-object Activity by Gaussian Processes

Chen Change Loy
ccloy@dcs.qmul.ac.uk
Tao Xiang
txiang@dcs.qmul.ac.uk
Shaogang Gong
sgg@dcs.qmul.ac.uk

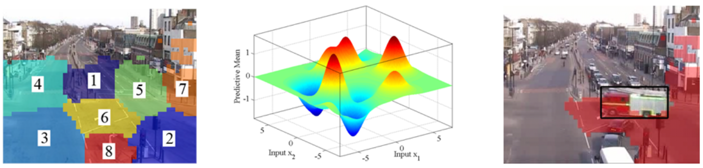
School of EECS
Queen Mary University of London
London E1 4NS, UK

Problem - This paper aims to address the problem of modelling multiple object activity captured in surveillance videos for the application of anomaly detection.

Related work - Most existing approaches [1, 5, 6] have been devoted to parametric models such as Dynamic Bayesian Networks (DBNs). In the context of complex multi-object activity modelling, learning a DBN structure with *appropriate complexity* (i.e. the number of hidden states, the state connectivity, and model topology) remains a difficult problem: (1) Automatic model selection criteria are inaccurate given sparse and noisy training data. (2) Specifying a model structure based on prior knowledge is challenging with surveillance video data as the activity states and dynamics are often not apparent and nor well defined. They also change over time. (3) Once the model complexity is fixed, its expressive power is hampered/limited by the initial model structure. Adjusting model structure complexity on-line is nontrivial for a DBN that requires re-learning new structure and re-estimating model parameters over time.

Our solution - In this paper, we present a new approach for activity modelling and anomaly detection based on non-parametric Gaussian Process (GP) models [4]. Our approach has the following advantages compared to the commonly deployed DBNs: (1) The use of a flexible, non-parametric model alleviates the difficult problem of selecting appropriate model complexity. (2) Our models need fewer parameters. Therefore they are less prone to overfitting problem. (3) Our models are able to cope with noise explicitly, resulting in superior robustness against the inevitable noise in activity representation.

We observed that a complex wide-area scene naturally consists of a set of semantic regions; each of the regions encapsulates different activity patterns which are correlated with each other either explicitly or implicitly. Our approach aims to discover these semantic regions and model non-linear relationships among activity patterns observed from the regions using GP. The understanding of these relationships is crucial in facilitating the detection of subtle anomalies that involve a group of objects, which are hard to detect by observing individual object alone.



(a) Semantic scene decomposition (b) Activity modelling using Gaussian Process models (c) Activity prediction and anomaly detection

Figure 1: Approach overview.

Approach overview - A method similar to [2] is employ to decompose a complex scene into N regions (see Fig. 1(a)) automatically according to the spatial-temporal distribution of activity patterns observed in a training set of video sequences. We then extract horizontal/vertical components of optical flow from each region over time and represent them as time series, $\hat{\mathbf{u}}_n$ and $\hat{\mathbf{v}}_n$. A detailed account on activity representation is given in the paper.

GP regression models are constructed for each region to model features $\hat{\mathbf{u}}_n$ and $\hat{\mathbf{v}}_n$ separately. Each model predicts the activity pattern from each region in the next time interval using activity patterns in other regions observed at present. A GP regression model is formally defined as $y = f(\mathbf{x}) + \varepsilon$, where $\mathbf{x}$ denotes an input vector at $t-1$ and y denotes a one-dimensional scalar output at t. Function $f(\mathbf{x})$ is a GP specified by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x},\mathbf{x}')$. The noise in the data is modelled explicitly in $\varepsilon \sim \mathcal{N}(0,\sigma_n^2)$, an independent Gaussian white noise with variance σ_n^2.

As our objective is to model relationships among activity patterns from different regions, we consider a squared exponential covariance function that implements Automatic Relevance Determination (ARD) [3] since it is capable of capturing the strength of influence among regions of a busy scene:

$$k_{\mathrm{SE}}(\mathbf{x},\mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}\left(\mathbf{x}-\mathbf{x}'\right)^\top \Sigma \left(\mathbf{x}-\mathbf{x}'\right)\right), \quad (1)$$

We also employ a neural network covariance function which allows a model to adapt to functions whose smoothness changes with the inputs:

$$k_{\mathrm{NN}}(\mathbf{x},\mathbf{x}') = \sigma_f^2 \sin^{-1}\left(\frac{2\tilde{\mathbf{x}}^\top \Sigma \tilde{\mathbf{x}}'}{\sqrt{\left(1+2\tilde{\mathbf{x}}^\top \Sigma \tilde{\mathbf{x}}\right)\left(1+2\tilde{\mathbf{x}}^\top \Sigma \tilde{\mathbf{x}}\right)}}\right), \quad (2)$$

where σ_f defines the magnitude and Σ encodes the relationships among activity patterns from different regions. To train the models, we estimate the hyper-parameters of a covariance function by maximising its marginal likelihood given the training data.

With the learned models, a novel one-step ahead prediction strategy is formulated to detect subtle anomalies. In particular, given a test vector $\mathbf{x}_*$ that consists of the past observations at $t-1$ from $N-1$ regions $\{r_j\}$, where $j = 1,\ldots,N, j \neq i$, the one-step ahead predictive distribution of region r_i at t is computed as

$$\begin{aligned} \overline{f}(\mathbf{x}_*) &= \mathbf{k}_*^\top \left(K + \sigma_n^2 I\right)^{-1} \mathbf{y}, \\ \mathbb{V}(f_*) &= k(\mathbf{x}_*,\mathbf{x}_*) - \mathbf{k}_*^\top \left(K + \sigma_n^2 I\right)^{-1} \mathbf{k}_*, \end{aligned} \quad (3)$$

where $\overline{f}(\mathbf{x}_*)$ is the mean and $\mathbb{V}(f_*)$ is the variance of the predictive distribution (*i.e.* uncertainty on prediction), whilst $\mathbf{k}_*$ denotes the vector of covariance between the test vector and the M training cases.

Anomaly is detected if the actual observation deviates largely from the predictive distribution (see Eqn. 3), which indicates that the learned relationship between different activity patterns is broken (see Fig. 2 for example). We compared two types of anomaly score for measuring the normality deviation, namely squared residual and predictive log-likelihood, with the latter taking the predictive uncertainty into account. Detailed explanation on anomaly detection is given in the paper.

The proposed approach is evaluated using a challenging public traffic scene featured complex multi-object interactions, activity patterns with changing complexity and noisy observations. Experimental results show that our GP models outperform DBNs for activity modelling and anomaly detection on sensitivity to anomaly, noise robustness and flexibility in learning from scarce training data.

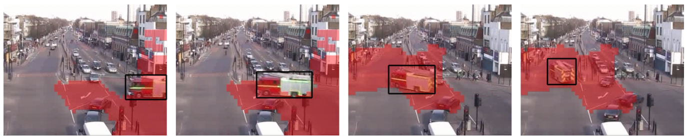


Figure 2: A fire engine (highlighted with a box) causing interruptions to left-right turn traffic flow. This anomaly is detected by the GP models because activity patterns from regions highlighted in red colour are contrary to the predictive distribution computed using the past observations from other regions.

[1] T. Duong, H. Bui, D. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-Markov model. In *CVPR*, pages 838–845, 2005.

[2] C. C. Loy, T. Xiang, and S. Gong. Multi-camera activity correlation analysis. In *CVPR*, pages 1988–1995, 2009.

[3] M. R. Neal. *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics. Springer, 1996.

[4] C. E. Rasmussen and C. K. I. Williams. *Gaussian Process for Machine Learning*. MIT Press, 2006.

[5] T. Xiang and S. Gong. Video behaviour profiling for anomaly detection. *TPAMI*, 30(5):893–908, 2008.

[6] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan. Semi-supervised adapted HMMs for unusual event detection. In *CVPR*, pages 611–618, June 2005.