# Modelling Activity Global Temporal Dependencies using Time Delayed Probabilistic Graphical Model

Chen Change Loy, Tao Xiang and Shaogang Gong
School of Electronic Engineering and Computer Science
Queen Mary University of London, London E1 4NS, UK
{ccloy,txiang,sgg}@dcs.qmul.ac.uk

## Abstract

*We present a novel approach for detecting global behaviour anomalies in multiple disjoint cameras by learning time delayed dependencies between activities cross camera views. Specifically, we propose to model multi-camera activities using a Time Delayed Probabilistic Graphical Model (TD-PGM) with different nodes representing activities in different semantically decomposed regions from different camera views, and the directed links between nodes encoding causal relationships between the activities. A novel two-stage structure learning algorithm is formulated to learn globally optimised time-delayed dependencies. A new cumulative abnormality score is also introduced to replace the conventional log-likelihood score for gaining significantly more robust and reliable real-time anomaly detection. The effectiveness of the proposed approach is validated using a camera network installed at a busy underground station.*

## 1. Introduction

Detecting video anomalies is non-trivial because abnormal behaviours are rare, difficult to define and can be subtle, ambiguous, and easily confused with noise. Compared with single view anomaly detection, detecting abnormal behaviours captured by a network of cameras is even more challenging particularly when the camera views are not overlapped. This is because anomalies can take place globally across multiple camera views even though they often appear normal if observed in isolated camera views. For instance, consider two cameras monitoring road junctions A and B which are one mile apart and a vehicle passing A will typically appear at B in 2 minutes; it is normal to observe either large or small volume of traffics in either views. However, if large volume of traffic is observed at Junction A but two minutes later only few vehicles can be seen in Junction B, it is possible to infer that something abnormal (*e.g.* a road accident) has just happened between A and B, provided that the time delayed causal relationship between activities in A and B can be discovered and quantified. To further complicate the matter, the network can be extended to include more junctions each of which can contribute to what is observed in Junction B to different extents by different time delays.

It is therefore essential to learn global activity dependencies and the associated time delays for detecting global anomalies in disjoint camera views. This requires not only discovering and interpreting the temporal and causal relationships between activities, but also collecting relevant vital cues for detecting global anomalies. In this context, a global anomaly can be detected when the learned normal time-delayed causal relationships are not supported by visual evidence collected cross camera views on-the-fly.

To this end, we propose a novel approach for modelling globally optimised time-delayed dependencies between distributed local activities by formulating a Time Delayed Probabilistic Graphical Model (TD-PGM), based on which we develop a framework for global anomaly detection in multiple disjoint camera views. In our model, each node represents activities in a semantically decomposed region from one of the camera views, and the directed links between the nodes encode the causal relationships between the activities. The time delayed dependencies among activities across camera views are globally optimised using a novel two-stage structure learning method. In the first stage, a prior structure of the graphical model is learned based on Time Delayed Mutual Information (TDMI). This is followed in the second stage by a scored-searching based structure learning method derived by modifying the K2 algorithm [6]. This two-stage method ensures accurate and efficient learning of globally optimised dependencies among activities.

Once learned, our model can be used for real time anomaly detection by examining the log-likelihood of visual evidence collected from all camera views given the model. However, since we are interested in analysing busy public scenes featured with severe occlusions and low image resolution both spatially and temporally, the detection could potentially be sensitive to noise resulting in large

number of false alarms. To overcome this problem, a new Cumulative Abnormality Score (CAS) is introduced to replace the conventional log-likelihood (LL) score for more robust and reliable real-time anomaly detection.

The novelties of this work are: (1) To the best of our knowledge, there is no reported study on modelling time delayed global activity dependencies for real-time detection of subtle and ambiguous global behaviour anomalies across distributed multi-camera views of busy public scenes. (2) A novel structure learning method is proposed to discover and quantify globally optimised time delayed dependencies; this is achieved without any prior knowledge and assumptions on the camera topology and the method is fully unsupervised. (3) A new CAS is formulated for more robust real-time anomaly detection. Comparative experiments are carried out to demonstrate the effectiveness of the proposed approach using 177 hours of videos from a camera network installed at a busy underground station with complex and diverse scenes including ticket hall, concourse, train platforms and escalators.

## 2. Related Work

Existing multi-camera activity modelling approaches can be categorised into two groups: tracking based [9, 12, 14, 17, 18] and event based [11, 19]. With a tracking-based approach, one must solve the camera topology inference problem [12, 14] and the trajectory correspondence problem [9]. Both problems are non-trivial and remain unsolved for a large number of cameras and complex activity patterns in a crowded public scene captured in low spatial and temporal resolutions [17]. Recently Wang *et al*. propose a method which bypasses the topology inference and correspondence problem [17]. However, the method still cannot cope with busy scenes. Moreover, their trajectory co-clustering method is based on Latent Dirichlet Analysis (LDA) and is thus limited to capture only co-occurrence relationships among activity patterns. Any temporal relationship is not discovered and quantified automatically but simply determined by a pre-defined temporal threshold.

Alternatively, an event-based approach aims to avoid explicit object-centred segmentation and tracking. Zhou and Kimber [19] detect blob events in each camera view and model them as a first-order Markov chain in a Coupled Hidden Markov Model (CHMM). The chain's connectivity is manually defined and labelled to reflect neighbouring relationships of cameras. The model becomes intractable even given a small number of camera views. Moreover, the model is restricted to capture first order temporal dependency, which is not suitable for modelling cross-camera activity dependencies with arbitrary time delays. In comparison, our approach learns the activity dependencies without any prior knowledge on the camera topology or top-down rules for labelling. In addition, it is computationally efficient and can handle arbitrary time delays among activities.

Our approach is closely related to our previous work [11], which learns pairwise time delayed correlations among multi-camera activities using Cross Canonical Correlation Analysis (xCCA). However, our approach differs from and is advantageous over that in [11] in the following ways: (1) Their method is limited to the discovery of pairwise linear correlations without considering multiple dependencies in a global context; it is thus not suitable for a complex network where activity can have multiple causes from different views. In contrast, our approach performs structure learning for discovering the dependencies globally among all activity patterns. (2) Compared to the correlation-based method, we explore TDMI to take into account possible non-linear dependencies among activity patterns across different views. (3) Global anomaly detection is not attempted in [11].

Our approach is centred around a novel two-stage structure learning algorithm for a TD-PGM. There are a large number of methods in the literature on learning the structure of a graphical model. They are conventionally categorised as either constraint-based methods [3, 5, 10], or scored-searching based methods [6, 8]. More recently, hybrid methods have been proposed to combine both methods above in order to improve computational efficiency and prediction accuracy in structure learning [1, 15, 16]. With the same objectives, our two-stage structure learning method differs from existing hybrid methods in that it is capable of learning graph dependencies among multiple time-series with unknown time delays.

## 3. Global Activity Modelling

### 3.1. Global Activity Representation

It is necessary to decompose each camera view into semantic regions where different activity patterns are observed (*e.g*. decompose a traffic junction into different lanes and waiting zones). To this end, the approach proposed in [11] is adopted which clusters a scene using spectral clustering based on correlation distances of local block spatio-temporal activity patterns. This results in $N$ regions across $K_c$ cameras. Given scene decomposition, activity patterns observed over time in a region $r_n$ with $n = 1, \ldots, N$, is represented as a bivariant time series: $\hat{\mathbf{u}}_n = (\hat{u}_{n,1}, \ldots, \hat{u}_{n,t}, \ldots, \hat{u}_{n,T})$, where $\hat{u}_{n,t}$ is the percentage of static foreground pixels within the $n$th region at time $t$ and $T$ is the total number of frames used in the learning process; and $\hat{\mathbf{v}}_n = (\hat{v}_{n,1}, \ldots, \hat{v}_{n,t}, \ldots, \hat{v}_{n,T})$, where $\hat{v}_{n,t}$ is the percentage of pixels within the region that are classified as moving foreground. Note that more sophisticated features such as optical flows can be considered if videos of high spatial and temporal resolution are available.

To obtain a more compact representation, we cluster the

original 2D feature space $(\hat{\mathbf{u}}_n, \hat{\mathbf{v}}_n)$ in each region independently using a Gaussian Mixture Model (GMM). The GMM is learned using Expectation-Maximisation (EM) with the number of component $K_n$ determined by automatic model order selection using Bayesian Information Criterion (BIC). The learned GMM is then used to classify activity patterns detected in each region at each frame into one of the $K_n$ components. Activity patterns in a region over time are thus represented using the class labels and denoted as a 1D vector $\mathbf{x}_n = (x_{n,1}, \ldots, x_{n,t}, \ldots)$, where $x_{n,t} \in \{1, 2, \ldots, K_n\}$ and $n = 1, \ldots, N$.

### 3.2. Time Delayed Mutual Information Analysis

TDMI is explored here to learn time delayed dependency between each pair of regional activity patterns. TDMI was originally proposed by Fraser and Swinney [7] for measuring the Mutual Information (MI) between a time series $\mathbf{x}(t)$ and a time shifted copy of itself $\mathbf{x}(t+\tau)$ as a function of time delay $\tau$. If we treat the regional activity patterns in the $i$th and $j$th regions as time series data and denote them as $\mathbf{x}_i(t)$ and $\mathbf{x}_j(t)$ respectively, the mutual information for a time delay $\tau$ between them is computed as follows:

$$\begin{aligned} \mathrm{I}(\tau) &= \sum_{\mathbf{x}_i(t)} \sum_{\mathbf{x}_j(t+\tau)} p\left(\mathbf{x}_i(t), \mathbf{x}_j(t+\tau)\right) \\ &\quad \cdot \log_2 \frac{p\left(\mathbf{x}_i(t), \mathbf{x}_j(t+\tau)\right)}{p\left(\mathbf{x}_i(t)\right) p\left(\mathbf{x}_j(t+\tau)\right)}, \end{aligned} \tag{1}$$

where $p\left(\mathbf{x}_i(t)\right)$ and $p\left(\mathbf{x}_j(t+\tau)\right)$ denote the marginal probability distribution functions of $\mathbf{x}_i(t)$ and $\mathbf{x}_j(t+\tau)$ respectively, and $p\left(\mathbf{x}_i(t), \mathbf{x}_j(t+\tau)\right)$ is the joint probability distribution function of $\mathbf{x}_i(t)$ and $\mathbf{x}_j(t+\tau)$. The probability distribution functions are approximated by constructing histograms with $K_n$ equal-width bins, each of which correspond to one aforementioned GMM class. Note that $\mathrm{I}(\tau) \geq 0$ with the equality if, and only if $\mathbf{x}_i$ and $\mathbf{x}_j$ are independent. If $\tau = 0$, TDMI is equivalent to MI between $\mathbf{x}_i$ and $\mathbf{x}_j$.

The time delay between $\mathbf{x}_i$ and $\mathbf{x}_j$ is estimated as:

$$\hat{\tau}_{\mathbf{x}_i\mathbf{x}_j} = \underset{\tau}{\operatorname{argmax}}\, \mathrm{I}(\tau), \tag{2}$$

and the corresponding TDMI is obtained as:

$$\hat{\mathrm{I}}_{\mathbf{x}_i\mathbf{x}_j} = \mathrm{I}\left(\hat{\tau}_{\mathbf{x}_i\mathbf{x}_j}\right). \tag{3}$$

We compute the time delay and TDMI for each pair of regional activity patterns to construct a TDMI matrix $\mathbf{I} = [\hat{\mathrm{I}}_{\mathbf{x}_i\mathbf{x}_j}]_{N\times N}$ and an associated time delay matrix $\mathbf{D} = [\hat{\tau}_{\mathbf{x}_i\mathbf{x}_j}]_{N\times N}$, to be used for the learning of globally optimised time-delayed dependencies as below.

### 3.3. Global Activity Dependency Modelling

Globally optimised time-delayed dependencies among regional activity patterns are modelled using a TD-PGM. This is achieved by taking two steps: (1) structure learning and (2) parameter learning. Let us first formally define the model. A TD-PGM is denoted as $B = \langle G, \Theta \rangle$ and consists of a directed acyclic graph (DAG) $G$ whose nodes represent a set of observations $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, in which an observation $\mathbf{x}_n$ corresponds to the activity patterns observed in the $n$th semantic region. The network is governed by a set of parameters denoted by $\Theta = \{\theta_n\}$. All the observations in the network are discrete variables due to the GMM clustering, allowing us to represent the conditional probability distribution $p(\mathbf{x}_n|\mathbf{pa}_n, \theta_n)$ between a child node and its parents $\mathbf{pa}_n$ in $G$ using a multinomial probability distribution. We assume conditional independence which implies that $\mathbf{x}_n$ is independent from its non-descendants given its parents. These relationships are represented through a set of directed edges $\mathbf{E}$, each of which points to a node from its parents on which the distribution is conditioned. Given any two observations $\mathbf{x}_i$ and $\mathbf{x}_j$, we represent a directed edge from $\mathbf{x}_i$ to $\mathbf{x}_j$ by writing $\mathbf{x}_i \rightarrow \mathbf{x}_j$, where $(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{E}$ and $(\mathbf{x}_j, \mathbf{x}_i) \notin \mathbf{E}$. Note that in our model each observation has an associated time index and the model aims to capture time-delayed dependencies among observation variables.

#### 3.3.1 Structure Learning

The optimal structure of the TD-PGM $B$ encodes the time delayed dependencies that we aim to discover and quantify. To learn this structure, we formulate a novel two-stage structure learning method which first generates a prior network structure based on the TDMI matrix $\mathbf{I}$ and the time delay matrix $\mathbf{D}$. Furthermore, we also re-formulate a scored-searching based method, the K2 algorithm [6], so that it can be employed to refine the prior network structure and produce a final dependency structure. This is necessary because of a limitation of the K2 algorithm as explained below.

The K2 algorithm is well suited for learning the dependency structure of a large camera network due to its superior computational efficiency compared to other methods such as Markov Chain Monte Carlo (MCMC) based structure learning. The computational speed up is gained through the use of topological ordering $\varphi$ that helps to reduce the network search space. The ordering $\varphi$ specifies that a node $\mathbf{x}_i$ can only be the parent of $\mathbf{x}_j$ if, and only if, $\mathbf{x}_i$ precedes $\mathbf{x}_j$ in $\varphi$. The conventional K2 algorithm is sensitive to $\varphi$. A randomly set $\varphi$ does not guarantee to give the most probable model structure. A straightforward solution to this problem is to apply the K2 algorithm exhaustively on all possible orderings to find a structure that maximises the score. However, this solution is clearly infeasible even for a moderate number of nodes, as the number of iterations required is $N!$ for a model with $N$ nodes .

Let us now describe in more details our structure learning method. Instead of setting $\varphi$ randomly, we derive it

from a prior network structure constructed based on the TDMI matrix $\mathbf{I}$ and time delay matrix $\mathbf{D}$. First, an optimal dependence tree is generated based on $\mathbf{I}$. This is achieved in two steps. (a) We assign weights following $\mathbf{I}$ to each possible edges of a weighted graph with node set $\mathbf{X}$ that encodes no assertion of conditional independence. (b) Prim's algorithm [13] is applied to find a subset of the edges that forms a tree structure that includes every node, in which the total weight $\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathbf{E}} \hat{\mathrm{I}}_{\mathbf{x}_i\mathbf{x}_j}$ of the tree is maximised. The output is an optimal dependence tree (Chow-Liu tree) [4] that best approximates the network joint probability.

Second, the edges of the tree structure are oriented. Typically, one can assign edge orientations by either selecting a random node as root node, or by performing conditional independence test [1] and scoring function optimisation over the graph [2]. These methods are either inaccurate or require exhaustive search on all possible edge orientations therefore computationally costly. To overcome these problems, we propose an effective and accurate approach to orient the edges by tracing the time delays for pair of nodes in the tree structure using $\mathbf{D}$ learned by the TDMI analysis. In particular, if the activity patterns observed in $\mathbf{x}_i$ are lagging the patterns observed in $\mathbf{x}_j$ with a time delay $\tau$, it is reasonable to assume that the distribution of $\mathbf{x}_i$ is conditionally dependent on $\mathbf{x}_j$. The edge is therefore pointed from $\mathbf{x}_j$ to $\mathbf{x}_i$. With a prior network structure defined by the edges, we can derive $\varphi$ by performing topological sorting. The whole process of obtaining the prior network structure and $\varphi$ is summarised in Alg. 1.

**Input**: An undirected weighted graph with a node set $\mathbf{X} = \{\mathbf{x}_n\}$, where $n = 1, 2, \ldots, N$, and edge set $\mathbf{E}$. TDMI matrix $\mathbf{I}$ and time delay matrix $\mathbf{D}$
**Output**: Prior network structure defined by $\mathbf{X}'$ and $\mathbf{E}'$ Topological ordering $\varphi$
$\mathbf{X}' = \mathbf{x}$, where $\mathbf{x}$ is an arbitrary node from $\mathbf{X}$;
$\mathbf{E}' = \emptyset$;
**while** $\mathbf{X}' \neq \mathbf{X}$ **do**
  Choose an edge $(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{E}$ with maximum weight $\hat{\mathrm{I}}_{\mathbf{x}_i\mathbf{x}_j}$, where $\mathbf{x}_i \in \mathbf{X}'$ and $\mathbf{x}_j \notin \mathbf{X}'$;
  $\mathbf{X}' = \mathbf{X}' \cup \{\mathbf{x}_j\}$;
  $\mathbf{E}' = \mathbf{E}' \cup \{(\mathbf{x}_i, \mathbf{x}_j)\}$;
  **if** $\mathbf{D}(\mathbf{x}_i, \mathbf{x}_j) > 0$ **then**
    $\mathbf{x}_i \rightarrow \mathbf{x}_i$;
  **else**
    $\mathbf{x}_i \leftarrow \mathbf{x}_j$;
  **end**
**end**
$\varphi$ = topological_sort ($\mathbf{E}'$);

**Algorithm 1**: Finding a prior network structure and the topological ordering.

We can now introduce a modified K2 algorithm (see Alg. 2) after obtaining $\varphi$. The algorithm iterates over each node $\mathbf{x}_n$ that has an empty parent set $\mathbf{pa}_n$ before the iterations. Note that a candidate parent is selected in accordance with the node sequence specified by $\varphi$. Parent nodes are greedily added to $\mathbf{pa}_n$ if the addition of the parent to $\mathbf{x}_n$ maximises the score of the network. A BIC score is adopted. Between each node $\mathbf{x}_n$ and its parent set $\mathbf{pa}_n$, the BIC score is computed as:

$$BIC(\mathbf{x}_n, \mathbf{pa}_n) = L_n - \frac{C_n}{2} \log T, \tag{4}$$

where $C_n$ is the number of parameters needed to describe the conditional distribution, whilst $L_n$ is the log probability of $\mathbf{x}_n$ given its parents set $\mathbf{pa}_n$, which is given as:

$$L_n = \log p(\mathbf{x}_n|\theta_n) = \sum_{t=1}^{T} \log p(x_{n,t}|\mathbf{pa}_n, \theta_n). \tag{5}$$

Different from the original K2 algorithm, in our algorithm parent's activity patterns are shifted with a relative delay to child node's activity patterns based on $\mathbf{D}$.

**Input**: A graph with a node set $\mathbf{X} = \{\mathbf{x}_n\}$, where $n = 1, 2, \ldots, N$, an ordering of the nodes $\varphi$, an upper bound $\eta$ on the number the parents a node may have, and time delay matrix $\mathbf{D}$
**Output**: Parents set of each node $\mathbf{pa}_n$
**for** $n = 1$ *to* $N$ **do**
  $\mathbf{pa}_n = \emptyset$;
  $score_{\text{old}} = BIC(\mathbf{x}_n, \mathbf{pa}_n)$;
  $OKToProceed = \mathbf{true}$;
  **while** $OKToProceed$ *and* $|\mathbf{pa}_n| < \eta$ **do**
    let $\mathbf{x}_m$ be the candidate parent of $\mathbf{x}_n$, $\mathbf{x}_m \notin \mathbf{pa}_n$, with activity patterns $\mathbf{x}_m(t + \tau)$, $\tau = \mathbf{D}(\mathbf{x}_n, \mathbf{x}_m)$, that maximises the $BIC(\mathbf{x}_n, \mathbf{pa}_n \cup \{\mathbf{x}_m\})$;
    $score_{\text{new}} = BIC(\mathbf{x}_n, \mathbf{pa}_n \cup \{\mathbf{x}_m\})$;
    **if** $score_{\text{new}} > score_{\text{old}}$ **then**
      $score_{\text{old}} = score_{\text{new}}$;
      $\mathbf{pa}_n = \mathbf{pa}_n \cup \{\mathbf{x}_m\}$;
    **else**
      $OKToProceed = \mathbf{false}$;
    **end**
  **end**
**end**

**Algorithm 2**: Our re-formulated K2 algorithm with a time delay factor being introduced.

### 3.3.2 Parameter Learning

There are two typical methods for estimating the parameters of a Probabilistic Graphical Model (PGM) given fully observed data, namely Maximum Likelihood Estimation (MLE) and Bayesian learning. In this study, we adopt the latter approach since it offers the flexibility of performing real-time learning. Specifically, we apply the BDeu prior (likelihood equivalent uniform Bayesian Dirichlet) [8] (a

conjugate prior of multinomial distribution) over model parameters. The use of conjugate prior allows us to update the network posterior distribution sequentially and efficiently. To account for a cross-region time delay factor, regional activity patterns are temporally shifted according to the time delay matrix $\mathbf{D}$ during the parameter learning stage.

### 3.4. Global Anomaly Detection

A conventional way for detecting anomalies is to examine the log-likelihood (LL), $\log p\left(\mathbf{x}_t|\Theta\right)$ of the observations given a model, *e.g*. [19]. Specifically, an unseen global activity pattern is detected as abnormal if

$$\log p\left(\mathbf{x}_t|\Theta\right) = \sum_{n=1}^{N} \log p\left(x_{n,t}|\mathbf{pa}_n, \theta_n\right) < \mathrm{Th}, \quad (6)$$

where $\mathrm{Th}$ is a pre-defined threshold, and $\mathbf{x}_t = \{x_{1,t}, \ldots, x_{n,t}, \ldots, x_{N,t}\}$ are observations at time slice $t$ for all $N$ regions from all the camera views in a camera network. However, given a busy public scene featured with severe occlusions and low image resolution both spatially and temporally, observations $\mathbf{x}_t$ inevitably contain noise and the LL-based method is likely to fail in discriminating the 'true' anomalies from noisy observations because both can contribute to a low value in $\log p\left(\mathbf{x}_t|\Theta\right)$, and thus cannot be distinguished by examining $\log p\left(\mathbf{x}_t|\Theta\right)$ alone.

We address this problem by introducing a Cumulative Abnormality Score (CAS) which alleviates the effect of noise by accumulating the temporal history of the likelihood of anomaly occurrences in each region over time. This is based on the assumption that noise would not persist over sustained period of time and thus can be filtered out when visual evidence is accumulated over time. Specifically, an abnormality score (set to zero at $t = 0$) is computed for each node in the TD-PGM on-the-fly to monitor the likelihood of abnormality for each region. Given observation $x_{n,t}$ for the $n$th region at time $t$, the log-likelihood of the regional activity $\mathbf{x}_n$ is computed as:

$$\log p\left(x_{n,t}|\mathbf{pa}_n, \theta_n\right). \quad (7)$$

If $\log p\left(x_{n,t}|\mathbf{pa}_n, \theta_n\right)$ is less than a threshold $\mathrm{Th}_n$, the abnormality score for $\mathbf{x}_n$, denoted as $c_{n,t}$, is increased as: $c_{n,t} = c_{n,t-1} + |\log p\left(x_{n,t}|\mathbf{pa}_n, \theta_n\right) - \mathrm{Th}_n|$. Otherwise it is decreased from the previous abnormality score: $c_{n,t} = c_{n,t-1} - \delta\left(|\log p\left(x_{n,t}|\mathbf{pa}_n, \theta_n\right) - \mathrm{Th}_n|\right)$ where $\delta$ is a decay factor controlling the rate of the decrease. $c_{n,t}$ is set to 0 whenever it becomes a negative number after a decrease. We therefore have $c_{n,t} \geq 0, \forall\{n,t\}$, with a larger value indicating higher likelihood of being abnormal.

A global anomaly is detected at each time frame when the total of cumulative abnormality score $\mathbb{C}_t = \sum_{n=1}^{N} c_{n,t}$ across all the regions is larger than a threshold $\mathrm{Th}$. Overall, there are two thresholds to be set. Threshold $\mathrm{Th}_n$ is set automatically to the same value for all the nodes as:

$$\mathrm{Th}_n = \overline{LL} - \sigma_{LL}^2. \quad (8)$$

where the $\overline{LL}$ and $\sigma_{LL}^2$ are the mean and variance of the log probabilities computed over all the nodes for every frames, which are obtained from a validation dataset. The other threshold $\mathrm{Th}$ is set according to the detection rate/false alarm rate requirement for specific application scenarios. Note that during the computation of log probability, the activity patterns of a parent node are referred based on the relative delay between the parent node and the child node.

Once a global anomaly is detected, the contributing regional activities can be identified by examining $c_{n,t}$. Particularly, $c_{n,t}$ for all regions are ranked in a descending order. The contributing regional activities of the anomaly is identified as being from the first $N_a$ regions in the rank that are accounted for a given fraction $P = [0, 1]$ of $\mathbb{C}_t$.

## 4. Experimental Results

**Dataset** – Our dataset contains fixed views from nine disjoint and uncalibrated cameras installed at a busy underground station with a ticket hall and a concourse leading to two single train platforms via escalators (see Fig. 1). Three cameras were placed in the ticket hall and two cameras were positioned to monitor the escalator areas. Both train platforms were covered by two cameras each. The video from each camera lasts around 20 hours from 5:42am to 01:19am the next day, giving a total of 177 hours of video footage at a frame rate of 0.7 fps. Each frame has a size of $320 \times 230$.

The dataset was divided into 10 subsets, each of which contains 5000 frames ($\approx$ 2-hour in length) per camera. Two subsets were used as validation data. For the remaining 8 subsets, 500 frames/camera from each subsets were used for training and the rest for testing, *i.e*. $10\%$ for training.

**Global Time-delayed Dependency Learning** – Using the training data, the 9 camera views were automatically decomposed into 65 semantically meaningful regions (see Fig. 1). Given the scene decomposition, the global activities, composed of 65 regional activities, are modelled using a TD-PGM. The model structure, which encodes the time-delayed dependencies among regional activities, was initialised using pairwise TDMI and then optimised globally using our two-stage structure learning method. The final structure of the TD-PGM is depicted in Fig. 2. As expected, most of the discovered dependencies are between regions from the same camera views which have short time-delays. However, a number of interesting dependencies between inter-camera regions were also discovered accurately. For instance, the entrance/exit region in Cam 6 (Region 41) has an edge pointing towards the bottom of the going-up escalator in Cam 5 (Region 33) with a time delay of 38 frames.

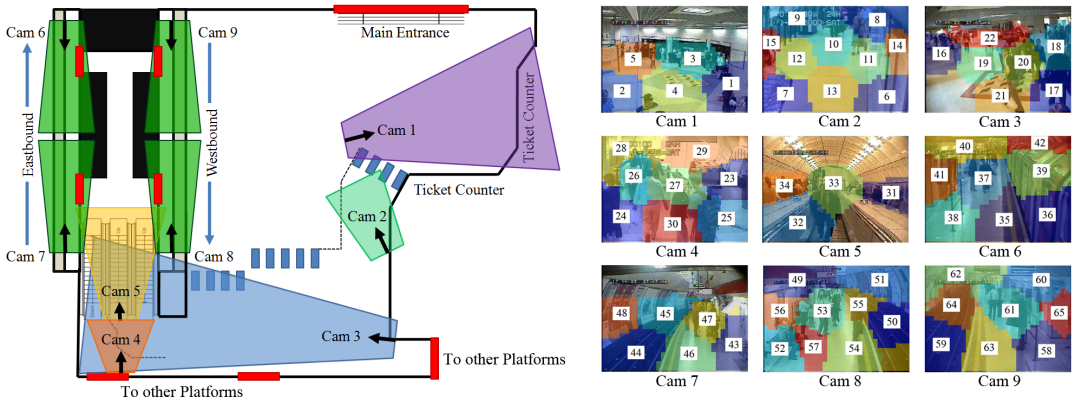


Figure 1. The underground station layout and the decomposed camera views for our dataset. Entry and exit points are shown in red bars. Passengers typically enter from the main entrance, walk through the ticket hall or queue up for tickets (Cam 1), enter the concourse through the ticket barriers (Cam 2, 3), take the escalators (Cam 4, 5), and enter one of the platforms. The opposite route is taken if they are leaving the station. The dataset is challenging due to: (1) complexity and diversity of the scene (*e.g.* behaviour on the platforms are very different from the behaviours in the ticket hall), (2) low video temporal and spatial resolution, (3) enormous number of objects appears in the scene especially during peak hours, (4) complex crowd dynamics, (*e.g.* passengers may appear in a group or individual, remain stationary at any point of the scenes, not to get on an arrived train etc.) (5) the existence of multiple entry and exit points that are not visible in the camera views.

This corresponds to the inter-camera activity of passengers leaving the northbound platform, walking along a corridor (invisible from the nine views), and taking the escalator up.

We compared our method (TDMI+SL) with three alternative dependency learning methods: (1) our two-stage structure learning method but initialised using MI rather than TDMI (MI+SL), to demonstrate the importance of encoding time-delay; (2) pairwise (without global) dependency learning method of [11] (xCCA without SL), to highlight the effectiveness of global dependency optimisation; (3) our structure learning method but initialised using xCCA rather than TDMI (xCCA+SL), to show the advantage of modelling non-linearity among activity dependencies using TDMI. Note that the same global activity representation described in Sec. 3.1 was used on both TDMI and xCCA based approaches. The results are shown in Fig. 3.

From Fig. 3(b), it is evident that without taking time delay into account, ML+SL yielded a number of incorrect dependency links such as $34 \rightarrow 23$ and $39 \rightarrow 45$ (highlighted with red circles), which are against the causal flow of activity patterns. Fig. 3(d) clearly shows that without global dependency optimisation, most of the dependencies discovered by method proposed in [11] were redundant. The result was greatly improved after applying the proposed structure learning method (see Fig. 3(c)). However, there were still a few missing links such as regions (31,23) and regions (2,4). This is mainly due to the use of pairwise linear correlations without taking into account non-linearity among activity dependencies across regions.

**Anomaly Detection** – For quantitative evaluation of our anomaly detection method, ground truth was obtained by exhaustive frame-wise examination on the entire test set.

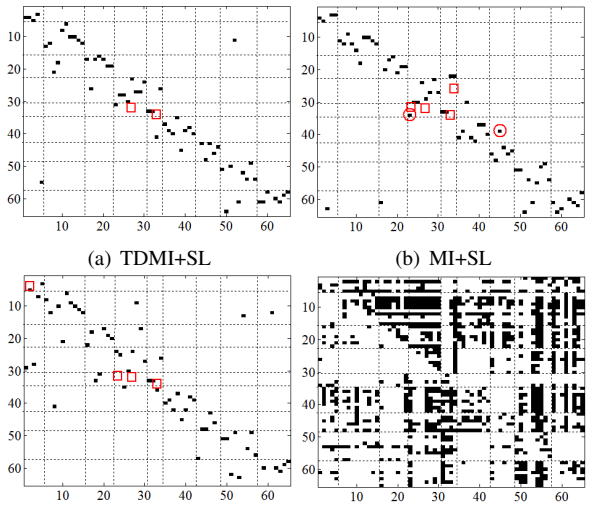


Figure 3. Activity global dependency structures learned using different methods. The y-axis represents the parent nodes, whilst the x-axis represents the child nodes. A black mark at (y,x) means $y \rightarrow x$. Missed edges and redundant edges are depicted using squares and circles respectively, except in (d) where there are too many.

Consequently, nine anomaly cases were found, each of them lasting between 34 to 462 frames with an average of 176 frames (254 secs). In total, there were 1585 anomalous frames accounting for 4.53% of the total frames in the test set. As shown in Table 1, these anomaly cases fall into three categories, all of which involve multiple regional activities.

A TD-PGM learned using our TDMI+SL method was employed for anomaly detection using the proposed CAS. One of the two free parameters in our approach, the de-

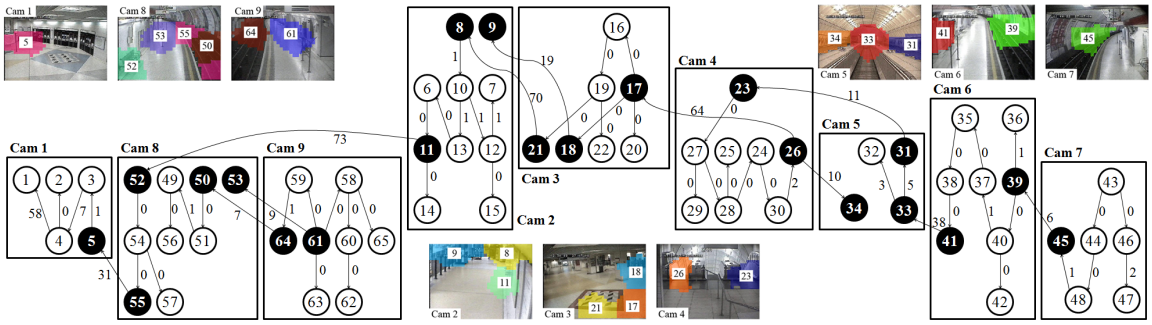


Figure 2. An activity global dependency graph learned using the two-stage structure learning method. Edges are labelled with the associated time delays discovered using the Time Delayed Mutual Information analysis. Regions and nodes with discovered inter-camera dependencies are highlighted.

| Cases | Anomaly Description | Cam | Total frames (% from total) |
|---|---|---|---|
| 1-6 | The queue in front of the ticket counters was built to a sufficient depth in Regions 1, 2, 4 that it blocked the normal route from Region 5 to 2 taken by passenger who did not have to buy ticket (see Fig. 6) | 1 | 1021 (2.92) |
| 7-8 | Faulty trains on the platforms in Cam 6 and 7 resulting in overcrowding. To prevent further congestion on the platform, passengers were disallowed to enter the platforms via the escalator (Region 34 in Cam 5). This in turn caused congestion in front of the escalator entry zone in Cam 4 (see Fig. 7) | 4,5,6,7 | 446 (1.27) |
| 9 | Train moved in reversed direction | 6,7 | 118 (0.34) |

Table 1. Ground truth.

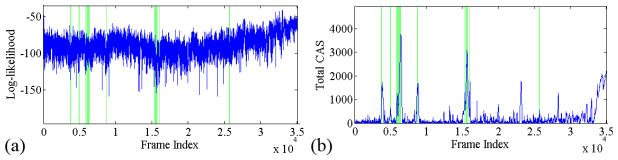


Figure 4. Anomaly scores computed using (a) log-likelihood (LL), and (b) cumulative abnormality score (CAS). Ground truths of anomalies are represented as bars in green colour.

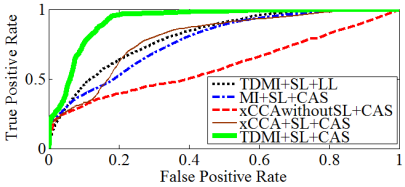


Figure 5. Receiving operating characteristic (ROC) curves obtained using different methods.

cay factor $\delta$ of CAS was set to 10 throughout the experiments. From our experiments, we found that consistent results were obtained with $\delta$ set beyond value 5. The performance of our approach (TDMI+SL+CAS) was evaluated using ROC curve by varying the other free parameter threshold Th. This was compared with four alternative approaches as shown in Fig. 5.

We first examine how effective the proposed CAS is for anomaly detection. Specifically our approach was compared with a method that use the same TD-PGM but with the conventional LL score, denoted as TDMI+SL+LL. As can be seen from Fig. 4, using the LL-based abnormality score, the true anomalies are overwhelmed by the noise collected from the large number of regions and thus difficult to detect. The poor performance is also evident from its ROC curve in Fig. 5. In contrast, the proposed CAS effectively alleviated the effect of noise, thus offering much more superior anomaly detection performance.

We further investigate how anomaly detection performance can be affected when the global time-delayed dependencies among regional activities are not learned accurately. More specifically, TD-PGMs were learned using MI+SL, xCCA without SL, XCCA+SL respectively as described above. CAS was then used for anomaly detection. These three methods are denoted as MI+SL+CAS, xCCAwithoutSL+CAS, and xCCA+SL+CAS respectively. For xCCAwithoutSL+CAS, the structure discovered by xCCA cannot be applied directly since the structure is not acyclic. We therefore followed steps described in Sec. 3.3.1 to generate an optimal dependence tree with pairwise correlations for comparison. It is evident from Fig. 5 that without accurate dependencies learned using our TDML+SL, all three methods yielded much poorer performance. In particular, the missing time-delayed dependencies shown in Fig. 3 caused missed detections of true anomalies whilst those redundant dependencies resulted in false positives.

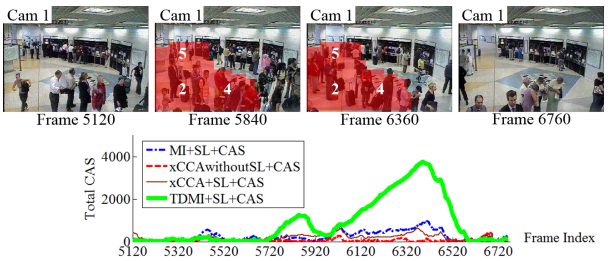


Figure 6. Example frames from detection output using the proposed framework on analysing anomalies caused by atypical long queues. The plot depicts the associated cumulative abnormality scores produced by different methods over the period. In ground truth, anomalies occurred at frames (5741-5853) and (5915-6376).

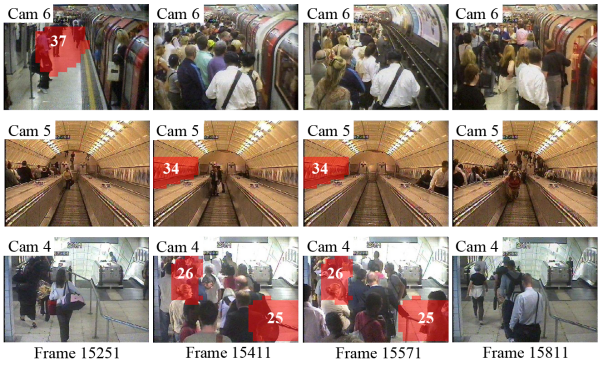


Figure 7. Global anomaly due to faulty train, which first occurred in Cam 6 and Cam 7 (not shown here), and later propagated to Cam 5 and Cam 4. In ground truth, anomalies occurred at frames (15340-15680).

An example of anomaly detection using the proposed approach is given in Fig. 6. The contributing anomalous regions are highlighted in red following the method described in Sec. 3.4 with $P = 0.8$. The atypical long queue led to a robust detection using our approach. In comparison, other methods yielded a weaker response or no response at all.

Another example of anomaly detection using our approach is shown in Fig. 7. This anomaly was Case 7 in Table 1. As can be seen in Fig. 7, our approach detected the anomaly across Cam 4, Cam 5 and Cam 6 successfully. Specifically, our model first detected abnormal crowd dynamics in Cam 6, *i.e*. all train passengers were asked to get off the train. From frame 15340 to frame 15680, passengers were disallowed to use the downward escalator and therefore started to accumulate at the escalator entry zone in Cam 4. The congestion led to a high CAS in Region 25. A large volume of crowd in Region 34 of Cam 4 was expected due to the high crowd density in Region 26 of Cam 4. However, the fact that Region 34 was empty violated the model's expected time delayed dependency, therefore causing a high CAS in Region 34. The proposed approach also associated the anomaly occurred in Region 34 with Region 26, which has an immediate and direct causal effect to it.

We also followed the method proposed in [19] to construct a CHMM with each chain corresponded to a region. However, the model is computationally intractable on our platform (dual-Core 3GHz processor with 4GB of RAM) due to the high space complexity during the inference stage.

## 5. Conclusions

We presented a novel approach to learn time delayed activity dependencies for global anomaly detection in multiple disjoint cameras. Time delayed dependencies are learned globally using a new two-stage structure learning method. Extensive experiments demonstrate that the new approach outperforms methods that disregard the time delay factor or without learning dependency structure globally. Finally, the proposed cumulative abnormality score has yielded superior result in achieving robust and reliable anomaly detection compared to conventional log-likelihood score.

## References


[1] X. Chen, G. Anantha, and X. Lin. Improving Bayesian network structure learning with mutual information-based node ordering in the K2 algorithm. *TKDE*, 20(5):628–640, 2008.

[2] X. Chen, G. Anantha, and X. Wang. An effective structure learning method for constructing gene networks. *Bioinformatics*, 22(11):1367–1374, 2006.

[3] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu. Learning Bayesian networks from data: an information-theory based approach. *Artif. Intell.*, 137(1):43–90, 2002.

[4] C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Trans. on Information Theory*, 14(3):462–467, 1968.

[5] G. F. Cooper. A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Min. Knowl. Discov*, 1(2):203–224, 1997.

[6] G. F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, (9):309–347, 1992.

[7] A. M. Fraser and H. L. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review*, 33(2):1134–1140, 1986.

[8] D. Heckerman, D. Geiger, and D. M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, pages 197–243, 1995.

[9] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *CVPR*, pages 26–33, 2005.

[10] M. Kalisch and P. Buhlmann. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *J. Machine Learning Research*, pages 613–636, 2007.

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

[12] D. Makris, T. Ellis, and J. Black. Bridging the gaps between cameras. In *CVPR*, pages 205–210, 2004.

[13] R. C. Prim. Shortest connection networks and some generalizations. *Bell Sys. Tech. J.*, 36:1389–1401, 1957.

[14] K. Tieu, G. Dalley, and W. E. L. Grimson. Inference of non-overlapping camera network topology by measuring statistical dependence. In *ICCV*, pages 1842–1849, 2005.

[15] I. Tsamardinos, L. E. Brown, and C. F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, 2006.

[16] M. Wang, Z. Chen, and S. Cloutier. A hybrid Bayesian network learning method for constructing gene networks. *Computational Biology and Chemistry*, 31:361–372, 2007.

[17] X. Wang, K. Tieu, and W. E. L. Grimson. Correspondence-free activity analysis and scene modeling in multiple camera views. *TPAMI*, In Press, 2009.

[18] E. E. Zelniker, S. Gong, and T. Xiang. Global abnormal behaviour detection using a network of CCTV cameras. In *IEEE Intl. Workshop on Visual Surveillance*, 2008.

[19] H. Zhou and D. Kimber. Unusual event detection via multi-camera video mining. In *ICPR*, pages 1161–1166, 2006.