# Detecting anomalous events at railway level crossings

Hajananth Nallaivarothayan, David Ryan, Simon Denman, Sridha Sridharan, Clinton Fookes and Andry Rakotonirainy

*Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 2013 227: 539
DOI: 10.1177/0954409713501296

Special Issue on work of the Cooperative Research Centre for Rail Innovation, Australia

Institution of MECHANICAL ENGINEERS

Journal of RAIL AND RAPID TRANSIT

Proc IMechE Part F:
J Rail and Rapid Transit
227(5) 539–553

DOI: 10.1177/0954409713501296
pif.sagepub.com

SAGE

# Detecting anomalous events at railway level crossings

**Hajananth Nallaivarothayan[1], David Ryan[1], Simon Denman[1], Sridha Sridharan[1], Clinton Fookes[1] and Andry Rakotonirainy[2]**

**Abstract**

Collisions between pedestrians and vehicles continue to be a major problem throughout the world. Pedestrians trying to cross roads and railway tracks without any caution are often highly susceptible to collisions with vehicles and trains. Continuous financial, human and other losses have prompted transport related organizations to come up with various solutions addressing this issue. However, the quest for new and significant improvements in this area is still ongoing. This work addresses this issue by building a general framework using computer vision techniques to automatically monitor pedestrian movements in such high-risk areas to enable better analysis of activity, and the creation of future alerting strategies. As a result of rapid development in the electronics and semi-conductor industry there is extensive deployment of CCTV cameras in public places to capture video footage. This footage can then be used to analyse crowd activities in those particular places. This work seeks to identify the abnormal behaviour of individuals in video footage. In this work we propose using a Semi-2D Hidden Markov Model (HMM), Full-2D HMM and Spatial HMM to model the normal activities of people. The outliers of the model (i.e. those observations with insufficient likelihood) are identified as abnormal activities. Location features, flow features and optical flow textures are used as the features for the model. The proposed approaches are evaluated using the publicly available UCSD datasets, and we demonstrate improved performance using a Semi-2D Hidden Markov Model compared to other state of the art methods. Further we illustrate how our proposed methods can be applied to detect anomalous events at rail level crossings.

**Keywords**

Computer Vision, Imaging/ Image Processing, Image Analysis, Intelligent Systems, Electronic Engineering, Computer Applications

Date received: 20 November 2012; accepted: 23 July 2013

## Introduction

Rapid development in the semi conductor industry has led to the ubiquitous deployment of surveillance cameras in public and secure places. Due to their affordability, CCTV cameras are installed anywhere where monitoring is required. But it is impractical to monitor all the video feeds with human operators. The need for several human operators and the difficulties in detecting events as they occur are the main difficulties currently faced in surveillance. Furthermore it is quite natural that human operators won't be able to continuously monitor the video footage due to fatigue and they won't capture all the important content in the surveillance video due to the nature of human visual perception. This can cause them to miss the most informative content of the video footage, such as any crucial events, and eventually results in failures and holes in the surveillance system. Hence, the rapid increase in the deployment of CCTV systems and the challenges posed by direct human monitoring have led to a greater demand for computer algorithms that are able to process the video feeds to extract information of interest for human operators.

Pedestrian safety is of paramount importance to the rail industry, and one of the biggest sources of risk for pedestrians and road and rail users alike is level crossings. The continued impact and resulting human, financial, and other economic costs associated with level-crossing incidents has driven various organizations to put more efforts into reducing the number

[1]Queensland University of Technology, School of Electrical Engineering and Computer Science, Australia
[2]Queensland University of technology, School of Psychology and Counselling, Australia

**Corresponding author:**
Hajananth Nallaivarothayan, [1]School of Electrical Engineering and Computer Science. [2]Queensland University of Technology, GPO Box 2434, 2 George St Brisbane, Queensland 4001, Australia.
Email: hajananth.nallaivarothayan@student.qut.edu.au

of incidents. However, despite the investment and motivation to improve safety, most analysis conducted to date continues to reveal that errors or violations on the part of the road user or pedestrian are the largest contributor to level crossing incidents.

Automated visual surveillance using computer vision technologies addresses real-time observation of people and vehicles within a busy environment, leading to a description of their actions and interactions.[1] This has received more research attention and funding due to increased global security concerns and an ever increasing need for effective monitoring of public places such as airports, railway stations, shopping malls, crowded sports arenas, military installations, and so on, or for use in smart health care facilities such as daily activity monitoring and fall detection in old people's homes.[2]

Major issues related to surveillance tackled by the current computer vision researchers are moving object detection and tracking, object classification, human motion analysis, and activity understanding, touching on many of the core topics of computer vision, pattern analysis, and artificial intelligence.[1] Human event detection and behavioural analysis enable use of the solutions provided for the above issues. Abnormal event detection is one sub-category of event detection where the human actions and interactions are categorized as either normal or abnormal.

Event detection is used to categorize one person's action into a pre-learned or pre-defined action. But when it comes to abnormal event detection, this consists of a binary classifying process where there can only be two possible results (normal or abnormal). An abnormal event is defined subjectively rather than predefined. In certain contexts an event can be abnormal while in other contexts it can be very normal. Here the objective is to detect, recognize or learn interesting events[2] which contextually may be defined as a 'suspicious event',[3] 'irregular behaviour',[4] 'uncommon behaviour',[5] 'unusual activity/event/behaviour',[6] 'abnormal behaviour',[7] 'anomaly',[8] and so on.

Feature extraction, training and learning of normal activity models based on the extracted features of the training video and finally the classification of new video as normal or abnormal are the core components of an anomalous event detection system. As this is an unsupervised classification process almost all the models used in existing research are clustering algorithms. Many of these algorithms fail to capture the temporal and spatial correlation of the activities through the models. While some of the researchers have used Hidden Markov Models (HMMs) to model the temporal behaviour,[9–11] the modelling of spatial causality is omitted in all but a minority of systems.[8,12]

In this paper we propose three different types of HMMs to model both the temporal and spatial causalities. They are the Semi-2D HMM, Full-2D HMM and Spatial HMM. The Semi-2D HMM models the current state as being not only dependent on the previous state in the temporal direction, but also dependent upon the previous states in adjacent spatial locations (either horizontal or vertical). The Full-2D HMM is similar to Semi-2D HMM with modifications made to take into account temporal information from the adjacent spatial locations through their main temporal sequences. Spatial HMMs model the current state as being only dependent on the previous state in the spatial direction.

For the Semi-2D HMM and Spatial HMM, two model structures are investigated, modelling the causalities in either the vertical or horizontal direction. Within the HMM, outliers are detected to locate abnormal events. The proposed approaches use features extracted from spatial blocks and spatio-temporal patches. The features used are the location of the spatio-temporal block to capture the location-specific abnormalities, flow features to capture speed related abnormalities, and textures of optical flow to capture the anomalies related to the motion characteristics.[13]

The remainder of this paper: summarises related work in this field; describes the features used in the proposed algorithm; describes the Semi-2D HMM algorithm, the Full 2D algorithm, and the Spatial HMM algorithm; presents an evaluation of the publicly available USCD database[14]; discusses how the proposed approach can be applied to railway level crossings; presents conclusions and directions for further work.

## Related work

Event detection using computer vision technologies has been an active research topic for several years. Anomalous event detection is a sub-topic of event detection, where the events are classified into normal and abnormal activities. Anomalous event detection is a challenging problem in that it is difficult to explicitly define an anomaly. Anomalies related to rail level crossings have their own definition. Most of the past research was conducted on a global perspective of anomalous detection but evaluated with certain datasets. Those methods can be investigated for the rail level-crossing related anomalies with a little modification according to the context. In this section we summarize the related work in anomalous event detection on a global perspective.

Over the years, there has been a paradigm shift from rule-based anomalous event detection to statistically based methods to achieve a robust framework to conceptualize semantically meaningful scene behaviours.[2] We can divide the work done in anomalous event detection into two categories, called rule-based methods and statistically based methods.

Early event detection research investigated rule-based systems where pre-defined rules are used

to define normal and abnormal activities.[15] Their performance was good but their limitation in defining more and more pre-defined rules restricted them to be used in only specific types of anomalous detection. So they are limited in terms of robustness and scalability especially for unseen events in the scene.[2]

Statistical methods provide a means to identify an anomalous event when it appears, despite the fact that it has never occurred before.[16] Statistical methods can be subdivided into those that first learn a model of normal behaviour and use it as a basis to detect anomalies, and those that automatically learn the normal and abnormal patterns from the statistical properties of the observed data, either offline or online.[2] All methods have a common two-step framework: feature extraction and classification using a learned (online or offline) model.

Feature extraction can be done using both bottom-up and top-down approaches.

In the context of event detection, a top-down approach means each individual in the scene is segmented and features are extracted separately. Anomalous event detection using object tracking is an example of this approach, where individuals' object trajectories are obtained and the individuals with abnormal trajectories are deemed to be performing an abnormal event. Despite the limitations of trajectory analysis, it has been widely used to detect abnormalities. Among the trajectory analysis works, Zhou et al. group similar trajectories using the Edit Distance (ED);[17] Hu et al. associate foreground pixel masks with extracted trajectories, providing a more descriptive representation of the activities than trajectories alone;[18] and Morris and Trivedi represent trajectories by a series of flow vectors.[19] Like Zhou et al., Morris and Trevedi group similar trajectories, and an HMM is trained to represent each characteristic trajectory.[19] Vasquez et al. modelled the object's motion in terms of an augmented continuous state vector, composed of two sets of variables describing its current and intended (goal) states based on the key observation that often, objects move as a function of their intention to reach a particular state (goal).[20] This approach can be effective in a sparsely crowded environment, though in dense crowds it is very challenging to track each individual separately due to clutter and dynamic occlusions.

Bottom-up approaches are stimulus-driven approaches. Instead of tracking individual objects, features are extracted that represent the underlying scene characteristics and crowd behaviour. These approaches can work very well in densely crowded environments amidst extensive clutter and dynamic occlusions. Features extracted for the bottom-up approaches are at pixel level and are generally referred to as low-level features.

Xiang and Gong have proposed the Pixel Change History (PCH) for measuring multi-scale temporal changes at each pixel.[21] Zaharescu and Wildes have used distributions of spatio-temporal oriented energy.[22] Andrade et al. used optical flow patterns,[9] and spatial histograms of the detected objects are used as the feature by Zhong et al.[6] Zhao et al. used histograms of gradients (HoG) and histograms of optical flow (HoF).[23] Computing these features can be slow due to the need to calculate dense optical flow fields for all frames at full resolution. Additionally, the motion patterns captured by these algorithms are often incomplete due to the dimensionality reduction or histogram binning process. Incomplete motion information will cause the anomaly detection algorithm to fail in some scenarios. Ryan et al. proposed a visual representation called textures of optical flow, which captures both the smoothness of the flow and the presence of motion.[13] This may be useful for detecting bicycles or vehicles in a pedestrian scene, for example the UCSD dataset.[14]

The various low-level features and object level features that are extracted are the input to a learning model. Popular learning models include HMM, Petri net,[24] LDA,[25,26] Support Vector Machine (SVM) and Markov Random Field (MRF).[27]

Typically, the algorithms group the video into different clusters for analysis. This can be done in the spatial domain, time domain, or a three dimensional spatial–temporal domain. Hierarchical Bayesian Models are used by Xiaogang et al. to detect anomalies in crowded scenes.[26] Similarly Mehran et al. use LDA and a bag of words methodology to learn a 'normal' model, after which frames can be classified as either abnormal or normal.[28] Adam et al. use histogram binning of the extracted features, while anomaly detection is done by using a cyclic buffer to determine the likelihood of new observations.[29] Kim and Grauman use a mixture of probabilistic principal component analysers to model their features.[27] Hamid et al. represent activities as bags of event n-grams, where global structural information of activities is analysed using local event statistics.[30] Zhao et al. proposed a fully unsupervised dynamic sparse coding approach for detecting unusual events in videos based on on-line sparse reconstructibility of query signals from a learned event dictionary, which forms a sparse coding base.[23] Further, Ryan et al. and Greenspan et al. utilized GMMs for their feature modelling,[13,31] while Zhong et al. used K-means clustering to group the video segments into disjointed sets.[6] Mahadevan et al. use a generative mixture of dynamic textures.[14] Reddy et al. model the motion and size features by an approximated version of kernel density estimation and the texture features by an adaptively grown codebook.[32]

The above analysed models generally do not capture the temporal behaviour of the crowd, such as repetitiveness and continuity of the activities as these techniques fail to model the interrelationship

between individual observations. This will result in important information relating to the pattern and duration of the normal activities not being captured by the learning model, making the detection of abnormalities more challenging.

HMMs provide a means to capture temporal dependencies within the detection process. Andrade et al. use a bank of HMMs trained on normal behaviours, and detect a sequence as anomalous when the likelihood falls below a threshold.[9,10] Kratz and Nishino used the symmetric KL divergence as a distance measure, and identified spatio-temporal cuboids in the video sequence by associating local spatio-temporal motion patterns that have a small distance between them.[11] They modified the parameters of the clusters of Gaussian distributions in an online manner using the KL distance. After deriving the prototypes of similar activities represented by the cuboids in the scene they modelled the temporal relationship by a Hidden Markov Model for every spatial location.

Zhang et al. proposed a semi-supervised adapted HMM framework.[33] Snoek et al. used an HMM to analyse the temporal progression of the affine features.[34] Jiang et al.[35] proposed an unusual video event-detection method based on unsupervised clustering of object trajectories, which are modelled by HMMs. Vasquez et al.[20] proposed growing HMMs which they describe as time-evolving HMMs, with continuous observation variables, where the number of discrete states, structure, and probability parameters are updated every time a new observation sequence is available. These models only capture the causality in the temporal direction while the information about the adjacent behaviour is missed.

Kratz and Nishino also used coupled HMMs to capture the spatial relationships.[8] They used separate HMMs for each spatial location and during the classification process they computed spatial confidence measures using the surrounding HMMs of the current HMM, and combined it with a temporal classifier for the detection of anomalous behaviour. Though they have considered the spatio-temporal cubes adjacent to the current cube during the classification, there is no information gathered about the spatial causality during the training process. Utasi and Czuni construct their models at two levels, a region-based continuous distribution HMM, and a higher-level HMM to inter-link those regional HMMs that form the first level.[12] Here, spatial information is missing at the low-level HMMs and only limited spatial causality can be captured by the high-level discrete HMM.

While a variety of approaches using HMMs have been proposed, none of these adequately capture both the spatial and temporal dependencies in the scene, leading to a loss of information, and potentially accuracy.

## Feature extraction

We use three features within the proposed system:

(a) location features (center coordinate of a spatial block) to detect the location-specific anomalies;
(b) motion information (summation of optical flow vectors inside a block) to identify the anomalies related to speed of movements of the objects;
(c) textures of optical flow to identify the anomalies related to the type of motion that is occurring.[13] For example, flow may be smooth and constant or highly variable and turbulent. This feature is useful for detecting anomalous objects, such as bicycles and vehicles, and can be computed in real time.

Features are extracted in spatial blocks as outlined in the next subsection.

To calculate the optical flow vectors, we have used Black and Anandan's algorithm.[36] To ensure the proposed approach is computationally efficient, we downsample the input video. In the proposed system, we place a greater emphasis on having an accurate optical flow estimate (i.e. using a robust estimator) than requiring high resolution optical flow images. We feel this is justified as the anomalous events and objects are still clearly visible even at lower resolutions.

The motion features across a block $B$ are given by

$$\sigma_u = \sum_{(x,y)\in B} u(x,y) \tag{1}$$

$$\sigma_v = \sum_{(x,y)\in B} v(x,y) \tag{2}$$

Textures of optical flow, which measures the uniformity of the motion, is computed from the dot product of flow vectors at different offsets. Having uniformity measures computed from different offsets inside a feature vector is useful for detecting objects of various sizes.[13]

We evaluate our proposed system with the following combinations of the three types of features:

1. All three features: textures of optical flow at various scales $\{\phi\}$, motion information $(\sigma_u, \sigma_v)$ and location features $(x, y)$,

$$\mathbf{f} = \left[\phi_{(1,1,0)}, \phi_{(3,3,0)}, \phi_{(5,5,0)}, \sigma_u, \sigma_v, x, y\right] \tag{3}$$

where $\phi_{(\delta,\delta,0)}$ is the uniformity feature value at $\delta$ offset.[13]

2. Optical flow vectors and location features alone,

$$\mathbf{f} = [\sigma_u, \sigma_v, x, y] \tag{4}$$

## Spatial blocks and observation sequences

The spatial blocks and observation sequences used for HMM input are extracted as follows.

We divide the video frames into non-overlapping spatial blocks and spatio-temporal patches of different configurable sizes. Features are extracted using each pixel within a block, and are summed to form the feature vector for the spatial block as well as for the spatio-temporal patch. During the training process an observation sequence of configurable length is created for each spatial location by collecting the feature vectors of the blocks belonging to the same spatial location for consecutive video frames from the training video data. The feature vector is then used in the testing to compute the likelihood of the observation sequence in the presence of the particular feature vector, and the block is classified as normal or abnormal based on the likelihood.

The size of the block ($7 \times 7$) is chosen, as it is similar to the size of an interesting object in the testing dataset used, and other previous work done using this dataset has used a similar block size.[13] The sequence length is chosen as 20 frames. The number of HMM nodes chosen is four for pedestrian two dataset and five for pedestrian one dataset which gave better performance than the other options.

## Semi-2D HMM

We propose a Semi-2D HMM to model the extracted observation sequences from the training video, and to detect abnormalities. Generally, HMMs are of one dimension and model the causality in this single direction. To capture causalities in more than one direction, various approaches that interconnect separate HMMs have been proposed, leading to alternate HMM-type models such as the Multi-Level HMM,[12] and coupled HMMs.[8] In the field of image classification, a form of 2D HMM has been used to capture the spatial causality of images in both vertical and horizontal directions.[37] However, for a video task, these 2D HMMs create too many observation sequences in different directions, making them computationally prohibitive. Here we propose a Semi-2D HMM which captures the causality in the temporal direction and the dependencies in adjacent spatial locations either horizontally (Figure 1(a)) or vertically (Figure 1(b)).

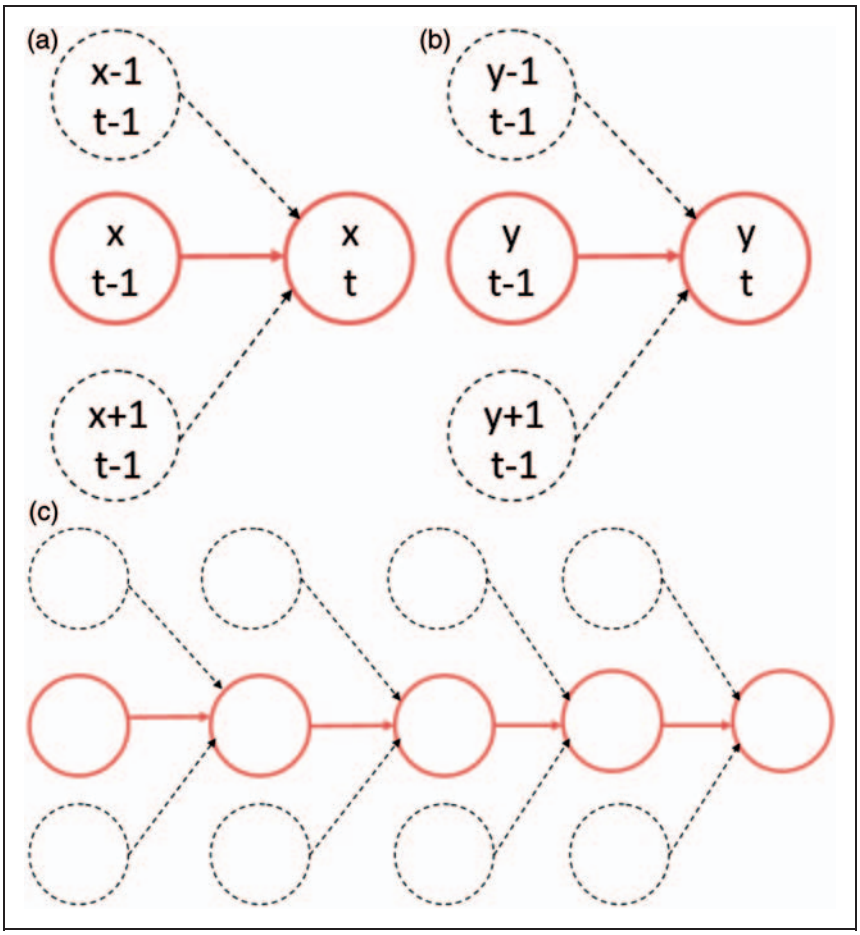


**Figure 1.** Schematic diagrams of the proposed Semi-2D HMM. (a) Temporal and spatial dependency diagram of the horizontal Semi-2D HMM, (b) Temporal and spatial dependency diagram of the vertical Semi-2D HMM and(c) Overall sequence diagram of a Semi-2D HMM.

## Assumptions of our HMM

The proposed approach makes three key assumptions. These are:

1. The current state is not only dependent on the previous state in the temporal direction, but also the previous states of the adjacent spatial locations.
2. The main observation sequence is in the temporal direction only (see Figure 1(c), the sequence in full circles is the main observation sequence).
3. Adjacent spatial observations in one sequence are part of another main temporal sequence.

## Parameters of the HMM

Our HMM consists of $N$ hidden-states which are visited in the sequence $Q = \{q_{t,x}\}_{t=1}^{T}$ at spatial location $x$ with the adjacent spatial dependency states $q_{t,x-1}$ and $q_{t,x+1}$ at time $t$. The set of observations $O = \{O_t\}_{t=1}^{T}$ is a Gaussian function of hidden states. Observations of adjacent spatial locations are denoted $O_{t,x-1}$ and $O_{t,x+1}$. Here both $q_{t,x}$ and $q_t$ denote the state at the $t^{th}$ time step at spatial location $x$, while $O_{t,x}$ and $O_t$ denote the relevant observation. Our model is based on the parameters detailed in the following subsections.

### Transition probabilities

The transition probability, $a_{g,i,h,j}$, denotes the probability of being in state $j$ at time $t+1$, given that the state of the same location at time $t$ is $i$ and the states of the adjacent spatial locations at time $t$ are $g$ and $h$. Adjacent locations in the horizontal direction are considered in the case of the horizontal HMM, and adjacent locations in the vertical direction are considered for the vertical HMM. The transition probability for the horizontal case is

$$a_{g,i,h,j} = p(q_{t+1,x} = j \,|q_{t,x-1} = g, q_{t,x} = i, q_{t,x+1} = h) \tag{5}$$

### Gaussian distribution parameters for likelihood of observations

The likelihood of an observation which belongs to a state $j$ is a Gaussian distribution with mean $\mu_j$ and covariance matrix $\Sigma_j$. The probability of an observation at time $t$, given that the state is $j$, is given by

$$b_j(O_t) = p(O_t|q_t = j) = \mathcal{N}(O_t|\mu_j, \Sigma_j) \tag{6}$$

### Initial probabilities

The initial probability of observing state $i$ is denoted by $\pi_j$

$$\pi_j = p(q_t = j) \tag{7}$$

## Algorithm

During the training process, model parameters are optimized in such a way to maximize the likelihood of the observed sequence. The Baum–Welch algorithm uses expectation maximization, where the likelihood of the observations is locally maximized by iteratively re-estimating the model parameters. The usual procedure for HMMs is slightly modified for the calculation of our model's parameters, as described below, with the remainder of the procedure remaining unchanged.[38]

### Forward procedure

This is the probability of observing the partial main observation sequence, $\{o_1, o_2, \ldots, o_t\}$ and $t^{th}$ observations at adjacent spatial locations $o_{t,x-1}$, $o_{t,x+1}$ with $q_t = i$

$$\alpha_t(i) = p(O_1, O_1, \ldots, O_t, O_{t,x-1}, O_{t,x+1}, q_t = i|\lambda) \tag{8}$$

The forward probability is calculated using an inductive algorithm in our work.

### Backward procedure

This is the probability of observing the main partial observation sequence from $t+1$ to the end of the sequence, and the $t^{th}$ observations at adjacent spatial locations $o_{t,x-1}$, $o_{t,x+1}$ given $q_t = i$

$$\beta_t(i) = p(O_{t+1}, O_{t+2}, \ldots, O_T, O_{t,x-1}, O_{t,x+1}|q_t = i, \lambda) \tag{9}$$

The backward probability is calculated using an inductive algorithm in our work.

### Expectation equations

The probability of being in state $i$ at time $t$, given the observations $O$ and the model parameters (collectively denoted $\lambda$) is given by

$$\gamma_i(t) = p(q(t) = i|O, \lambda) \tag{10}$$

The probability of being in state $i$ at time $t$, $j$ at time $t+1$ and in states $g$, $h$ at time $t$ at spatial locations $x-1$, $x+1$ respectively is denoted as $\epsilon_{g,i,h,j}(t)$ and given by equation (22).

Variables of the expected equations are calculated based on the forward and backward variables in each iteration of the inductive algorithm.

## Full-2D HMM

This model is similar to the model described in the previous section with modifications in the backward and forward variables to take account of temporal

information from the adjacent spatial locations through the main temporal sequence as shown in Figure 2.

## Parameters of the HMM

This HMM consists of $N$ hidden-states which are visited in the sequence $Q = \{q_{t,x}\}_{t=1}^{T}$, at spatial location $x$ with the adjacent spatial dependency states $q_{t,x-1}$ and $q_{t,x+1}$, at time $t$. The set of observations $O = \{O_t\}_{t=1}^{T}$ is a Gaussian function of hidden states. Observations of adjacent spatial locations are denoted as $O_{t,x-1}$ and $O_{t,x+1}$. Here both $q_{t,x}$ and $q_t$ denote the state at the $t^{th}$ time step at spatial location $x$, while $O_t$ and $O_{t,x}$ denote the relevant observation. This model is based on the following parameters:

### Transition probabilities

The transition probability, $a\_down_{g,j}$, denotes the probability of being in state $j$ at time $t+1$, given that the state of the adjacent spatial location $(x-1)$ at time $t$ is $g$.

The transition probability, $a\_direct_{i,j}$, denotes the probability of being in state $j$ at time $t+1$, given that the state of the same spatial location $(x)$ at time $t$ is $i$.

The transition probability, $a\_up_{h,j}$, denotes the probability of being in state $j$ at time $t+1$, given that the state of the adjacent spatial location $(x+1)$ at time $t$ is $h$.

Adjacent locations in the horizontal direction are considered in the case of the horizontal HMM, and adjacent locations in the vertical direction are considered for the vertical HMM. The transition probabilities for the horizontal case are

$$a\_up_{g,j} = p(q_{t+1,x} = j \,|\, q_{t,x-1} = g) \tag{11}$$

$$a\_direct_{i,j} = p(q_{t+1,x} = j \,|\, q_{t,x} = i) \tag{12}$$

$$a\_down_{h,j} = p(q_{t+1,x} = j \,|\, q_{t,x+1} = h) \tag{13}$$

### Gaussian distribution parameters for likelihood of observations

The likelihood of an observation which belongs to a state $j$ is a Gaussian distribution with mean $\mu_j$ and covariance matrix $\Sigma_j$. The probability of an observation at time $t$, given that the state is $j$, is given by

$$b_j(O_t) = p(O_t | q_t = j) = \mathcal{N}(O_t | \mu_j, \Sigma_j) \tag{14}$$

## Algorithm

During the training process, model parameters are optimized in such a way to maximize the likelihood of the observed sequence. The Baum–Welch algorithm uses expectation maximization, where the likelihood of the observations is locally maximized by iteratively re-estimating the model parameters. The procedure used in the previous main section is slightly modified for the calculation of this model's parameters, as described below, to take account of temporal information from the adjacent spatial locations through the main temporal sequence with the remainder of the procedure remaining unchanged.

### Forward procedure

This is the probability of observing the partial observation sequence, $\{o_1, o_2, \ldots, o_t\}$ and observation sequences at adjacent spatial locations are $\{o_{1,x-1}, o_{2,x-1}, \ldots, o_{t,x-1}\}$, $\{o_{1,x+1}, o_{2,x+1}, \ldots, o_{t,x+1}\}$ with $q_t = i$. The formula is given by equation (23), below, where $i$ is the state number.

The forward probability is calculated using an inductive algorithm in our work.

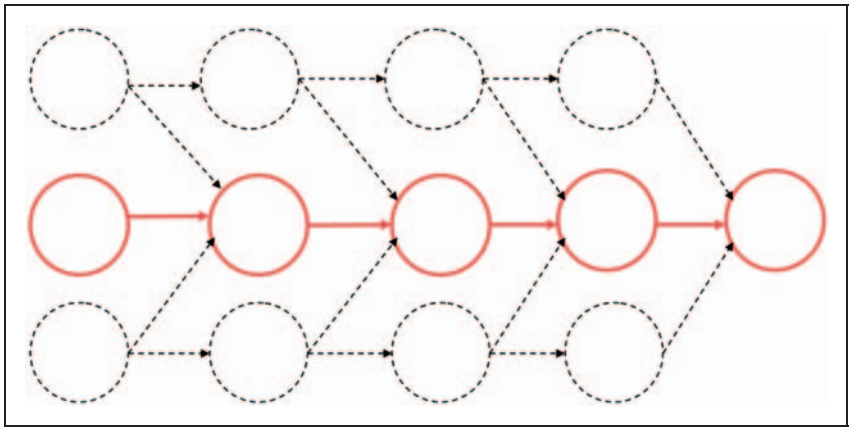

**Figure 2.** Schematic diagrams of the proposed Full-2D HMM.

### Backward procedure

This is the probability of observing the partial observation sequences from $t+1$ to the end of the sequences at locations $x$, $x-1$ and $x+1$ given $q_t = i$. The formula is given by equation (24).

The backward probability is calculated using an inductive algorithm in our work.

### Expectation equations

The probability of being in state $i$ at time $t$, given the observations $O$ and the model parameters (collectively denoted $\lambda$) is given by

$$\gamma_i(t) = p(q(t) = i|O, \lambda) \tag{15}$$

The probability of being in state $g$ at spatial location $x-1$ at time $t$, $j$ at spatial location $x$ at time $t+1$ is denoted as $\epsilon\_down_{g,j}(t)$ and given by

$$\epsilon\_down_{g,j}(t) = p(q_{t,x-1} = g, q_{t+1,x} = j\,|O, \lambda) \tag{16}$$

The probability of being in state $i$ at spatial location $x$ at time $t$, $j$ at spatial location $x$ at time $t+1$ is denoted as $\epsilon\_direct_{i,j}(t)$ and given by

$$\epsilon\_direct_{i,j}(t) = p(q_{t,x} = i, q_{t+1,x} = j\,|O, \lambda) \tag{17}$$

The probability of being in state $h$ at spatial location $x+1$ at time $t$, $j$ at spatial location $x$ at time $t+1$ is denoted as $\epsilon\_up_{h,j}(t)$ and given by

$$\epsilon\_up_{h,j}(t) = p(q_{t,x+1} = h, q_{t+1,x} = j\,|O, \lambda) \tag{18}$$

Variables of the expected equations are calculated based on the forward and backward variables in each iteration of the inductive algorithm.

## Spatial HMM

A Spatial HMM was designed to model the causality of the scene in the spatial direction. Features extracted from the spatio-temporal cuboids are used to create the observation sequences in both the spatial directions. Two HMMs each for a specific direction are created based on the spatial observation sequences from the particular directions. To model the horizontal dependencies, the observation sequence is created for a $y$-value (being constant) by considering the sequential blocks in the $x$-direction, and observation sequences are created for each $y$-value. Similarly the vertical modelling was done by keeping the $x$-value constant and by considering the sequential blocks in the $y$-direction. Figure 3 depicts the block diagram of the model.

In the following subsections, the design of the horizontal HMM is described.

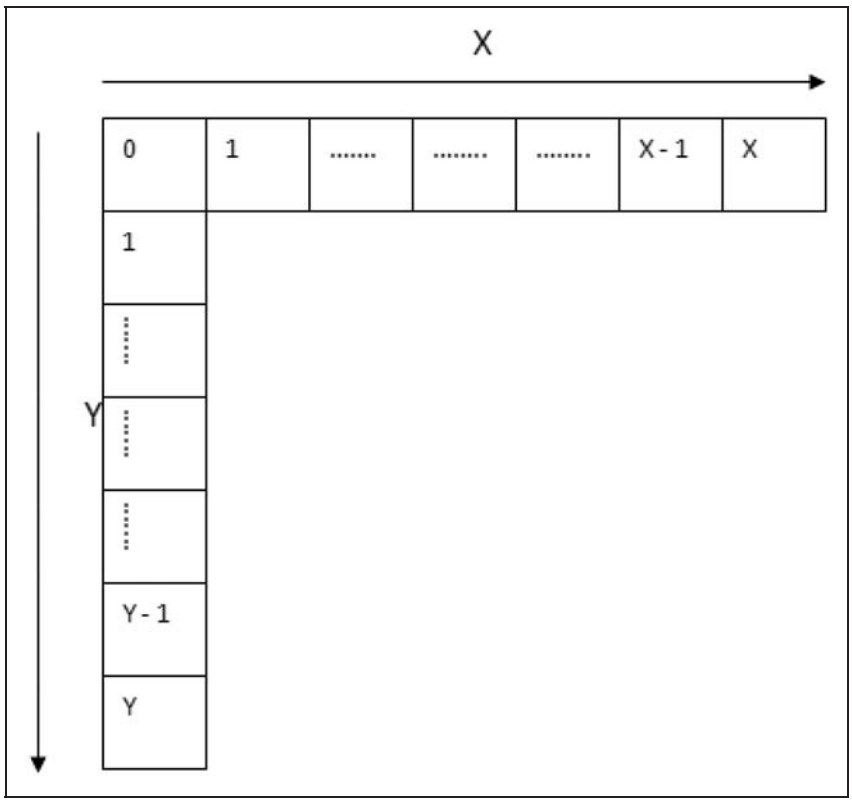


**Figure 3.** Spatial HMM.

## Assumptions

1. The current state is only dependent on the previous state in the horizontal spatial direction.
2. State transition probabilities and emission probabilities don't vary with the spatial location.

## Parameters of the HMM

This HMM consists of $N$ hidden states which are visited in the sequence $Q = \{q_x\}_{x=1}^{X}$. The set of observations $O = \{O_x\}_{x=1}^{X}$ is a Gaussian function of hidden states. Here $q_x$ denotes the state at the spatial location x at $t^{th}$ time step while $O_x$ denotes the relevant observation, and $X$ is the sequence length in the horizontal direction. Further observation sequences are extracted from each frame that is processed. This model is based on the parameters in the following sections.

### Transition Probabilities

The transition probability, $a_{i,j}$, denotes the probability of being in state $j$ at spatial location $x+1$, given that the state at spatial location $x$ is $i$. The transition probability is given by

$$a_{i,j} = p(q_{x+1} = j \,|\, q_x = i) \tag{19}$$

### Gaussian distribution parameters for likelihood of observations

The likelihood of an observation which belongs to a state $j$ is a Gaussian distribution with mean $\mu_j$ and covariance matrix $\Sigma_j$. The probability of an observation at spatial location $x$, given that the state is $j$, is given by

$$b_j(O_x) = p(O_x|q_x = j) = \mathcal{N}(O_x|\mu_j, \Sigma_j) \tag{20}$$

### Initial probabilities

The initial probability of observing state $i$ is denoted by $\pi_j$

$$\pi_j = p(q_x = j) \tag{21}$$

$$\begin{aligned}\epsilon_{g,i,h,j}(t) = p(q_{t,x-1} = g, q_{t,x} = i, q_{t,x+1} \\ = h, q_{t+1,x} = j \,|\, O, \lambda)\end{aligned} \tag{22}$$

$$\begin{aligned}\alpha_i(t) = p\big(O_{1,x}, \ldots, O_{t,x}, O_{1,x-1}, \ldots, O_{t,x-1} \\ , O_{1,x+1}, \ldots, O_{t,x+1}, q_t = i|\lambda\big)\end{aligned} \tag{23}$$

$$\begin{aligned}\beta_i(t) = p(O_{t+1,x}, \ldots, O_{T,x}, O_{t+1,x-1}, \ldots, \\ O_{T,x-1}, O_{t+1,x+1}, \ldots, O_{T,x+1}|q_t = i, \lambda)\end{aligned} \tag{24}$$

## Algorithm

During the training process, model parameters are optimized in such a way to maximize the likelihood of the observed sequence. The Baum–Welch algorithm uses expectation maximization, where the likelihood of the observations is locally maximized by iteratively re-estimating the model parameters. The usual procedure for HMMs is followed for the calculation of our model's parameters, as described below.[38]

### Forward procedure

This is the probability of observing the partial observation sequence $\{o_1, o_2, \ldots, o_x\}$ with $q_x = i$

$$\alpha_i(x) = p(O_1, \ldots, O_x, q_x = i|\lambda) \tag{25}$$

The forward probability is calculated using an inductive algorithm in our work.

### Backward procedure

This is the probability of observing the partial observation sequence from $x+1$ to the end of the sequence, given $q_x = i$

$$\beta_i(x) = p(O_{x+1}, O_{x+2}, \ldots, O_X|q_x = i, \lambda) \tag{26}$$

The backward probability is calculated using an inductive algorithm in our work.

### Expectation equations

The probability of being in state $i$ at spatial location $x$, given the observations $O$ and the model parameters (collectively denoted $\lambda$) is given by

$$\gamma_i(x) = p(q(x) = i|O, \lambda) \tag{27}$$

The probability of being in state $i$ at spatial location $x, j$ at time $x+1$ is denoted as $\epsilon_{i,j}(x)$ and given by

$$\epsilon_{i,j}(x) = p(q_x = i, q_{x+1} = j \,|\, O, \lambda) \tag{28}$$

Variables of the expected equations are calculated based on the forward and backward variables in each iteration of the inductive algorithm.

## Evaluation

### Model training

The model is trained on a large video data set containing normal pedestrian activities. Observation sequences, each of length $T$, are created from the feature vectors of the spatial blocks of $T$ consecutive video frames for the Semi-2D HMM and Full-2D HMM, while observation sequences for Spatial

HMMs are created from the feature vectors of the spatio-temporal cubes aligned in the spatial directions. Created observation sequences are used to train the proposed HMM models. As mentioned above there are two instances of HMMs which are trained to capture both the horizontal and vertical spatial causality.

A large number of frames in the training video data results in a huge number of observations being created, thus making the computation process time consuming. To avoid this, observation sequences which don't have any motion information, i.e that have no foreground pixels, are filtered out. Filtering is done based on the number of foreground pixels in the particular sequence.[39] A sequence which contains fewer foreground pixels than a threshold is omitted from being added to the training process.

The number of states for the HMMs are chosen, and individual observations from all the created observation sequences are hard clustered initially using the K-Means++ algorithm, to find the initial parameters of the Gaussian distributions belonging to each state.[40] Then, the modified version of the Baum–Welch algorithm is used to train the model until it reaches convergence or until the maximum number of specified iterations is reached.

## Experimental evaluation

We have tested our algorithms with the publicly available UCSD datasets.[14] This video dataset contains bi-directional pedestrian traffic from two camera viewpoints. Several video sequences (each of 200 frames duration) which contain normal pedestrian movements are used for the training. The testing video sequences contain abnormalities, such as the presence of abnormal objects, anomalous pedestrian motions and spatial abnormalities, and are annotated with frame-level ground truth.

We use two different threshold values for our horizontal and vertical HMMs to detect the abnormal blocks and the frame is classified as abnormal if it contains an abnormal block. Detection from both HMMs in our algorithm is compared with the annotated ground truth at frame level and threshold values are varied to generate an ROC curve. Corresponding equal error rates (EER) and the area under the curve (AUC) are obtained.

In order to examine the exact effects of our proposed HMMs, we compare the performance of our methods (using vertical and horizontal configurations) with a regular HMM (1D) which does not capture spatial causalities. All other parameters are equal (block size $7 \times 7$, sequence length 20 frames). Results (EER and AUC) are shown in Table 1, and Figures 4–6 show the ROC curves for the Semi-2D, Full-2D and Spatial HMMs respectively. Table 1 shows that the vertical Semi-2D HMM performs better than the other HMMs and the one dimensional version of the proposed approach, and the vertical version of the Semi-2D HMM performs the best overall. In both training and testing videos, the majority of moving objects are humans and their height is larger than their width. So the motion information of humans is spread in the vertical direction rather than the horizontal direction. This results in adjacent locations in the horizontal direction having less useful motion information than the adjacent locations in the vertical direction, leading to the poor performance of the horizontal Semi-2D HMM when compared to the vertical Semi-2D HMM. The performance of the Full-2D HMM is good for the Ped1 dataset (close to that of the Semi-2D HMM), though it performs poorly for the Ped2 dataset, while Spatial HMM performs well

**Table 1.** Comparison of proposed 2D-HMM with regular HMM (1D). Different combinations of features are shown: 'ToF' stands for Textures of Optical Flow[13] and 'O/F' stands for Optical Flow based features (equations (1) and (2)).

| | | (Ped1) | | (Ped1) | |
|---|---|---|---|---|---|
| Classifier | Features | EER (%) | AUC (Ped2) | EER (%) | AUC |
| **Proposed Semi-2D-HMM (vertical)** | **ToF, O/F and location** | 27.64 | 0.780 | **11.67** | **0.928** |
| **Proposed Semi-2D-HMM (vertical)** | **O/F and location** | **21.68** | **0.859** | 16.62 | 0.883 |
| Proposed Semi-2D-HMM (horizontal) | ToF, O/F and location | 27.42 | 0.790 | 22.32 | 0.882 |
| Proposed Semi-2D-HMM (horizontal) | O/F and location | 22.79 | 0.816 | 31.18 | 0.702 |
| Proposed Full 2D-HMM (vertical) | ToF, O/F and location | 25.92 | 0.818 | 39.5 | 0.691 |
| Proposed Full 2D-HMM (vertical) | O/F and location | 25.76 | 0.827 | 36.12 | 0.654 |
| Proposed Full 2D-HMM (horizontal) | ToF, O/F and location | 25.32 | 0.821 | 31.94 | 0.753 |
| Proposed Full 2D-HMM (horizontal) | O/F and location | 25.32 | 0.821 | 27.26 | 0.788 |
| Proposed Spatial-HMM (vertical) | ToF, O/F and location | 30.79 | 0.760 | 14.64 | 0.916 |
| Proposed Spatial-HMM (vertical) | O/F and location | 33.85 | 0.735 | 19.91 | 0.881 |
| Proposed Spatial-HMM (horizontal) | ToF, O/F and location | 32.90 | 0.757 | 22.95 | 0.861 |
| Proposed Spatial-HMM (horizontal) | O/F and location | 36.75 | 0.705 | 25.23 | 0.846 |
| HMM (1D) | ToF, O/F and location | 30.12 | 0.780 | 16.2 | 0.921 |
| HMM (1D) | O/F and location | 22.42 | 0.831 | 31.18 | 0.716 |

with Ped2 dataset (close to the performance of the Semi-2D HMM), though it performs poorly for the Ped1 dataset.

Textures of optical flow features work well for Ped2 but not Ped1 as the poor resolution in the far field of Ped1 is poorly suited to textural type features, leading to poor detection in the far field and lower performance overall. Poor performance of Full-2D HMM in Ped2 dataset is due to lack of training data, and in future it will be tested with datasets containing sufficient data to evaluate accurate performance. Spatial HMMs perform well for Ped2 but not Ped1 due to poor resolution in the far field of Ped1. The Semi-2D HMM outperforms the Spatial HMM as it better represents both spatial and temporal dependencies. Reasons for the better performance of

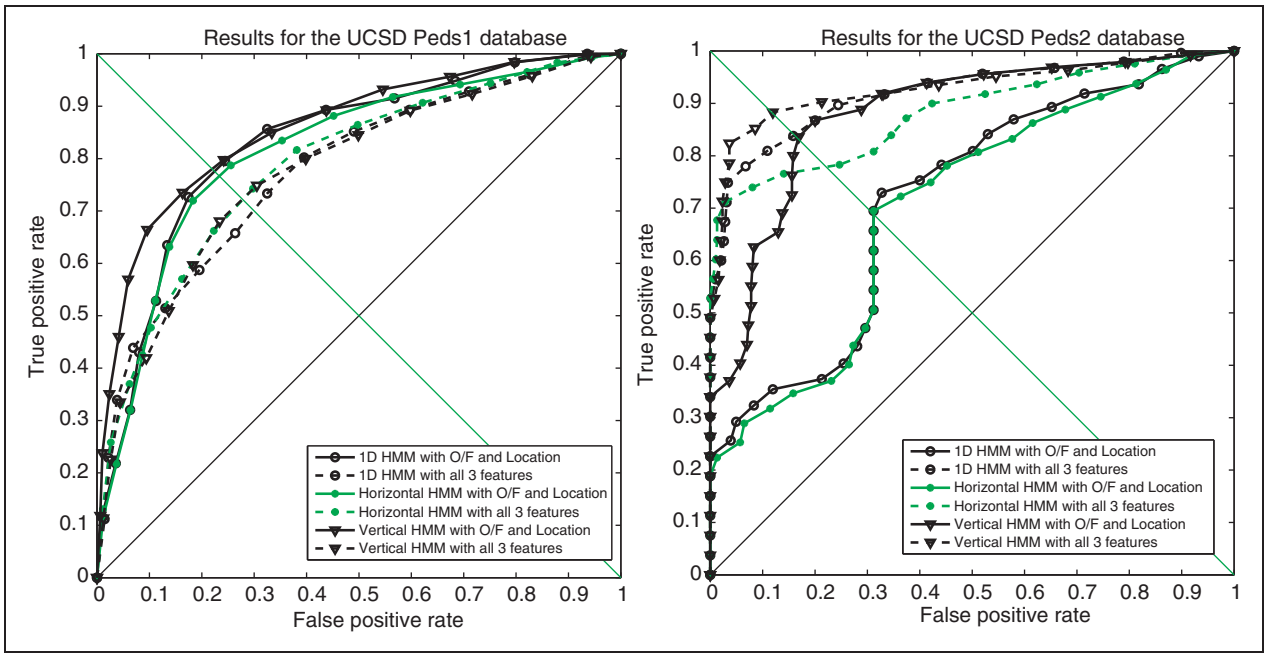


**Figure 4.** ROC curves of Ped1 and Ped2 of both Semi-2D HMMs with different feature combinations. (a) ROC curves of Ped1 of both Semi-2D HMMs with different feature combinations and (b) ROC curves of Ped2 of both Semi-2D HMMs with different feature combinations.

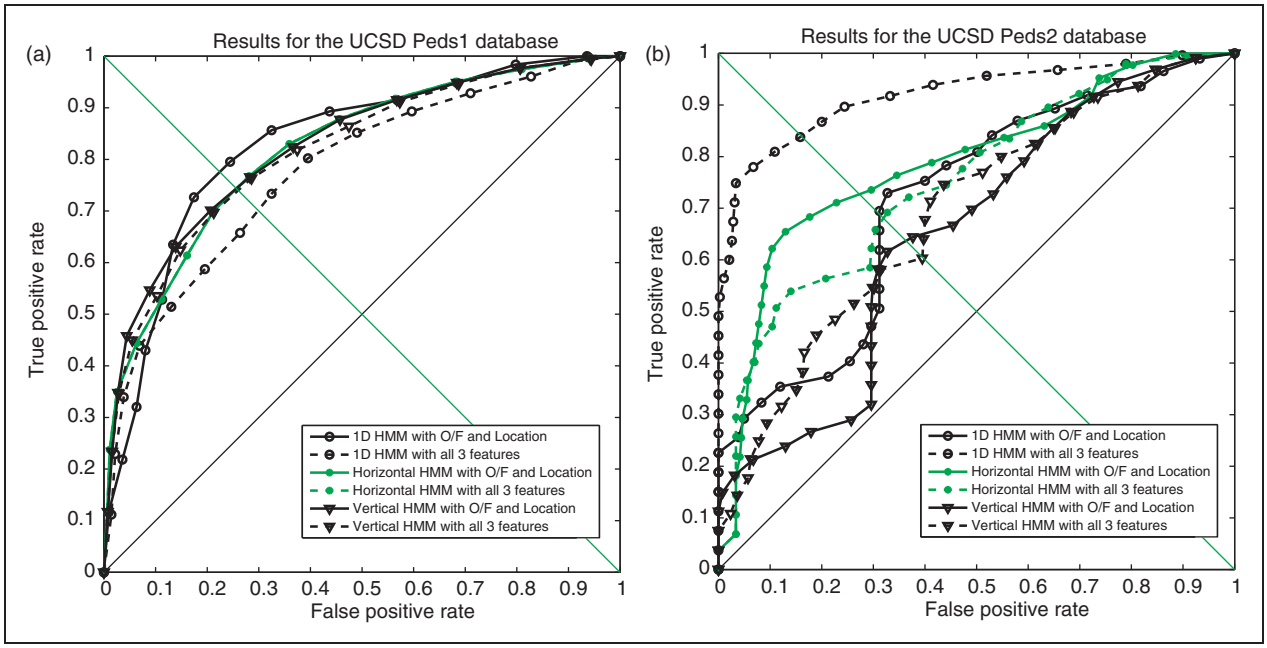


**Figure 5.** ROC curves of Ped1 and Ped2 of both Full-2D HMMs with different feature combinations. (a) ROC curves of Ped1 of both Full-2D HMMs with different feature combinations and (b) ROC curves of Ped2 of both Full-2D HMMs with different feature combinations.

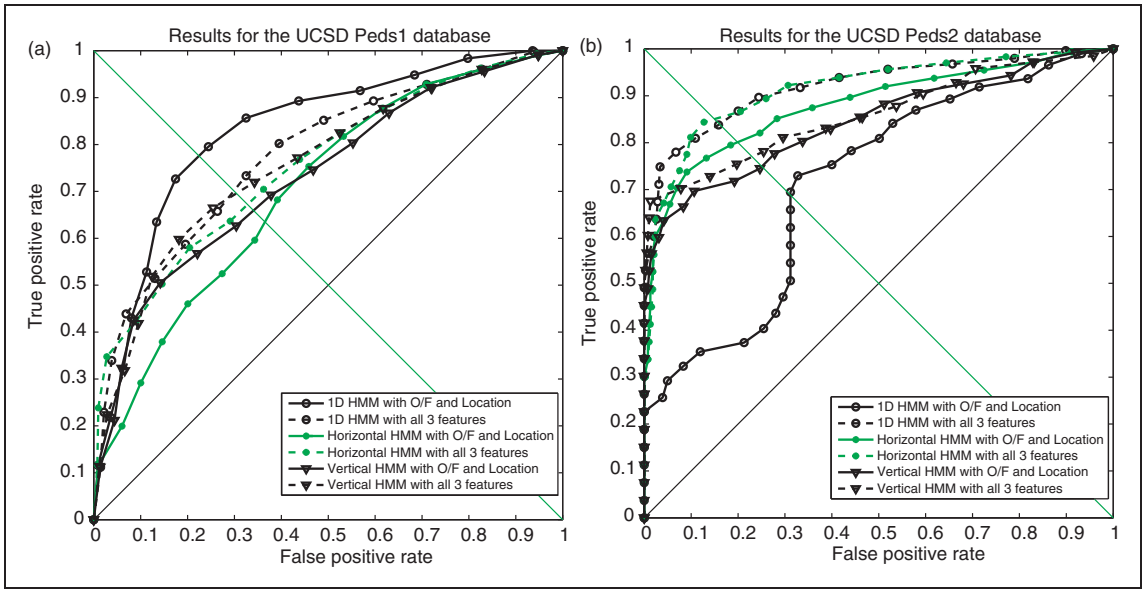


**Figure 6.** ROC curves of Ped1 and Ped2 of both Spatial HMMs with different feature combinations. (a) ROC curves of Ped1 of both Spatial HMMs with different feature combinations and (b) ROC curves of Ped2 of both Spatial HMMs with different feature combinations.

the Semi-2D HMM as compared to the Full-2D HMM will be investigated in the future.

The performance of our algorithm is compared with the outcomes of other previous work – the social force model,[28] the MPPCA model of optical flow,[27] the normalized combination of SF-MPPCA,[14] the pixel monitoring approach of Adam et al.,[29] mixture of dynamic textures,[14] textures of optical flow[13] and cell-based analysis of foreground speed, size and texture[32] – in Table 2. Values of the EER and AUC obtained by the above works are depicted in the table. EER for the Ped1 dataset from the above works lies between 22.5% and 40% while that of the Ped2 dataset lies between 12.7% and 42%.[13,32]

Our method's performance using the Semi-2D vertical HMM is also shown in Table 2. When all features are used, the method performs competitively with existing approaches, with an EER of 27.64% for Ped1 and 11.67% for Ped2. Omitting the textures of optical flow feature degrades performance slightly for the Ped2 dataset, but improves performance on Ped1 with an EER of 21.68%.

Our system performs well, detecting the anomalies such as bicycles of various speeds, vans, skateboarders, as well as spatial abnormalities and any combination of these anomalies. Figure 7 shows some video frames from both Ped1 and Ped2 datasets with blocks detected as containing anomalies highlighted.

Regarding the speed of our algorithm, on average it takes 0.09 s to process a frame (11 fps) on a computer with 2.53 GHz Intel i5 processor and 4 GB memory, running in a single threaded configuration, making the algorithm suitable for real-time deployment (the UCSD dataset is captured at 10 fps).

**Table 2.** Performance on the UCSD datasets.[14] Equal error rate (EER) is reported. ToF stands for Textures of Optical Flow[13] and O/F stands for Optical Flow based features (equations (1) and (2)).

| | EER (%) | |
|---|---|---|
| System | Ped1 | Ped2 |
| SF[28] | 31 | 42 |
| MPPCA[27] | 40 | 30 |
| SF-MPPCA[14] | 32 | 36 |
| Adam[29] | 38 | 42 |
| MDT[14] | 25 | 25 |
| Ryan[13] | 23.1 | 12.7 |
| Reddy[32] | 22.5 | 20 |
| Proposed Semi-2D HMM (With ToF, O/F and Location) | 27.64 | **11.67** |
| Proposed Semi-2D HMM (With O/F and Location) | **21.68** | 16.62 |

## *Application to railway crossings*

The models proposed in this work can be used to model the pedestrian activities at a railway level crossing. Temporal causalities and spatial causalities can both be captured by the proposed approaches. Models can be trained from video footage containing the normal activities of the pedestrians, and, during the real-time monitoring of the surveillance video, outliers of the learned models will be detected as anomalies.

At a railway level crossing we can define two main contexts relating to pedestrian activities: time periods

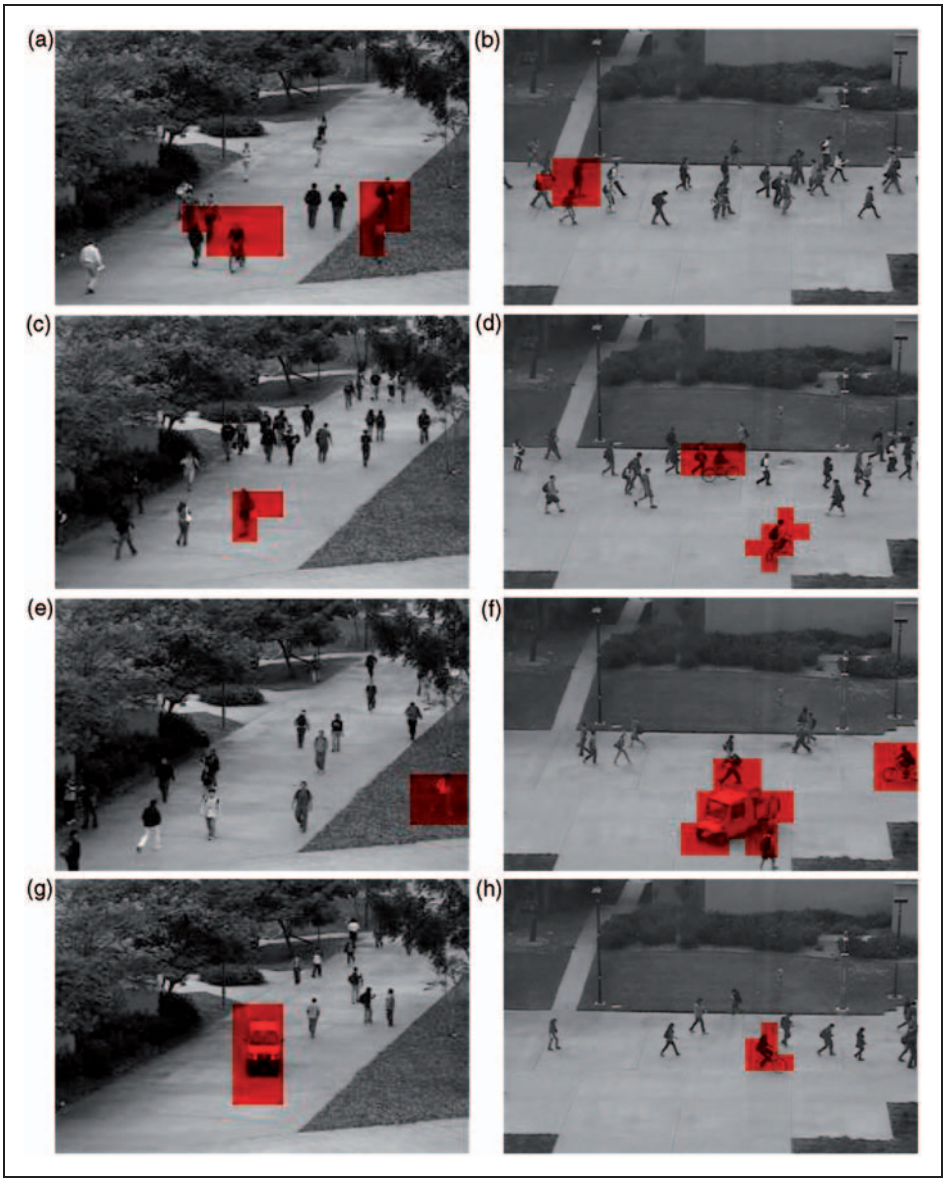


**Figure 7.** Representative frames demonstrating the proposed anomaly detection algorithm. The left column is from dataset 'Ped1' and the right column is from 'Ped2'.[14] (a) Bicycle (bottom center) and spatial anomaly (bottom right), (b) Skateboard is detected, (c) Skateboard is detected, (d) Two bicycles are detected, (e) Spatial abnormality, (f) Vehicle (centre) and bicycle (right) are detected, (g) Vehicle is detected, and (h) Bicycle moving slowly is detected.

during the arrival of the train and all other time periods.

During the arrival of the train, rail gates are closed and people are not supposed to move through the level crossings. In this context, the presence of any human object should be detected as an anomaly. Using our developed model, a spatial anomaly can be located in the scene as shown in Figure 7(e). As the proposed system is capable of real-time operation, an alarm can be flagged as soon as such an event is detected. This alarm could result in a variety of measures, including systems to alert both pedestrians and railway guards, and potentially even alerting the train driver of the potential danger (ultimately these measures would depend on what other subsystems were present). While the utility of this alarm may vary from site to site and event to event (i.e. if a pedestrian enters the tracks at the same moment as the train arrives there is little that can be done), it is reasonable to expect it to be of benefit in environments where there is a delay between the crossing being closed and trains arriving (for instance, in Queensland, Australia, level crossings close 1–2 m before a train arrives). Furthermore, the act of forcing or climbing over the gate is also likely to trigger an alarm as it constitutes abnormal behaviour, thus potentially offering a few extra seconds warning.

During the second context (that is, the rest of the time period where trains are not supposed to arrive and gates are opened so that pedestrians can cross the track), anomalies may include a person running, a person falling and the presence of other abnormal objects such as bicycles or vehicles. Using our

developed model, anomalies such as these can be detected (see Figure 7(c), a speed related anomaly caused by a skateboarder is detected; and Figure 7(g), where a vehicle is detected). In this second context, the proposed approach could be particularly useful for detecting incidents involving people with a disability, children and elderly people. Furthermore, the proposed approach would also aid in gathering statistics of events and near-misses. Such data could be helpful in awareness programs related to railway pedestrian safety, or in identifying ways to further improve safety.

The context could be switched based on the traffic signals, i.e. the first context can be enabled when a train arrives and the other context can be set by default. Future work will focus on the development of models that enable event detection in multiple contexts, as well as evaluations of abnormal event detection with railway level-crossing video footage.

## Conclusion and future work

We have proposed new variants of 2D HMM techniques for anomaly detection. These approaches capture both the temporal and spatial causality of a training sequence, and the Semi-2D HMM performs well when detecting the anomalies compared to other state of the art algorithms as well as the equivalent 1D HMM in terms of accuracy and speed. The Full-2D HMM and Spatial HMM also perform well in certain scenarios, though are less consistent than the Semi-2D HMM due to lack of training data and lack of ability to model both the temporal and spatial causalities respectively. Further reasons will be investigated in the future.

The proposed methods can be used to capture abnormal events such as a person falling, a person running, the presence of skate boarders and presence of other vehicles at the railway level crossings. Further, these methods can be used to study normal behaviour on rail crossings in different contexts.

Future work will involve investigating different features and combinations of features, as well as evaluations of other datasets, specifically railway level-crossing data. Multi-context models and the automatic detection of different contexts will also be investigated.

### Funding

This work was supported by the CRC for Rail Innovation (established and supported under the Australian Government's Cooperative Research Centres program). (Research title: Video based detection of normal and anomalous behaviour of individuals.)

### References

1. Ko T. A survey on behavior analysis in video surveillance for homeland security applications. In: *37th IEEE applied imagery pattern recognition workshop (AIPR '08)*, Washington DC, 15–17 October 2008, pp.1–8. IEEE Computer Society.
2. Popoola OP and Wang K. Video-based abnormal human behavior recognition – A review. *IEEE T Syst Man Cy C* 2012; 42(6): 865–878. New York, USA: ACM.
3. Lavee G, Khan L and Thuraisingham B. A framework for a video analysis tool for suspicious event detection. *Multimed Tools Appl* 2007; 35: 109–123. MA, USA: Kluwer Academic Publishers Hingham.
4. Zhang Y and Liu Z. Irregular behavior recognition based on treading track. In: *International conference on wavelet analysis and pattern recognition (ICWAPR '07)*, Beijng, 2–4 November 2007, Vol. 3, pp.1322–1326. Heidelberg, Berlin: Springer.
5. Wiliem A, Madasu V, Boles W and Yarlagadda P. Detecting uncommon trajectories. In: *Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, ACT, 1–3 December 2008, pp.398–404. Washington DC, USA: IEEE Computer Society.
6. Zhong H, Shi J and Visontai M. Detecting unusual activity in video. In: *IEEE conference on computer vision and pattern recognition*, 2004, Vol. 2, pp. 819–826. Washington DC, USA: IEEE Computer Society.
7. Hung Y, Chiang C, Hsu SJ and Chan C. Abnormality detection for improving elder's daily life independence. In: *8th international conference smart homes health telematics*, 2010, pp.186–194. Heidelberg, Berlin: Springer.
8. Kratz L and Nishino K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: *IEEE conference on computer vision and pattern recognition*, 2009, pp.1446–1453. IEEE.
9. Andrade E, Blunsden S and Fisher R. Hidden Markov Models for optical flow analysis in crowds. In: *18th international conference on pattern recognition (ICPR 2006)*, Hong Kong, 20–24 August 2006, Vol. 1 pp.460–463.
10. Andrade E, Blunsden S and Fisher R. Modelling crowd scenes for event detection. In: *18th international conference on pattern recognition*, Hong Kong, 2006, Vol. 1, pp.175–178. Washington DC, USA: IEEE Computer Society.
11. Kratz L and Nishino K. Spatio-temporal motion pattern modeling of extremely crowded scenes. In: *1st international workshop on machine learning for vision based motion analysis (MLVMA '08)*, Marseilles, France, 2008, pp. 263–274. London, Springer.
12. Utasi A and Czuni L. Detection of unusual optical flow patterns by multilevel hidden Markov models. *Opt Eng* 2010; 49(1): 017201. SPIE publications.
13. Ryan D, Denman S, Fookes C and Sridharan S. Textures of optical flow for real-time anomaly detection in crowds. In: *8th IEEE international conference on advanced video and signal based surveillance*, 2011, pp.230–235.
14. Mahadevan V IEEE Computer Society. Li W, Bhalodia V and Vasconcelos N. Anomaly detection in crowded scenes. In: *IEEE conference on computer vision and pattern recognition*, 2010, pp.1975–1981. IEEE. Available at: http://www.svcl.ucsd.edu/projects/anomaly/
15. Anderson D, Luke RH, Keller JM, Skubic M, Rantz M and Aud M. Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Comput Vis Image Und* 2009; 113(1): 80–89. New York, USA: Elsevier Science Inc.

16. Jiang F, Yuan J, Tsaftaris SA and Katsaggelos AK. Anomalous video event detection using spatiotemporal context. *Comput Vis Image Und* 2011; 115(3): 323–333.
17. Zhou Y, Yan S and Huang TS. Detecting anomaly in videos from trajectory similarity analysis. In: *IEEE International Conference on Multimedia and Expo*, Beijing, 2–5 July 2007, pp.1087–1090. IEEE.
18. Hu W, Xiao X, Fu Z, Xie D, Tan T and Maybank S. A system for learning statistical motion patterns. *IEEE T Pattern Anal* 2006; 28: 1450–1464. Washington DC, USA: IEEE Computer Society.
19. Morris BT and Trivedi MM. Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach. *IEEE T Pattern Anal* 2011; 33(11): 2287–2301. Washington DC, USA: IEEE Computer Society.
20. Vasquez D, Fraichard T and Laugier C. Incremental learning of statistical motion patterns with growing hidden Markov models. *IEEE T Intell Transp* 2009; 10(3): 403–416. New Jersey, USA: IEEE Press Piscataway.
21. Xiang T and Gong S. Beyond tracking: Modelling activity and understanding behaviours. *Int J of Comp Vis* 2006; 67(1): 21–51. Kluwer Academic Publishers.
22. Zaharescu A and Wildes RP. Anomalous behaviour detection using spatiotemporal oriented energies, subset inclusion histogram comparison and event-driven processing. In: *Proceedings of the 11th European conference on computer vision: Part I (ECCV '10)*, 2010. Heidelberg, Berlin: Springer.
23. Zhao B and Xing EP. Online detection of unusual events in videos via dynamic sparse coding. In: *24th IEEE conference on computer vision and pattern recognition (CVPR '11)*, Colorado Springs, CL, 20–25 June 2011, pp.3313–3320. Washington DC, USA: IEEE Computer Society.
24. Borzin A, Rivlin E and Rudzsky M. Surveillance event interpretation using generalized stochastic petri nets. In: *8th international workshop on image analysis for multimedia interactive services (WIAMIS '07)*, 2007. IEEE Computer Society.
25. Tipping M and Bishop C. Mixtures of probabilistic principal component analyzers. *Neural Comput* 1999; 11(2), 443–482. MA, USA: MIT Press Cambridge.
26. Xiaogang W, Xiaoxu M and Grimson W. Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. *IEEE T Pattern Anal* 2009; 31(3): 539–555. Washington DC, USA: IEEE Computer Society.
27. Kim J and Grauman K. Observe locally, infer globally: A space–time MRF for detecting abnormal activities with incremental updates. In: *IEEE conference on computer vision and pattern recognition*, 2009, pp.2921–2928. IEEE.
28. Mehran R, Oyama A and Shah M. Abnormal crowd behavior detection using social force model. In: *IEEE conference on computer vision and pattern recognition*, 2009, pp.935–942. IEEE.
29. Adam A, Rivlin E, Shimshoni I and Reinitz D. Robust realtime unusual event detection using multiple fixed-location monitors. *IEEE T Pattern Anal* 2008; 30(3): 555–560. Washington DC, USA: IEEE Computer Society.
30. Hamid R, Johnson A, Batta S, Bobick A, Isbell C and Coleman G. Detection and explanation of anomalous activities: Representing activities as bags of event n-grams. In: *IEEE conference on computer vision and pattern recognition*, 2005, Vol. 1, pp.1031–1038. Washington DC, USA: IEEE Computer Society.
31. Greenspan H, Goldberger J and Mayer A. Probabilistic space–time video modeling via piecewise GMM. *IEEE T Pattern Anal* 2004; 26(3): 384–396. Washington DC, USA: IEEE Computer Society.
32. Reddy V, Sanderson C and Lovell BC. Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture. In: *IEEE conference on computer vision and pattern recognition*, 2011, pp.55–61.
33. Zhang D, Gatica-Perez D, Bengio S and McCowan I. Semisupervised adapted HMMs for unusual event detection. In: *emphIEEE conference on computer vision and pattern recognition*, 2005, Vol. 1, pp.611–618. Washington DC, USA: IEEE Computer Society.
34. Snoek J, Hoey J, Stewart L and Zemel RS. Automated detection of unusual events on stairs. In: *3rd Canadian Conference on Computer and Robot Vision*, 2006, Vol. 27, p. 5. MA, USA: Butterworth-Heinemann Newton.
35. Jiang F, Wu Y and Katsaggelos AK. A dynamic hierarchical clustering method for trajectory-based unusual video event detection. *IEEE T Image Process*, 2009, Vol. 18(4), pp.907–913. Piscataway, New Jersey, USA: IEEE Press.
36. Black MJ and Anandan P. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Comput Vis Image Und* 1996; 63(1): 75–104. New York, USA: Elsevier Science Inc.
37. Ma X, Schonfeld D and Khokhar A. A general two-dimensional hidden Markov model and its application in image classification. In: *IEEE International Conference on Image Processing (ICIP 2007)*, 2007, pp.41–44. IEEE.
38. Bilmes J. A gentle tutorial on the em algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.613 (1997).
39. Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. In: *17th International Conference on Proceedings of the Pattern Recognition (ICPR '04)*, 2004, Vol. 2, pp.28–31. Washington DC, USA: IEEE Computer Society.
40. Arthur D and Vassilvitskii S. k-means++: the advantages of careful seeding. In: *18th annual ACMSIAM symposium on discrete algorithms (SODA '07)*, Philadelphia, PA, 2007, pp.1027–1035. PA, USA: Society for Industrial and Applied Mathematics Philadelphia.