Anomalous Event Detection using a Semi-Two Dimensional Hidden Markov Model

Hajananth Nallaivarothayan, David Ryan, Simon Denman, Sridha Sridharan and Clinton Fookes
Image and Video Research Laboratory,
Queensland University of Technology,
GPO Box 2434, 2 George St.
Brisbane, Queensland 4001.
{*h.nallaivarothayan*, *david.ryan*, *s.denman*, *s.sridharan*, *c.fookes*}@qut.edu.au

***Abstract*—The rapid increase in the deployment of CCTV systems has led to a greater demand for algorithms that are able to process incoming video feeds. These algorithms are designed to extract information of interest for human operators. During the past several years, there has been a large effort to detect abnormal activities through computer vision techniques. Typically, the problem is formulated as a novelty detection task where the system is trained on normal data and is required to detect events which do not fit the learned 'normal' model. Many researchers have tried various sets of features to train different learning models to detect abnormal behaviour in video footage. In this work we propose using a Semi-2D Hidden Markov Model (HMM) to model the normal activities of people. The outliers of the model with insufficient likelihood are identified as abnormal activities. Our Semi-2D HMM is designed to model both the temporal and spatial causalities of the crowd behaviour by assuming the current state of the Hidden Markov Model depends not only on the previous state in the temporal direction, but also on the previous states of the adjacent spatial locations. Two different HMMs are trained to model both the vertical and horizontal spatial causal information. Location features, flow features and optical flow textures are used as the features for the model. The proposed approach is evaluated using the publicly available UCSD datasets and we demonstrate improved performance compared to other state of the art methods.**

I. INTRODUCTION

As a result of an increased focus on security, as well as advances in the electronics and semi-conductor industry, there are a growing number of CCTV cameras in public places. Continuous operation of numerous cameras in a single place has resulted in a huge amount of video data to be processed. This large amount of data makes it very difficult, if not impossible, for human operators to effectively observe and detect all events in the video footage in real time. Hence there is significant interest in developing computer vision technologies as a solution to these problems.

Event detection using computer vision technologies has been an active research topic for several years. Anomalous event detection is a sub-category in the field 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. It is possible that we may need to identify an anomalous event when it appears, despite the fact that it has never occurred before [12].

The features that are extracted, and the models which are used to classify the extracted features as normal or abnormal, are the two 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 to model the temporal behaviour [3], [2], [15], the modelling of spatial causality is omitted in all but a minority of systems [14], [24].

In this paper we propose a Semi Two Dimensional Hidden Markov Model to model both the temporal and spatial causalities. This 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. Two model structures are investigated, modeling the causalities in either the vertical or horizontal direction. Within the HMM, outliers are detected to locate abnormal events. The proposed approach uses features extracted from spatio-temporal blocks. 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 [22] to capture the anomalies related to the motion characteristics.

The remainder of this paper is structured as follows: Section II summarises related work in this field; Section III describes the features used in the proposed algorithm; Section IV describes the Semi-2D HMM algorithm; Section V presents experimental results on the publicly available USCD database [17]; and Section VI presents conclusions and directions for future work.

II. RELATED WORK

Abnormal event detection falls into two categories, bottom-up and top-down approaches. In the context of event detection, a top-down approach segments each individual in the scene and features are extracted separately. Anomalous event detection using object tracking is an example of this approach, where individuals' object trajectories are obtained and those with abnormal trajectories are deemed to be performing an abnormal event.

Among the trajectory analysis works, Zhou *et al*. [30] groups similar trajectories using the Edit Distance (ED). Hu *et al*. [11] associates foreground pixel masks with extracted trajectories, providing a more descriptive representation of the activities than trajectories alone. Morris *et al*. [19] represents trajectories by a series of flow vectors. Like Zhou [30], similar trajectories are grouped together, and a HMM is trained to represent each characteristic trajectory. 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 underling scene characteristics and crowd behaviour. Point-based, block based and patch-based methods are the main three categories of this approach. These approaches can work very well in densely crowded environments amidst extensive clutter and dynamic occlusions. Features which can be used include the pixel location, pixel intensity, intensity changes, velocities, motion textures and other low level features [20].

In bottom-up approaches, features can be extracted at a pixel level, block level or patch level. Xiang *et al*. [25] has proposed the Pixel Change History (PCH) for measuring multi-scale temporal changes at each pixel. Andrei *et al*. [27] have used distributions of spatio-temporal oriented energy. Andrade *et al*. [4] used optical flow patterns, and spatial histograms of the detected objects are used as the feature by Zhong *et al*. [29]. Zhao *et al*. [28] used histograms of gradients (HoG) and histograms of optical flow (HoF). 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*. [22] proposed a visual representation called textures of optical flow, which captures both the smoothness of the flow and the presence of motion. This may be useful for detecting bicycles or vehicles in a pedestrian scene, for example the UCSD dataset [17].

The various low level features serve as the input to a learning model. Popular learning models include HMM [24], [3], [2], [15], [14], Petri net [8], LDA [23], [26] and Markov Random Field (MRF) [13].

Hierarchical Bayesian Models are used by Wang *et al*. [26] to detect anomalies in crowded scenes. Similarly Mehran *et al*. [18] uses LDA and a bag of words methodology to learn a 'normal' model, after which grames can be classified as either abnormal or normal. Adam *et al*. [1] uses histogram binning of the extracted features, while anomaly detection is done by using a cyclic buffer to determine the likelihood of new observations. Kim *et al*. [13] uses a mixture of probabilistic principal component analyzers to model their features. Hamid *et al*. [10] represents activities as bags of event n-grams where global structural information of activities is analyzed using local event statistics. Zhao *et al*. [28] 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. Further, Ryan *et al*. [22] and Greenspan *et al*. [9] utilized GMMs for their feature modeling while Zhong *et al*. [29] used K-means clustering to group the video segments into disjoint sets. Mahadevan *et al*. [17] uses a generative mixture of dynamic textures. Vikkas *et al*. [21] models the motion and size features by an approximated version of kernel density estimation and the texture features by an adaptively grown codebook.

These 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 inter-relationship between individual observations. This will result in important information relating to the pattern and duration of the normal activities, making the detection of abnormalities more challenging.

Hidden Markov Models (HMMs) provide a means to capture temporal dependencies within the detection process. Andrade [3], [2] uses a bank of Hidden Markov Models trained on normal behaviours, and detects a sequence as anomalous when the likelihood falls below a threshold. Kratz *et al*. [14] proposes the same distributed location-based HMMs but with different features and an alternative state clustering technique and emission probability distributions. These models only capture the causality in the temporal direction while the information about the adjacent behaviour is missed.

Kratz *et al*. [14] also used coupled HMMs to capture the spatial relationships. 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 *et al*. [24] constructs their models at two levels, a region-based continuous distribution HMM, and a higher layer HMM to inter-link those regional HMMs that form the first level. Here, spatial information is missing at the low level HMMs and only limited spatial causality can be captured by the high level discrete HMM.

III. Features Extraction

We use three features within the proposed system:

1) Location features (center coordinate of a spatial block) to detect the location-specific anomalies.
2) Motion information (summation of optical flow vectors inside a block) to identify the anomalies related to speed of movements of the objects.

3) Textures of optical flow [22] to identify the anomalies related to the type of motion that is occurring. 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 Section III-A.

To calculate the optical flow vectors, we have used Black and Anandan's algorithm [7]. 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 measure the uniformity of 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 [22].

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

1) All three features: textures of optical flow (ToF) at various scales $\{\phi\}$, motion information (σ_u, σ_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 uniformity feature value at δ offset [22].

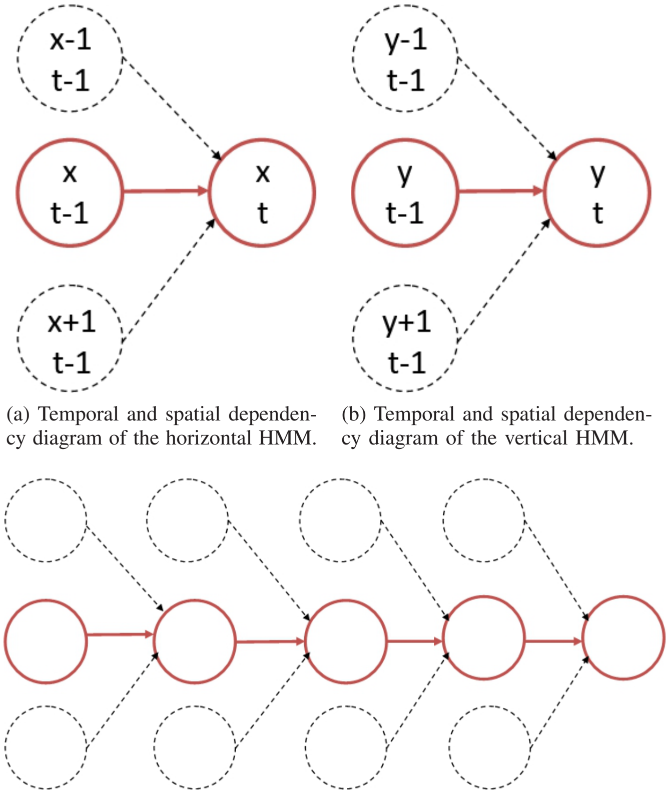
2) Optical flow vectors and location features alone,

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

A. 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 blocks of different configurable sizes. Features are extracted using each pixel within a block, and are summed to form the feature vector for the block. 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.



(a) Temporal and spatial dependency diagram of the horizontal HMM.

(b) Temporal and spatial dependency diagram of the vertical HMM.

(c) Overall sequence diagram of a Semi 2D HMM.

Fig. 1: Schematic diagrams of the proposed HMM.

The size of the block (7 X 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 [22]. The sequence length is chosen as 20 frames.

IV. Semi Two Dimensional Hidden Markov Models

We propose a semi-two dimensional HMM to model the extracted observation sequences from the training video, and to detect abnormalities. Generally, Hidden Markov Models 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 Hidden Markov Models have been proposed leading to alternate HMM-type models such as the Multi Level HMM [24], and coupled HMMs [14]. In the field of image classification, a form of two dimensional Hidden Markov Models has been used to capture the spatial causality of images in both vertical and horizontal directions [16]. However for a video task, these 2D HMMs create too many observation sequences in different directions, making it 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 1a) or vertically (Figure 1b).

A. Assumptions of our Hidden Markov Model

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 1c the sequence drawn in red is the main observation sequence).
3) Adjacent spatial observations in one sequence are part of another main temporal sequence.

B. Parameters of the Hidden Markov Model

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 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. Our model is based on the following parameters.

1) 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 horizontal direction are considered in case of the horizontal HMM, and adjacent locations in vertical direction are considered for vertical HMM. The transition probability for the horizontal case is,

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

2) Gaussian Distribution Parameters for Likelihood of Observations : The likelihood of an observation which belongs to a state j is a Gaussian distribution with mean μ_j and covariance matrix Σ_j. The probability of an observation at time t, given that the state is j, is given by,

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

3) Initial Probabilities: The initial probability of observing state i is denoted by π_j,

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

C. Algorithm

During the training process model parameters are optimised 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 [6] is slightly modified for the calculation of our model's parameters, as described below, with the remainder of the procedure remaining unchanged.

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

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

The forward probability is calculated using an inductive algorithm as follows.

a) Initialization:

$$\alpha_i(1) = \pi_i(1) b_i(O_1), \tag{9}$$

where i is the state number.

b) Induction: Equation 10 states the induction step for calculating the forward probability, where j is the current state, i is the previous state and g,h are the previous states of the adjacent spatial locations $x-1$ and $x+1$ respectively.

2) 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 $q_t = i$,

$$\beta_t(i) = p(O_{t+1}, O_{t+2}, ..., O_T, O_{t,x-1}, O_{t,x+1} | q_t = i, \lambda). \tag{11}$$

The backward probability is calculated using an inductive algorithm as follows.

a) Initialization:

$$\beta_i(T) = 1, \tag{12}$$

where i is the state number and T is the sequence length.

b) Induction: Equation 13 states the induction step for calculating the backward probability, where i is the state at time t, j is the state at time $t+1$ and g,h are the states at time t in adjacent spatial locations $x-1$, $x+1$ respectively.

3) Expectation equations: The probability of being in state i at time t, given the observations O and the model parameters (collectively denoted λ) is given by,

$$\gamma_i(t) = p(q(t) = i | O, \lambda), \tag{14}$$

$$= \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{N} \alpha_j(t)\beta_j(t)}. \tag{15}$$

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 the relevant formulas are given in Equations 16 and 17.

$$\alpha_{j+1}(t+1) = [\sum_{i=1}^{N} \sum_{g=1}^{N}, \sum_{h=1}^{N} \alpha_i(t) a_{g,i,h,j} b_g(O_{t,x-1}) b_h(O_{t,x+1})] b_j(O_{t+1,x}), \tag{10}$$

$$\beta_i(t) = \sum_{j=1}^{N} \sum_{g=1}^{N} \sum_{h=1}^{N} a_{g,i,h,j} b_j(O_{t+1,x+1}) \beta_j(t+1) b_g(O_{t,x-1}) b_h(O_{t,x+1}), \tag{13}$$

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

$$\epsilon_{g,i,h,j}(t) = \frac{\alpha_i(t) a_{g,i,h,j} b_g(O_{t,x-1}) b_h(O_{t,x+1}) b_j(O_{t+1,x}) \beta_j(t+1)}{\sum_i^N \sum_g^N \sum_h^N \sum_k^N \alpha_i(t) a_{g,i,h,j} b_g(O_{t,x-1}) b_h(O_{t,x+1}) b_k(O_{t+1,x}) \beta_k(t+1)}. \tag{17}$$

D. Training of the model

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 blocks of T consecutive video frames, and are used to train this 2D HMM model. 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 are filtered out. Filtering is done based on the number of foreground pixels [31] in the particular sequence. A sequence which contains less 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 [5], to find the initial parameters of the Gaussian distributions belonging to each state. 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.

V. Experimental Results

We have tested our algorithm with the publicly available UCSD datasets [17]. This video dataset contains bi-directional pedestrian traffic from two camera view points. Several video sequences (each of 200 frames duration) which contain normal pedestrian movements are used for the training. The testing video sequences which contains abnormalities, such as the presence of abnormal objects, anomalous pedestrian motions and spatial abnormalities are annotated with frame-level ground truth.

We use two different threshold values for our horizontal and vertical versions of HMM 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.

System	EER	
	Ped 1	Ped 2
SF [18]	31%	42%
MPPCA [13]	40%	30%
SF-MPPCA [17]	32%	36%
Adam [1]	38%	42%
MDT [17]	25%	25%
Ryan [22]	23.1%	12.7%
Vikas [21]	22.5%	20%
Proposed method (With ToF, O/F and Location)	27.64%	**11.67**%
Proposed method (With O/F and Location)	**21.68**%	16.62%

TABLE I: Performance on the UCSD datasets [17]. Equal error rate (EER) is reported. ToF stands for Textures of Optical Flow [22] and O/F stands for Optical Flow based features (equations 1-2).

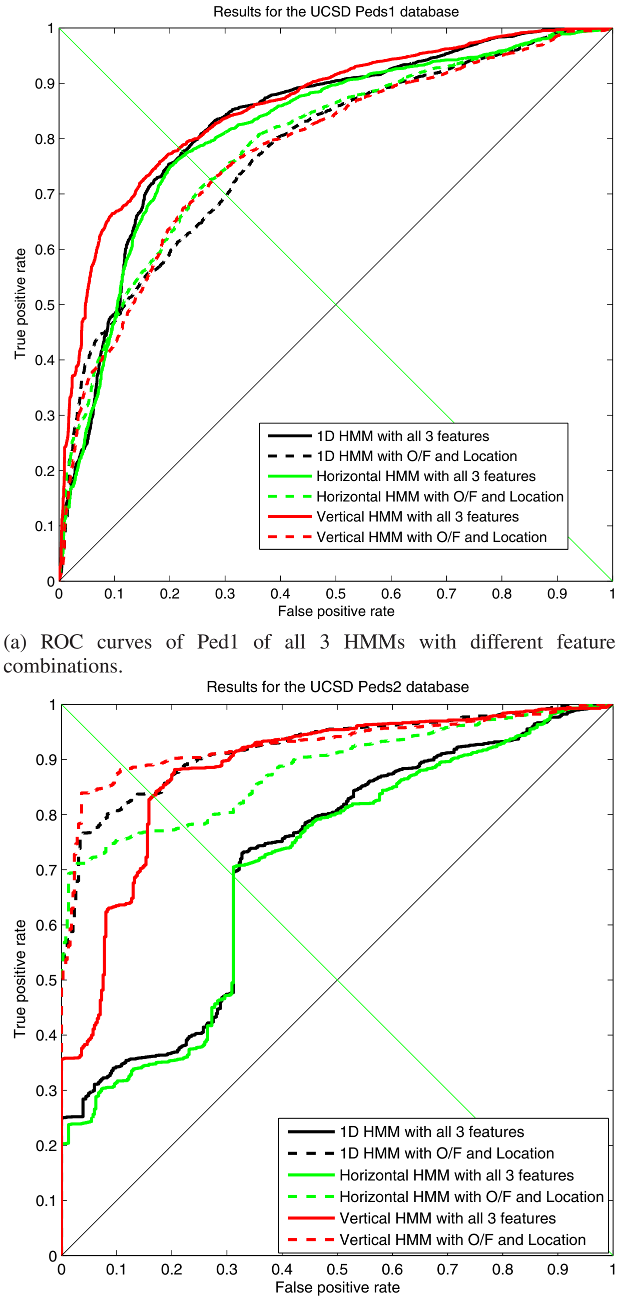
The performance of our algorithm is compared with the outcomes of the other previous work: social force model [18], the MPPCA model of optical flow [13], the normalized combination of SF-MPPCA [17], the pixel monitoring approach of Adam [1], mixture of dynamic textures [17], textures of optical flow [22] and Cell-based Analysis of Foreground Speed, Size and Texture [21] in Table I. Values of the EER and AUC obtained by the above works are depicted in the table. Equal error rate for the Ped1 dataset from the above works lies between 22.5 - 40% while that of the Ped2 dataset lies between 12.7 - 42% [22], [21].

Our method's performance using the vertical HMM is also shown in Table I. 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 (ToF) degrades performance slightly for the Ped2 dataset, but improves performance on Ped1 with an EER of 21.68%.

In order to examine the exact effects of the 2D HMM, we compare the performance of our method (using vertical and horizontal configurations) with a regular HMM (1D) which does not capture spatial causalities. All other parameters are equal (block size 7 × 7, sequence length 20 frames). Table

Classifier	Features	EER (Ped1)	AUC (Ped1)	EER (Ped2)	AUC (Ped2)
Proposed 2D-HMM (Vertical)	ToF, O/F and Location	27.64%	0.780	11.67%	0.928
Proposed 2D-HMM (Vertical)	O/F and Location	21.68%	0.859	16.62%	0.883
Proposed 2D-HMM (Horizontal)	ToF, O/F and Location	27.42%	0.790	22.32%	0.882
Proposed 2D-HMM (Horizontal)	O/F and Location	22.79%	0.816	31.18%	0.702
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

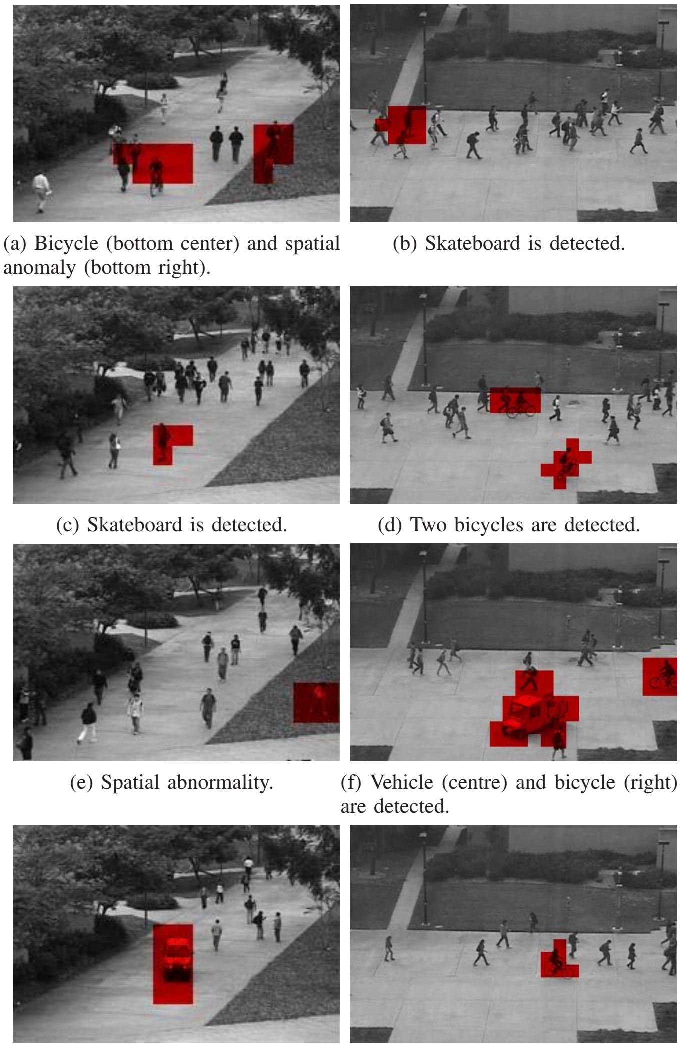
TABLE II: Comparison of proposed 2D-HMM with regular HMM (1D). Different combinations of features are shown: ToF stands for Textures of Optical Flow [22] and O/F stands for Optical Flow based features (equations 1-2).



(a) ROC curves of Ped1 of all 3 HMMs with different feature combinations.

(b) ROC curves of Ped2 of all 3 HMMs with different feature combinations.

Fig. 2: ROC curves of Ped1 and Ped2 of all 3 HMMs with different feature combinations



(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.

(h) Bicycle moving slowly is detected.

Fig. 3: Representative frames demonstrating the proposed anomaly detection algorithm. The left column is from dataset 'Ped1' and the right column is from 'Ped2' [17].

II shows that the vertical HMM performs better than the horizontal HMM and the one dimensional version of the proposed approach. In both training and testing videos moving objects are mostly humans and their height is larger than their width. So the motion information of humans is spread more in the vertical direction than in 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 HMM when compared to the vertical. The ROC curves of all the three HMMs with different feature combinations are depicted for both UCSD Ped1 and UCSD Ped2 datasets separately in Figures 2a and 2b respectively.

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 3 shows some video frames from both Ped1 and Ped2 datasets with blocks detected as containing anomalies highlighted in red.

Regarding the speed of our algorithm, on average it takes 0.09 sec 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.

VI. Conclusion and Future Work

We have proposed a new Semi-Two Dimensional Hidden Markov Model technique for anomaly detection. This approach captures both the temporal and spatial causality of a training sequence and 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.

Future work will involve implementing a full two dimensional Hidden Markov Model to capture the full spatial and temporal causality information in the video. Different features and combinations of features will be tried, as well as evaluations on other datasets.

References

[1] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 3, pp. 555 –560, mar. 2008.

[2] E. Andrade, S. Blunsden, and R. B. Fisher, "Hidden markov models for optical flow analysis in crowds," *ICPR (1) 2006*, pp. 460–463, 2006.

[3] E. Andrade, S.Blunsden, and R.Fisher, "Modelling crowd scenes for event detection," *ICPR 2006*, vol. 1, p. 4, 2006.

[4] E. L. Andrade, S. Blunsden, and R. B. Fisher, "Hidden markov models for optical flow analysis in crowds," *ICPR (1) 2006*, pp. 460–463, 2006.

[5] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[6] J. Bilmes, "A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," 1997. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.613

[7] M. J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *Comput. Vis. Image Underst.*, vol. 63, no. 1, pp. 75–104, 1996.

[8] A. Borzin, E. Rivlin, and M. Rudzsky, "Surveillance event interpretation using generalized stochastic petri nets," *Image Analysis for Multimedia Interactive Services, 2007. WIAMIS '07. Eighth International Workshop*, 2007.

[9] H. Greenspan, J. Goldberger, and A. Mayer, "Probabilistic space-time video modeling via piecewise gmm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.

[10] R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, and G. Coleman, "Detection and explanation of anomalous activities - representing activities as bags of event n-grams," *CVPR (1) 2005*, pp. 1031–1038, 2005.

[11] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006.

[12] F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Computer Vision and Image Understanding*, vol. 115, no. 3, 2011.

[13] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time mrf for detecting abnormal activities with incremental updates," in *Computer Vision and Pattern Recognition, 2009. IEEE Conference on*, jun. 2009, pp. 2921 –2928.

[14] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition CVPR '09.*

[15] ——, "Spatio-temporal motion pattern modeling of extremely crowded scenes," *The 1st International Workshop on Machine Learning for Vision-based Motion Analysis - MLVMA'08.*

[16] X. Ma, D. Schonfeld, and A. Khokhar, "A general two-dimensional hidden markov model and its application in image classification," *Image Processing 2007 ICIP 2007 IEEE International Conference*, vol. 6, p. 4, 2007.

[17] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition, 2010 IEEE Conference on*, jun. 2010, pp. 1975 –1981, http://www.svcl.ucsd.edu/projects/anomaly/.

[18] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Computer Vision and Pattern Recognition, 2009. IEEE Conference on*, jun. 2009, pp. 935 –942.

[19] B. T. Morris and M. M. Trivedi, "Trajectory learning for activity understanding : Unsupervised, multilevel, and long-term adaptive approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[20] O. P. Popoola and K. Wang, "Video-based abnormal human behavior recognition—a review," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. PP, no. 99, pp. 1 –14, 2012.

[21] V. Reddy, C.Sanderson, and B. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," *MLvMA Workshop, IEEE Conf. Computer Vision and Pattern Recognition (CVPR), USA*, 2011.

[22] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, "Textures of optical flow for real-time anomaly detection in crowds," *AVSS 2011*, p. 6, 2011.

[23] M. Tipping and C. Bishop, "Mixtures of probabilistic principal component analyzers." *Neural Computation*, vol. 11, no. 2, p. 443 482, 1999.

[24] A. Utasi and L. Czuni, "Detection of unusual optical flow patterns by multilevel hidden markov models," *Optical Engineering*, vol. 49, no. 1, p. 017201, 2010.

[25] T. Xiang and S. gong, "Beyond tracking : Modelling activity and understanding behaviour," *International Journal of Computer Vision*, vol. 67, no. 1, 2006.

[26] W. Xiaogang, M. Xiaoxu, and W. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 31, no. 3, pp. 539–555., 2009.

[27] A. Zaharescu and R. Wildes, "Anomalous behaviour detection using spatiotemporal oriented energies, subset inclusion histogram comparison and event-driven processing," *ECCV'10 Proceedings of the 11th European conference on Computer vision: Part I*, 2010.

[28] B. Zhao, L. Fei-Fei, and E. P.Xing, "Online detection of unusual events in videos via dynamic sparse coding," *CVPR11*, pp. 3313–3320, 2011.

[29] H. Zhong, J. Shi, and M. Visontai, "Detecting unusual activity in video," *In Proceedings of CVPR (2)*, pp. 819–826, 2004.

[30] Y. Zhou, S. Yan, and T. S. Huang, "Detecting anomaly in videos from trajectory similarity analysis," *ICME'07*, 2007.

[31] Zivkovic and Zoran, "Improved adaptive gaussian mixture model for background subtraction," in *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 2*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 28–31.