A Video-Based Abnormal Human Behavior Detection for Psychiatric Patient Monitoring

Shih-Chung Hsu ,
Dept. of Electrical Engineering,
National Tsing-Hua University,
Hsin-Chu, Taiwan
chvjohnff@gmail.com

Cheng-Hung Chuang
Dept. of Computer Sci. and Information Eng.
Asia University,
Tai-Chung, Taiwan.
chchuang@asia.edu.tw

Chung-Lin Huang
Dept. of M-commerce and Multimedia App.
Asia University,
Tai-Chung, Taiwan.
clhuang@asia.edu.tw

Por-Ren Teng and Miao-Jian Lin
Psychiatry Department,
Chang Bing Show Chwan Memorial
Hospital, Changhua, Taiwan,
porenten@ms23.hinet.net.

Abstract — This paper proposes an abnormal human behavior detection system for monitoring psychiatric patient. A normal behavior can be characterized by the spatial and temporal features of human activities. The difficulty of abnormal behavior detection is that human behavior is unpredictable and complicated. It varies in both motion and appearance. The human behavior video stream is interspersed with transition of abnormal and normal events. Here, we propose an unsupervised learning using the N-cut algorithm along with the SVM to label the video segments and then apply the Condition random field (CRF) with an adaptive threshold to distinguish the normal and abnormal events.

Keywords —Abnormal Human Behavior Detection; Support Vector Machine (SVM); Conditional Random Field (CRF).

I. INTRODUCTION

Monitoring the abnormal behavior of psychiatric patients is a labor intensive job for caregivers in the mental hospital. Some system requires the patients to wear embedded sensors (such as wrist communicators, motion detector, or passive infrared sensors). Different with the invasive sensors, video sensors is noninvasive sensor which provides sufficient and reliable information for detecting abnormal events. The video surveillance system can be a close system and the privacy will not be the major concern for the mental hospital. How to prevent the unpredictable behavior of the patients is an important issue. Recently, Kinect has also been used to recognize abnormal human behavior [7].

The abnormal event is unpredictable and undefined. The abnormal event is the event that has not occurred or occurs less frequently. If the previously detected abnormal event occurs frequently afterward, then it will become as a normal event. The abnormal event detection model has to adjust adaptively for the event occurrence. The definition of an abnormality differs with the context, and the events which are considered as abnormal typically occur very rarely compared to normal events. Due to this scarcity of the abnormal event for training, the problem of abnormal event detection is typically formulated as a novelty detection. System is trained using normal events in an unsupervised manner and the events which do not fit the learned 'normal' model are detected.

To detect abnormal events, most researches try to detect the moving object and extract the motion features such as the trajectories [1~5, 11, 19]. However, these methods have the occlusion problem. Most of abnormal event detection methods are effective only in sparsely crowded scenarios, and will degrade in densely crowded areas due to the challenges of brightness change and occlusion. Features containing shape, size and texture related information can also be used to detect the anomaly [16].

Mahadevan *et al.* [14] propose a linear dynamical system to model the dynamic textures of crowd activity as well as the spatial and temporal texture variations. Hidden Markov Model (HMM), Markov Random Field (MRF), and Conditional Random Field (CRF) are used to model human behavior. Multi-observation HMM [13] is proposed to account for the temporal causality of the activities. Coupled HMMs [22] and Semi-2D HMMs [25] are proposed to account for both the spatial and temporal causality of the activities. MRF-based methods [20, 21] cannot completely model both the spatial and temporal causality, which has high data requirements due to the use of location specific model parameters. In [24], they evaluate the performance of different state of the art features to detect the presence of the abnormal events. In [15], *CRFs* are proposed for human motion recognition which is a finite state model for all normal human behaviors.

Ryan *et al* [12] model the motion information based on the textures of optical flow. The smoothness of the optical flow field across a region is applied for the detection of anomaly in crowds. In [24], texture features extracted through Gabor wavelets are added to represent the abnormality. The histogram of optical flow has also been used in [16, 17, 21] to detect abnormal events. Kratz *et al* [22] use the distribution of spatio-temporal gradients to model the motion patterns. Mahadevan *et al*. [23] propose a detection framework using mixture of dynamic textures to represent both motion and appearance features.

Andrade *et al* [18] calculate a dense optical flow field followed by dimensionality reduction. The histogram of optical flow orientation as the descriptors or feature vectors [10] for classification using the one-class *SVM* classification method. Here, we describe the motion information based on the Motion Energy Image (MEI) which is described by Hu moment [16] for CRF model. The difficult of using HMM for the abnormal event detection is the thresholding problem. HMM model can be used to model a normal behavior. However, it is difficult to model abnormal behaviors using the HMM models because of large variations of the spatial/temporal features. Here, we propose an adaptive threshold CRF model to overcome the weakness of fixed threshold HMM method. A CRF is initially trained without using the abnormal behaviors. There exhibit similar postures and motion activities in different behaviors.

Here, we extract the motion vectors which can be converted to a label using N-cut algorithm [9] as the observation for *CRFs*.

978-1-5386-2615-3/18/$31.00 ©2018 IEEE

CRFs are initially trained based on the videos of normal behaviors. In the training process, we train the *CRFs* using the videos of normal behaviors. Our method consists of (a) extracting the motion feature of the moving human object; (b) using the BoWs to label the input motion information, and (c) training the CRFs model using these labeled features. Given an input video segment, we (a) extract the motion features of the object, (b) use BoWs to label the feature, (c) apply the CRFs to compute the likelihood, and (d) use the adaptive threshold method to identify the abnormal behavior.

II. THE CRF MODEL

CRFs are a framework based on conditional probability approaches for labeling the sequential data. Let **X** is a random variable over data sequence and **Y** is a random variable over the corresponding labeled sequence. The random variables **X** and **Y** are jointly distributed. A conditional model p(**Y**|**X**) present the paired observation and label sequences. CRF is a random field globally conditioned on observation **X**. Let G=(*V*, *E*) be a graph, such that $\mathbf{Y}=\{\mathbf{Y}_v | v\in V\}$ so that $\mathbf{Y}_v$ is indexed by the set of vertices *V*, then $(\mathbf{Y}, \mathbf{X})$ is a conditional random field in case, when conditioned on X, the random variables $\mathbf{Y}_v$ follow the Markov process. Here, we define $\mathbf{X}=(\mathbf{X}_1, \ldots \mathbf{X}_n)$ and $\mathbf{Y}=(\mathbf{Y}_1, \ldots \mathbf{Y}_n)$. If the graph *G* of **Y** is a tree, its cliques are edges and vertices. Let C(**Y**, **X**) be a set of maximal clique of *G* using the random field, The distribution over joint labels **Y** given observation **X** and parameter θ can be written as

$$p_{\boldsymbol{\theta}}(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z_{\boldsymbol{\theta}}(\mathbf{X})} \prod_{\boldsymbol{c}\in C(\boldsymbol{Y},\boldsymbol{X})} \phi_{\theta}^{c}(\mathbf{Y}_{\mathrm{c}}, \mathbf{X}_{\mathrm{c}}) \quad (1)$$

where $\phi_{\theta}^{c}(\cdot)$ is a positive value potential function of clique *c* and $Z_\theta(\mathbf{X})$ is the observation dependent normalization as

$$Z_{\boldsymbol{\theta}}(\mathbf{X}) = \sum_{X} \prod_{\boldsymbol{c}\in C(\boldsymbol{Y},\boldsymbol{X})} \phi_{\theta}^{c}(\mathbf{Y}_{\mathrm{c}}, \mathbf{X}_{\mathrm{c}}) \quad (2)$$

For a linear chain (first-order state dependency), the cliques include pairs of neighboring states (y_{t-1}, y_t) whereas the connectivity among the observation is unrestricted. For a model with T time-steps, CRF can be rewritten in terms of exponential feature functions $F_{\boldsymbol{\theta}}(y_t, y_{t-1}, \mathbf{x}, t)$ computed in terms of weighted sums over the features of cliques as

$$p_{\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x}) = \frac{1}{Z_{\boldsymbol{\theta}}(\mathbf{x})} \exp\left(\sum_{t=1}^{T} F_{\boldsymbol{\theta}}(y_t, y_{t-1}, \mathbf{x}, t)\right) \quad (3)$$

$$Z_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{\mathbf{y}} \exp\left(\sum_{t=1}^{T} F_{\boldsymbol{\theta}}(y_t, y_{t-1}, \mathbf{x}, t)\right) \quad (4)$$

The probability of a particular label sequence **y** given observation sequence **x** can be defined by a normalized product of potential functions as

$$F_{\boldsymbol{\theta}}(y_{i-1}, y_i, \mathbf{x}, i) = \sum_{v} \lambda_v\, t_v(y_{i-1}, y_i, \mathbf{x}, i) + \sum_{m} \mu_m\, s_m(y_i, \mathbf{x}, i) \quad (5)$$

which $t_v(y_{i-1}, y_i, \mathbf{x}, i)$ is the transition feature function of the entire observation sequence and the labels at position *i* and *i-1* in the label sequences. $s_m(y_i, \mathbf{x}, i)$ is state feature function of the label at position *i* and the observation. λ_v and μ_m are model parameters to be estimated. We define CRF in terms of $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \ldots, \lambda_{N_T}; \mu_1, \mu_2, \ldots, \mu_{N_S})$, where N_T and N_T are the number of transition feature functions and state feature functions.

To define state feature functions, we construct a set of real-value features $b(\boldsymbol{x}, i)$ indicates whether a feature value is observed at a particular label. The state feature function is defined as

$$s_m(y_i, \mathbf{x}, i) = \begin{cases} b(\mathbf{x}, i), & \text{if } y_i = Y_\alpha, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where Y_α is a label of CRF, y_i is a label of observation **x** at position *i* and *b*(**x**, *i*) is defined as

$$b(\mathbf{x}, i) = \begin{cases} 1 & \text{if the } i^{th}\text{feature of } \mathbf{x} = \text{ r} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where *r* is the feature value. If the label of observation sequence **x** equals to the label of the model, then the state feature function will be one, otherwise it is zero.

A transition feature function indicates whether a feature value is observed between two state or not. It is defined as

$$t_v(y_{i-1}, y_i, \mathbf{x}, i) = \begin{cases} 1, & \text{if } y_{i-1} = Y_\alpha \text{ and } y_i = Y_b, \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

where Y_α and Y_b are labels of CRF, y_{i-1} and y_i are labels of observation sequence **x** at position *i-1* and *i*.

III. THE PROPOSED METHOD

The normal/abnormal behaviors can be discriminated by a threshold model CRF. In the video of abnormal events, we may find the occurrence of larger moment of motion activities. Therefore, we use the optical flow to represent the motion information of the moving object. We compute the Motion Energy Image (MEI) in various directions which can be described by Hu moment as the feature vector.

A motion information

The moving objects is described by motion information using optical flow method [8]. Abnormal behavior detection can be treated as a behavior matching with the normal behavior which can be represented by a behavior template. For an input behavior, we may apply the behavior template matching if it does not match with the template then it is an abnormal behavior. Therefore, we represent the image by a motion vector field and then accumulate a number of motion vector fields as MEI. For every 10 frames, we accumulate ten motion vector fields as the motion energy. To represent the characteristics of behavior, we divide the motion vector field into nine different directions.

B Hu moments

With MEI obtained from the input image, we may represent the region of MEI using Hu moment [16] which is a moment invariant representation. The image moment is a weighted average of the pixels. Based on the nine directional MEI, we compute the Hu moments in nine directions. There are 7 different Hu moments, so each MEI is described by a 63 (9×7) dimension vector. The advantages of using Hu moment (ϕ_1~ϕ_7) are the scale, translation, and rotation invariant.

C. Labeling by using N-cut

To label the extracted feature of the video automatically, we apply the N-cut clustering algorithm and then use Support Vector Machine (SVM) to further classify the features. To apply the *N-cut* algorithm, we find the distance between every two features and computes the similarity matrix. we solve the eigenvalue problem to find *N* eigenvectors corresponding to the *N* smallest eigenvalues. Then, we use the k-means algorithm to cluster the eigenvectors and label the observations of each video segment. Here, we apply SVM to further classify the labeled observations. Similar to [6, 10,14], we apply the non-supervised clustering based on the similarity matrix of the feature vectors.

Bi-cut clustering algorithm is based on the graph theory. The set of feature points in feature space can be described by a undirected graph, *i.e.*, G=(V, E). G consists of a set of vertices V={1,2,…|V|} and a set of edges {(*i, j*)} with edge weight e_{ij}. The adjacency matrix is defined as **A** of which the component A_{ij} is

$$A_{ij} = \begin{cases} e_{ij} & ,if\ there\ is\ an\ edge\ \{i,j\} \\ 0 & ,otherwise \end{cases} \qquad (9)$$

where e_{ij} is the L_2 distance between linked vertices *i* and *j*. The similarity matrix can be used to cut the G into two non-overlapped clusters g_a and g_b. where $g_a=\{n_1, n_2\}$ and $g_b=\{n_3, n_4, n_5\}$ with $g_a \cup g_b$=G and $g_a \cap g_b=\Phi$. The sum of the weighted linkages between g_a and g_b to decide the cut between these two clusters 1s

$$cut(g_a, g_b) = \sum_{i\in g_a, j\in g_b} e_{ij} \qquad (10)$$

Here, we find the clustering as the best cut finding to minimize *Cut*(g_a, g_b). The *Laplacian* matrix L_G is an $n\times n$ symmetry matrix. The between-cluster of g_a and g_b makes contribute to *Cut*(g_a, g_b). We normalize the above equation and simplify the minimum cut as Rayleigh quotient problem as

$${}_{\min} cut(g_a, g_b) = \frac{n}{4}\frac{\mathbf{x}^T L_G \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \qquad (11)$$

However, minimum cut may create the isolated nodes problem. To solve isolated nodes problem, we compute all the between-cluster weighted *Cut*(g_a, g_b) and within-cluster weight *weight*(g_a) and *weight* (g_b), and modify the above equation as

$$Q(g_a, g_b) = \frac{cut(g_a, g_b)}{weight(g_a)} + \frac{cut(g_a, g_b)}{weight(g_b)} \qquad (12)$$

The above equation is also called the normalized cut. Minimizing $Q(g_a, g_b)$, we may find the best cut and avoid the isolated node problem. $Q(g_a, g_b)$ can be further simplified to satisfy $L_G\mathbf{x} = \lambda D\mathbf{x}$, of which the 2nd eigenvector is obtained to minimize Q defined as

$${}_{\min} Q(g_a, g_b) = \min_{\mathbf{x}^T D \mathbf{1}=0} \frac{\mathbf{x}^T L_G \mathbf{x}}{\mathbf{x}^T D \mathbf{x}} \qquad (13)$$

The clustering problem can be simplified by solving eigenvalue problem as $L_G\mathbf{x}=\lambda D\mathbf{x}$. By solving the second smallest eigenvalue and the corresponding eigenvector. The components of the eigenvector are used to cluster graph G into two halves. Ideally, the value of the components is discrete (*i.e.*, -1or +1). Actually, they are not discrete so that we need to the median value as the cutting value to separate the components into two groups. Besides the second eigenvector, we also consider the third and fourth eigenvectors. By using the clustering algorithm, we may separate the components into two group. Figure 1 shows the clustering results. The data is grouped into 4 groups by using the N-cut and K-means algorithm.

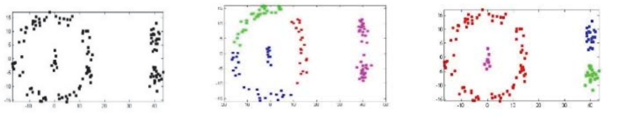

Figure 1. (a) the samples in 2-D features space. (b) the 4 k-means, (c) the 4 cuts.

D. Threshold Model with CRF

Most of the existing abnormal behavior detection methods employ a fixed threshold that best discriminates abnormal and normal behaviors. However, it is difficult to select a fixed threshold. A threshold model with CRF (*T-CRF*) is constructed by adding the label for normal behaviors patterns *N* in the original CRF using the weight of state and the transition feature function of the original CRF. Therefore the label for T-CRF are {A_1, A_2, …A_l, *N*} where *l* is the number of labels of the CRF and *G* is the label for normal patterns.

a) The weight of the state feature function. In CRF, the weights of a state feature function are distributed over several labels when it variance is small. By assigning the weight of state feature function of the label for normal behavior pattern G based on the variance of weights of state feature function

b) The weight of the transition feature function. The frequencies of normal behaviors are larger than those of abnormal behaviors. Therefore, the weights for feature functions of label for normal behavior patterns G should be higher than those of the other label for abnormal behaviors.

E. Implementation

Here, we propose an outlier detection to identify the abnormal event of which the probability is below a certain threshold under normal conditions. Labeling features by artificial way is time-consuming and unrealistic. The features clustering for the normal behaviors model with unsupervised learning is crucial for our method.

IV. EXPERIMENTAL RESULTS

We have applied our method for three testing videos captured in three different scenarios. The first video sequence consists of five actions as shown in Figure 2 including normal walking, staying and sitting, and the abnormal wandering and attack behavior. The corresponding response is shown in Figure 3. The response of wandering is close to the normal walking because there are such similar actions between walking and wandering. In the other abnormal event, the attack action generates small response obviously because the attack action is unknown in the pre-trained model.

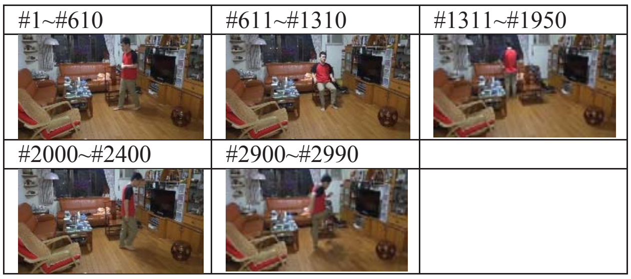

Figure 2. The sequence of normal/abnormal behaviors.

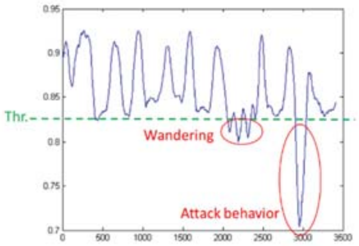

Figure 3. The response of home behavior sequence.

The second video is obtained from the local mental hospital (Show-Chwan Hospital). The video sequence are obtained with the IRB verification. First, we need find the human object in the videos. We compare the object detection method with ours. The results of conventional object detection are shown in Figure 4(a), whereas the results of using our method are shown in Figure 4(b). Then we apply our method to find the abnormal events in the videos of the guest room in the rehabilitation center of Show-Chwan mental hospital. There are two fighting events between two patients in the scene which are detect accurately as shown in Figures 5.

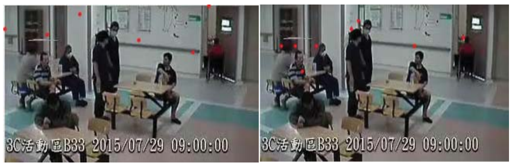
3C活動區B33 2015/07/29 09:00:00
3C活動區B33 2015/07/29 09:00:00

Figure 4. Object detection.

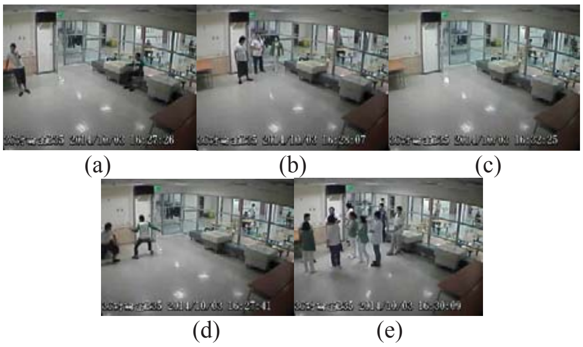

Figure 5. (a)~(c) are the normal events, (d) and (e) are the abnormal events.

The third video is taken from the corridor of the rehabilitation center. There is two queues of patients waiting for the caregivers to give them the medicine. The abnormal events of the patient stealing the drugs are detect accurately as shown in Figures 6.

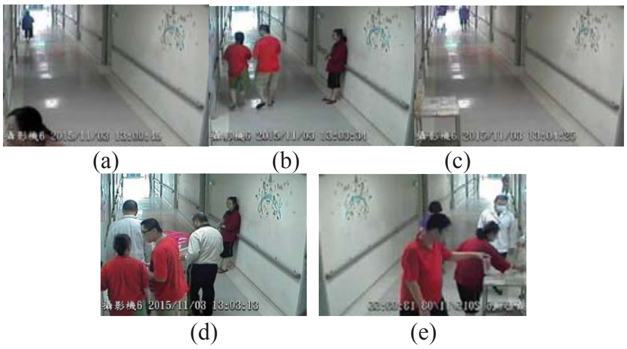

Figure 6. (a)~(c) are the normal events, (d) and (e) are the abnormal events.

REFERENCES

1. C. Beleznai and H. Bischof, "Fast Human Detection in Crowded Scenes by Contour Integration and Local Shape Estimation," *IEEE CVPR*, 2009, pp.2246-2253.
2. T. Wang, J. Chen, Yi Zhou, and H. Snoussi," Online Least Squares One-Class Support Vector Machines-Based Abnormal Visual Event Detection," *Sensors*, *13*(12), 2013.
3. S. C. Lee and R. Nevatia "Hierarchical Abnormal Event Detection by Real Time and Semi-Real Time Multi-Tasking Video Surveillance System," Machine Vision and Applications, Vol. 25(1), pp 133-143, January 2014.
4. O. P. Popoola and K. Wang, "Video-Based Abnormal Human Behavior Recognition: A Review," *IEEE Trans. on SMC*, Part C, 2011.
5. Y. Qian and W. Zhang, "Human Abnormal Behavioral Detection for Video Surveillance", *3rd Int. Conf. on Materials Engineering, Manufacturing Tech. and Control* (ICMEMTC) 2016.
6. H.-D. Yang and S.-W. Lee, "Sign Language Spotting with a Threshold Model based on Conditional Random Fields," IEEE Trans. on PAMI, Vol. 31, No. 7, July 2009
7. F. Ghanbarnezhad1, S. Hosseini and M. Hosseini, "Recognition of Abnormal Human Behavior using Kinect Case Study: Tehran Metro Station," *Indian Journal of Science and Technology*, Vol 9(43), 2016.
8. Berthold K. P. Horn and Brian G. Schunck,"Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185--203, 1981.
9. I. S. Dhillon. "Co-clustering documents and words using bipartite spectral graph partitioning", *ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Francisco, August 2001.
10. T. Wang and H. Snoussi, "Detection of Abnormal Visual Events via Global Optical Flow Orientation Histogram," IEEE Trans. on Inf. Forensics and Security, Vol. 9, No. 6, June 2014.
11. M. J. Roshtkhari and M. D. Levine, "Online Dominant and Anomalous Behavior Detection in Videos," IEEE CVPR 2013.
12. D. Ryan, S. Denman, C. Fookes, and S. Sridharan, "Textures of optical flow for real-time anomaly detection in crowds," AVSS, pages 230–235, 2011.
13. T. Xiang and S. Gong. "Incremental and Adaptive Abnormal Behavior Detection," CVIU, vol. 111, pages 59–73, 2008.
14. V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos. "Anomaly Detection in Crowded Scenes," *IEEE CVPR*, June 2010.
15. C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas, "Conditional Models for Contextual Human Motion Recognition," *IEEE ICCV*, Oct. 2005.
16. V. Reddy, C. Sanderson, and B. C. Lovell, "Improved Anomaly Detection in Crowded Scenes via Cell-Based Analysis of Foreground Speed, Size and Texture," CVPRW, 2011.
17. A.Adam, E.Rivlin, I.Shimshoni, and D.Reinitz, "Robust real time unusual event detection using multiple fixed-location monitors," IEEE Trans. PAMI, vol. 30, pages 555–560, Mar 2008.
18. E. Andrade, S. Blunsden, and R. Fisher, "Modelling Crowd Scenes for Event Detection," IEEE ICPR, 2006.
19. A. Basharat, A. Gritai, and M. Shah, "Learning Object Motion Patterns For Anomaly Detection and Improved Object Detection," *CVPR*, IEEE, Jan., 2008.
20. H. Nallaivarothayan, C. Fookes, S. Denman, S. Sridharan, "An MRF based Abnormal Event Detection Approach using Motion and Appearance Features," IEEE AVSS, 2014.
21. J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates," IEEE CVPR, pages 2921–2928, 2009.
22. L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," IEEE CVPR, pages 1446–1453, 2009.
23. V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," *IEEE CVPR*, 2010.
24. H. Nallaivarothayan, D.Ryan, S.Denman, S.Sridharan, and C. Fookes, "An evaluation of different features and learning models for anomalous event detection," DICTA 2013.
25. H. Nallaivarothayan, D. Ryan, S. Denman, S. Sridharan, and C. Fookes,"Anomalous Event Detection using a Semi-Two Dimensional Hidden Markov Model," DICTA, pages 1 – 7, 2012.