Abnormal Event Detection via Adaptive Cascade Dictionary Learning

Hui Wen[1,2], Shiming Ge[1], Shuixian Chen[1], Hongtao Wang[1,2], Limin Sun[1,*]
[1]Beijing Key Laboratory of IOT Information Security Technology, Institute of Information Engineering, CAS, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China

I. Abstract

Detecting abnormal events plays an essential role in video content analysis and has received increasing attention in surveillance system. One of the major problems in abnormal event detection is the imbalanced classification issue due to the rare abnormal samples. Another problem is the difficulty of detecting anomalies within a reasonable amount of computation time. To address these problems, we propose an adaptive cascade dictionary learning framework for detecting the anomalies. The framework considers anomaly detection as an one-class classification problem with a cascade of dictionaries. Each stage of the cascade constructs an adaptive dictionary to detect the anomalies with costless least square optimization solution. The experiments on benchmark datasets demonstrate that the proposed method has a better performance while comparing with several state-of-the-art methods.

II. Introduction

Abnormal event detection has been received increasing attention in recent years [1]. Although approaches to unusual event detection have been studied for years, there still exist two challenging problems. One of the major problems in anomaly detection is rare or none abnormal samples, while it is hard to build a classification model for detecting the unusual objects or motions. Another problem is the long response time in anomaly detection, which makes the proposed methods unpractical in surveillance system.

While abnormal event detection has difficulty to capture all of the abnormal classes by the lack of enough abnormal samples, most of the researches consider the detection as an one-class classification or outlier detection problem that can be solved by detecting the events deviated from the normal pattern. Specifically, the deviation approaches model usual activity, then detecting unusual events as those that different from the pre-trained model [2]–[5]. Following this line, we easily find that the normal pattern can be represented by motion-based feature, which contains rich individual behavior information. Typical motion-based video event detection method estimates object motion from optical flows [6] [7] or spatio-temporal volumes [8] [9]. Optical flow-based approaches [10] extract trajectories to represent normal patterns by tracking object of interest. The algorithms consider trajectories with lower probability as anomaly by fitting a probability model over the training data. For instance, Zhang et al. [11] proposed a semi-supervised method that detecting motion deviations with HMM and Mahadevan et al. [12] used multivariate gaussian mixture to generate probability model of normal pattern . Spatio-temporal feature based approaches [13] [14] use spatio-temporal video volumes to construct a behavioral model based on scene analysis in the form of BOV (bag of visual words) [15] [16]. Compared with optical flow feature, the spatio-temporal feature that capture activity within local scene regions can overcome the variation of activities caused by the large number of individuals [5].

To reduce the computation cost of the anomaly detection, representation in low dimension subspace should be considered [17]. Recently, sparsity-based anomaly detection models [3] [18] [19] have achieved considerable performance with short detection time. Lu et al. [20] proposed a fast abnormal event detection model that achieves the millisecond level in anomaly detection. The algorithm trained a set of dictionaries to represent normal pattern using sparse combination learning method. However, the proposed method needs to ensure each dictionary with a fixed dimension size, which affect the performance of the anomaly detection with unknown optimal dimension. We propose an adaptive cascade dictionary learning method with low dimension constraint, which can learn the dimension size literately.

In this paper, we propose an effective cascade dictionary framework to detect the abnormal events. The proposed method uses gradient based spatio-temporal feature to represent the behaviors of individuals and provides a cascade dictionary method that learning different levels of representation. Furthermore, the proposed approach achieves high performance on public datasets while comparing with several state-of-the-art algorithms.

III. Approach

In this section, we present the algorithmic details and theoretical analysis. Figure.(1) illustrates the framework of our algorithm. The main steps of adaptive cascade dictionary learning framework are summarized as follows: First, we extract multi-scale 3D gradient feature [12] from given videos to form a set of data cubes Y. Then, we train a group of dictionaries with different dimension size and choose an optimal dictionary with minimal error cost. Next, we take out

*Limin Sun is the corresponding author.

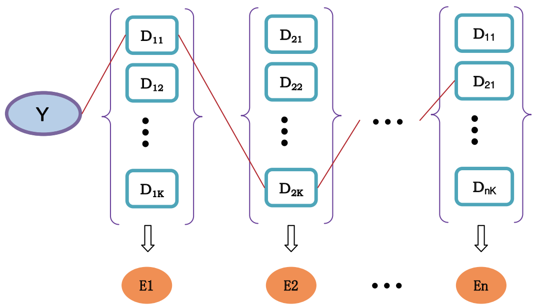

Fig. 1: The architecture of our approach. Y denotes training data cubes. Each stage of the cascade has k candidate dictionaries. The proposed method selects only one dictionary with minimal reconstruction error E in each stage.

the samples that have a high reconstruction error from the current optimal dictionary and train another optimal dictionary to represent them. Finally, we repeat the previous step until all elements in Y are represented by trained dictionaries.

A. Learning Adaptive Cascade Dictionary

The 3D gradient features $y_j \in R^{w\times h\times t}$ are constructed by assuming a cube of size $w \times h \times t$ within the local cuboid region for representing the non-uniform local spatio-temporal motion patterns. $w \times h$ is the size of the spatial window and t is the depth of the video cube in time. For each cube location, the features in all frames are denoted as $Y = \{y_1, ..., y_m\} \in R^{m\times l}$, where m is the number of the samples and l is the length of feature vector y_i. Based on the hypothesis that the normal pattern can be represented by dictionaries, our approach aims to find the corresponding cascade of dictionaries $\{D_1, ..., D_n\}$ for classifying the abnormalities. The dictionaries in cascade are selected from the dictionary pool $\boldsymbol{D} = \{D_1^1, D_1^2, ..., D_1^k, D_2^1, .., D_n^k\}$ with each $D_i^p \in R^{l\times s_k}$ containing s_k basis vectors. The parameter s_k controls the dictionary sparsity and be set by method of dichotomic classification, such as $\{s_1 = \frac{1}{2}l, s_2 = \frac{1}{4}l, s_3 = \frac{1}{8}l, s_4 = \frac{3}{8}l...\}$.

Our adaptive cascade dictionary learning method treats the abnormal event detection as a cascade of classification problem. Compared with single dictionary method, the cascade method can achieve decreased false detection performance [21]. However, while learning different levels of representation, we observe that larger dimension size of the dictionary would make smaller reconstruction error. Thus, we add a regularization term to balance multi-subspace learning:

$$E(y) = \min_{\alpha,\beta,D} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{p=1}^{k} \alpha_j^i \beta_i^p (||y_j - D_i^p x_j^{ip}||_2^2 + \lambda||D_i^p||_0)$$
$$s.t. \sum_{i=1}^{n} \alpha_j^i = 1, \alpha_j^i = \{0,1\}$$
$$\sum_{p=1}^{k} \beta_i^p = 1, \beta_i^p = \{0,1\} \qquad (1)$$

where $\alpha = \{\alpha_1, ..., \alpha_m\}$, $\alpha_j = \{\alpha_j^1, ..., \alpha_j^n\}$ and $\beta = \{\beta_1, ..., \beta_n\}$, $\beta_i = \{\beta_i^1, ..., \beta_i^k\}$. Each α_j^i indicates whether or not the i^{th} dictionary D_i is chosen for representing data j and each β_i^k shows that the k^{th} dictionary D_i^k in the dictionary pool D_i is the optimal error minimization choice in stage i. x_j^{ip} is the corresponding sparse feature for representing y_j with dictionary D_i^p. The constraints $\sum \alpha_j^i = 1, \alpha_j^i = \{0,1\}$ requires that the data y_j only can be represented by one dictionary and the constraints $\sum \beta_i^k = 1, \beta_i^k = \{0,1\}$ requires that stage i must choose only one proper dictionary in the dictionary pool with different dimension s_k. λ is a weight parameter, and the l_0-norm $||D||_0$ is a regularization term for controlling the dimension of representation in lowest value as well as the reconstruction error.

Consider that the l_0-minimisation is regarded as an NP-hard problem, l_0-minimisation problem always be relaxed to another norm problem such as l_1-minimisation and l_2-minimisation. Compared with l_1-norm or l_2-norm, the low rank constraint shows a better representation of the sparsity of matrix. In this case, we use Frobenius-norm to constraint the dictionary sparsity level, which can balance the weight of the reconstructed error in different subspace. Thus, we update the function (1) as:

$$E(y) = \min_{\alpha,\beta,D} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{p=1}^{k} \alpha_j^i \beta_i^p (||y_j - D_i^p x_j^{ip}||_2^2 + \lambda||D_i^p||_F^2)$$
$$s.t. \sum_{i=1}^{n} \alpha_j^i = 1, \alpha_j^i = \{0,1\}$$
$$\sum_{p=1}^{k} \beta_i^p = 1, \beta_i^p = \{0,1\} \qquad (2)$$

where $||D_i^p||_F^2 = Tr(D_i^{p\,T} D_i^p)$ represent the sparsity of the dictionary D by using the rank of the matrix. An advantage of using Frobenius-norm is that the dictionary can be trained with least square optimization solution.

The overall form of the adaptive cascade dictionary architecture seems like a degenerate decision tree. The method considered positive results as anomalous pattern and negative results as normal pattern. A positive result from the first dictionary reconstruction process triggers the second dictionary while the reconstruction errors less than threshold. Then, the second dictionary triggers the third dictionary, and so on. A negative outcome at each stage leads to immediate rejection of the data cube, which is considered as a normal motion pattern.

B. Optimization for Adaptive Cascade Dictionary Learning

Based on cascade learning method, the proposed algorithm automatically ends in stage n while all data are represented by trained dictionaries. Each dictionary in dictionary learning is obtained by setting a reconstruction error upper bound ζ, which is uniformly for all elements in D. The optimization of dictionary learning is considered as convergence while the reconstruction error is smaller than ζ. We formulate the process as:

$$E(y_j) = \underset{\alpha,\beta,D}{min} \sum_{i=1}^{n} \sum_{p=1}^{k} \alpha_j^i \beta_i^p \{||y_j - D_i^p x_j^{ip}||_2^2 - \zeta\} \leq 0$$
$$s.t. \sum_{i=1}^{n} \alpha_j^i = 1, \alpha_j^i = \{0,1\} \quad (3)$$
$$\sum_{p=1}^{k} \beta_i^p = 1, \beta_i^p = \{0,1\}$$

For each data cube y_j, α_j^i and β_i^p indicate the data cube y_j represented by p-th dictionary in stage i. The parameters are trained in an iterative manner with the cascade architecture. In each stage of the cascade, dictionary is learned with sparsity constraint by Eq.(2). and updated for representing training data as many as possible. After that, remaining training data cubes that cannot be well represented by the dictionary in this stage are sent to the next stage, which will be used for training another dictionary. This process ends while all training data are computed and the corresponding dictionaries converge to the optimal solution with reconstruction error less than ζ.

Specifically, in the i^{th} stage, the training process contains two steps. The first step trains a set of dictionaries as a dictionary pool $D_i = \{D_i^1, D_i^2 ... D_i^k\}$ with each $D_i^k \in R^{l \times s^k}$ containing different dictionary basis vectors s^k. Each dictionary is trained by setting the reconstruction error upper bound ζ. The second step aims to select one dictionary in the dictionary pool with minimal reconstruction error. The objective function can be expressed as:

$$E_i(y) = \underset{\alpha,\beta,D}{min} \sum_{j=1}^{m} \sum_{p=1}^{k} \alpha_j^i \beta_i^p (||y_j - D_i^p x_j^{ip}||_2^2 + \lambda ||D_i^p||_F^2 - \zeta)$$
$$s.t. \sum_{i=1}^{n} \alpha_j^i = 1, \alpha_j^i = \{0,1\}$$
$$\sum_{p=1}^{k} \beta_i^p = 1, \beta_i^p = \{0,1\} \quad (4)$$

where $E_i(y)$ shows the minimal error result obtained by using an optimal dictionary selection strategy and the process can be easily solved step by step. Given the leftover training data from stage $i-1$, we set $\alpha_j^i = 1$ to pick up the unrepresented data in stage i and train the dictionaries $D_i = \{D_i^1, D_i^2 ... D_i^k\}$ with the condition constraint $\{\forall D_i^p \in D_i, ||y_j - D_i^p x_j^{ip}||_2^2 - \zeta \leq 0\}$. Then, we choose a dictionary D_i^p from the dictionary pool with minimal error strategy as Eq.(2) proposed by setting $\beta_i^p = 1$.

Different from general multi-dictionaries training, we solve this optimization problem with three step iterations. In each stage i, the proposed method updating $\{D_i^p\}$, $\{\alpha_j^i\}$ and $\{\beta_i^p\}$ as follows:

Update$\{D_i^p\}$ With fixed α, β, Eq.(2) becomes a quadratic optimization function with a dictionary regularization term:

$$\underset{D}{min} \sum_{j=1}^{m} \sum_{p=1}^{k} \alpha_j^i \beta_i^p (||y_j - D_i^p x_j^{ip}||_2^2 + \lambda ||D_i^p||_F^2) \quad (5)$$

where x_j^{ip} can be represented as $(D_i^{p\,T} D_i^p)^{-1} D_i^{p\,T} y_j$ with least square optimization method and the objective function can be solved by optimizing the variable D_i^p with block-coordinate descent method. For simple expression, we use matrix Y_i to represent all of the training data in stage i and matrix X_i^p to represent the corresponding sparse representation. The problem can be expressed as:

$$\underset{D}{min} \sum_{p=1}^{k} \beta_i^p (||Y_i - D_i^p X_i^p||^2 + \lambda ||D_i^p||_F^2) \quad (6)$$

For each dictionary D_i^p, the optimization solution of the dictionary learning is updating the dictionary iteratively by block-coordinate descent method:

$$D_i^p = Proj_c(D_i^p - \tau((D_i^p X_i^p - Y_i) X_i^{p\,T} + \lambda D_i^p)) \quad (7)$$

where $Proj_c$ is an orthogonal projecting function that normalize the dictionary to be unit-norm. τ is a step size of the gradient descent by setting to $1/||X_i X_i^T||$. Anyway, the energy of the objective function decreases fastest in the direction of the negative gradient of Eq.(7) at D_i^p in each iteration.

Update$\{\alpha_j^i\}$ In stage i, α_j^i controls training data y_j whether can be represented by a dictionary D_i^p. The update step of the parameter α_j^i is expressed as:

$$\alpha_j^i = \begin{cases} 1 & \text{if } ||y_j - D_i^p x_j^{ip}||_2^2 < \zeta \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Update$\{\beta_i^p\}$ With a set of learned dictionaries $D_i = \{D_i^1, D_i^2 ... D_i^k\}$ in stage i, β_i^p decide which dictionary D_i^p in dictionary pool can represent the training data. For simple expression, we consider that $f(D_i^p) = ||Y_i - D_i^p X_i^p||^2 + \lambda ||D_i^p||_F^2$. The optimization solution of the parameter β_i^p can be expressed as:

$$\beta_i^p = \begin{cases} 1 & \text{if } f(D_i^p) \leq f(D_i^{\hat{p}}) \quad \forall \hat{p} = 1, ..., k \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Generally, we learn several dictionaries with different dimension size in each stage and only take one dictionary with the minimal reconstruction error to represent the current data. We repeat the process of dictionary learning to obtain a cascade of dictionaries until the training data set is empty. This scheme uses cascade structure to reduce unnecessary classification and makes the learned dictionaries fall into a mixture of low dimension subspaces for better representation.

C. Abnormal Event Detection

The abnormal event detection can be calculated quickly by checking the least square optimization method. The optimization results of $\{\alpha_j^i\}$, $\{\beta_i^p\}$ decide which $\{D_i^p\}$ can be used in cascade. Thus, we can replace the sparse feature with the closed-form solution $x_j = (D_i^{p\,T} D_i^p)^{-1} D_i^{p\,T} y_j$ in the quadratic function. The reconstruction error in D_i^p can be expressed as:

$$||y_j - D_i^p x_j^p||^2 = ||(I - D_i^p (D_i^{p\,T} D_i^p)^{-1} D_i^{p\,T}) y_j||^2 \quad (10)$$

TABLE I: Comparison of frame-level detection precision and number of the learned dictionary under the different sparse level of the dictionaries.

Dimension	Dim-10	Dim-15	Dim-20	Dim-25	Dim-30	Dim-40	Dim-50	Ours
Num	317	201	150	40	7	3	4	8
AUC	61.1%	66.5%	68.0%	68.0%	68.1%	67.8%	66.6%	68.3%

TABLE II: Comparison of frame-level detection precision under the different overlap threshold of the detection.

Overlap	0.2	0.3	0.4	0.5	0.6	0.7	0.8
MDL	70%	67.3%	63.3%	59.3%	57.5%	55.7%	54.4%
Ours	70%	68.3%	66.2%	64.2%	63.2%	62.8%	62.4%

where I is a identity matrix. As the Eq.(10) shows, x_j is considered as a normal pattern while the reconstruction error for D_i^p is less than threshold τ. In other words, data cube x_j is considered as an abnormal pattern while the reconstruction error of cube x_j is greater than threshold by all of the stages.

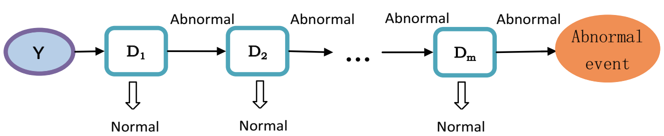

Fig. 2: Abnormal event detection with a cascade of dictionaries.

IV. Experiments

We evaluate our approach on two different datasets: Avenue dataset [20] and UCSD pedestrian dataset [12]. In the experiments, the proposed method compute three-scale 3D gradient features with resolution $10 \times 10 \times 5$ and set reconstruction error upper bound $\zeta = 0.1$ experimentally.

A. Adaptive Cascade Performance on Avenue Dataset

To validate the performance of adaptive cascade structure, we conduct quantitative comparison with general multi-dictionaries learning method (MDL) on Avenue dataset. The dataset contains 15 training sequences and totally 35,240 frames with resolution of 640×320. Moreover, Avenue have 14 unusual event sequences that including running, throwing objects and jumping.

Basically, the main difference between MDL and adaptive cascade dictionaries is whether or not the dictionary dimension size is fixed. Thus, we set different dimension of MDL compared with the proposed adaptive learning method. In the experiments, we consider that frame is detected if the overlap of detected area and ground truth is greater than 0.3. We also record the number of the learned dictionaries for describing the complexity of the two models. Table I shows that our method can train a small set of dictionaries for representing training data and achieve a better performance on the dataset. Experimental results demostrate that an improper dimension size of the dictionary will result in learning a large number of dictionaries and worse detection performance.

To evaluate the robustness of the detection method, we set different overlap threshold to calculate the detection precision of the methods. Table II shows that our method have a better detection result while compared with the MDL method. The decreasing degree of the detection precision is lower than general multi-dictionaries learning method, which demonstrate that our approach have a more robust performance.

B. Anomaly Detection Performance on UCSD Ped1 Dataset

To validate the performance of anomaly detection, we evaluate the proposed method on UCSD Ped 1 dataset with several algorithms. The dataset contains 34 short clips for training and 36 clips for testing. The video sequences shows abnormal events occur with different crowd densities. In this dataset, the anomalous patterns are the presence of non-pedestrians on the walkway. In order to make a quantitative comparison, the performance is evaluated on frame-level and pixel-level with the equal error rate(EER), equal detected rate (EDR) and area under roc curve (AUC) as suggested by literature [12].

The results in Table III,IV indicate that the proposed algorithm achieves better performance on UCSD pedestrian dataset at both frame level detection and pixel level localization while comparing with several state-of-art algorithms. Furthermore, our approach also achieves almost 100 fps in experiments, while others [12], [18], [22] achieve 30 fps at most.

TABLE III: Comparison of pixel-level EDR and AUC on the UCSD Ped1 dataset.

	MPPCA [12]	SF [12]	MDT [12]	Sparse [18]	Subspace [22]	Ours
EDR	18%	21%	45%	46%	39.3%	61%
AUC	21.3%	19.7%	44.1%	13.3%	43.2%	65.3%

TABLE IV: Comparison of frame-level EDR and AUC on the UCSD Ped1 dataset.

	MPPCA [12]	SF [12]	MDT [12]	Sparse [18]	Subspace [22]	Ours
EER	40%	31%	25%	19%	29.6%	17%
AUC	59%	67.5%	81.8%	86%	68.4%	88%

V. Conclusion

This paper presents a novel approach for abnormal event detection via adaptive cascade dictionary learning method. The algorithm is centered on three main ideas: (1) Designing a cascade dictionary learning method for representing the normal pattern; (2) Providing an automatic dictionary dimension selection strategy with low-rank dictionary constraint; (3) Accelerating abnormal event detection with cascade structure.

Acknowledgements

This work was supported in part by the Strategic Priority Research Program of CAS (No. XDA06040101), the National Natural Science Foundation of China (No. 61472418) and the Foresight Development Program of CAS (No. Y4Z0032102).

References

[1] K. W. Popoola, O.P., "Video-based abnormal human behavior recognitiona review," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 6, pp. 865 – 878, 2012.

[2] 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, 2008.

[3] B. Zhao, L. Fei-Fei, and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3313–3320.

[4] P. Cui, L.-F. Sun, Z.-Q. Liu, and S.-Q. Yang, "A sequential monte carlo approach to anomaly detection in tracking visual events," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[5] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1446–1453.

[6] E. L. Andrade, S. Blunsden, and R. B. Fisher, "Modelling crowd scenes for event detection," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 1. IEEE, 2006, pp. 175–178.

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

[8] 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, pp. 323–333, 2011.

[9] X. Cui, Q. Liu, M. Gao, and D. N. Metaxas, "Abnormal detection using interaction energy potentials," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3161–3167.

[10] 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. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2921–2928.

[11] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, "Semi-supervised adapted hmms for unusual event detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 611–618.

[12] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1975–1981.

[13] M. J. Roshtkhari and M. D. Levine, "Online dominant and anomalous behavior detection in videos," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 2611–2618.

[14] M. Bertini, A. Del Bimbo, and L. Seidenari, "Multi-scale and real-time non-parametric approach for anomaly detection and localization," *Computer Vision and Image Understanding*, vol. 116, no. 3, pp. 320–329, 2012.

[15] Y. Benezeth, P.-M. Jodoin, and V. Saligrama, "Abnormality detection using low-level co-occurring events," *Pattern Recognition Letters*, vol. 32, no. 3, pp. 423–431, 2011.

[16] T. Hospedales, S. Gong, and T. Xiang, "Video behaviour mining using a dynamic topic model," *International journal of computer vision*, vol. 98, no. 3, pp. 303–323, 2012.

[17] C. C. L. C. S. C. Mei Kuan Lim, Ven Jyn Kok, "Crowd saliency detection via global similarity structure," *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pp. 3957 – 3962, 2014.

[18] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3449–3456.

[19] J. Xu, S. Denman, S. Sridharan, C. Fookes, and R. Rana, "Dynamic texture reconstruction from sparse codes for unusual event detection in crowded scenes," in *Proceedings of the 2011 joint ACM workshop on Modeling and representing events*. ACM, 2011, pp. 25–30.

[20] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in matlab," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2720–2727.

[21] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[22] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2790–2797.