Real-time abnormal event detection in crowded scenes

Article *in* Journal of Theoretical and Applied Information Technology · September 2018

Ahmed Nady
Helwan University

Ayman Atia
Helwan University

Journal of Theoretical and Applied Information Technology
30th September 2018. Vol.96. No 18
© 2005 – ongoing JATIT & LLS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REAL-TIME ABNORMAL EVENT DETECTION IN CROWDED SCENES

[1]AHMED NADY, [1,2]AYMAN ATIA, and [1]AMAL ELSAYED ABUTABL

[1]Department of Computer Science, Faculty of Computers and Information, Helwan University, Cairo, Egypt.
[2] Misr International University. Cairo, Egypt

E-mail: {ahmednady, ayman, amal.aboutabl}@fci.helwan.edu.eg

ABSTRACT

Detecting unusual events in crowded scenes has drawn considerable research interest lately. In this paper, an unsupervised method that relies on a spatio-temporal descriptor and a clustering technique is presented to tackle this problem. We employ space-time auto-correlation of gradients (STACOG) descriptor to extract spatio-temporal motion features from video sequence. Following that, the K-medoids clustering algorithm is used to partition the STACOG descriptors of training frames into a set of clusters. The frame abnormality is defined by distances between the center of the clusters and the motion feature extracted by STACOG. We have conducted experiments on various benchmark datasets and the results show that the proposed method obtains comparable results: 98.48% AUC for UMN, and 92.13% accuracy for PETS 2009, at the frame level. In addition, fast computation time of our method that satisfies the demand of real-time processing.

Keywords: *STACOG, K-medoids, 3D gradient, Abnormal event detection, Visual surveillance*

1. INTRODUCTION

Surveillance cameras have become ubiquitous by reason of growing security matters and low costs of equipment. They are deployed at numerous public places, like airports, train stations, city centers, and shopping malls. Based on what the surveillance cameras capture, visual surveillance systems attempt to understand and describe what happen in the scene [1]. And it has an extensive variety of potential applications, such as important building security, Elderly care and traffic road monitoring [2].

Conventional visual surveillance which relies heavily on the manpower to analyze videos prove ineffective as the excessive number of cameras screen to monitor, fatigue due to lengthy monitoring and lack of a beforehand knowledge for what to look for. Together with the tremendous amount of video data which is generated per diem. This has prompted the need for an automated visual surveillance system to determine and recognize unusual events in real time.

So, detecting the unusual events which are rare and have a low probability of occurring is the key

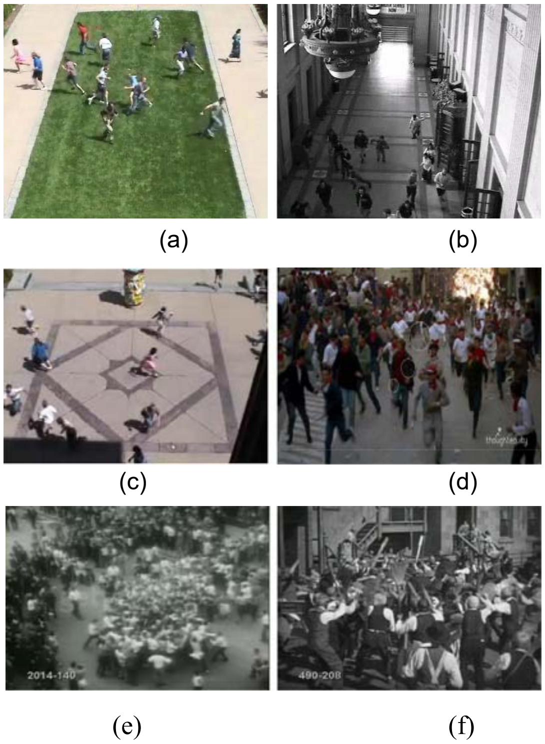


Figure. 1: Some examples of abnormal crowd event where people start to run suddenly (a-d), clashing (e), or fighting (f).

purpose of an automated visual surveillance system. Thus, Abnormality detection can be defined as a technique of finding events which are unusual in regard to previously observed events in the video. One common approach to anomaly detection is to train a model by normal image sequences to learn patterns of a normal event. Then, the model marks the event as abnormal if it deviated from learned patterns [3]. In this research, we focus on the real-time detection of anomalous behavior in crowded scenes~~.~~ Figure. 1 illustrates some examples of abnormal events in crowded scenes. Abnormal crowd event can be classified into two groups on the basis of a scale of interest: global anomalous event and local anomalous event. The global anomalous event (GAE) concentrates on deciding if the whole scene is abnormal or not, whereas the local anomalous event (LAE) determines the regions where the abnormal event is happening within the scene [4]. A considerable range of methods has been designed to identify either local or global abnormal events. Like, Mehrsan et al. [2] that propose an unsupervised method for learning the normal patterns and finding local anomalous ones occurring within a scene via a probabilistic scheme. On the other hand, Mehran et al. [5] introduce a social force model which depicts the crowd behavior as the consequence of objects interaction to detect global anomalous crowd behaviors. The limitation of the previous works that detect either local or global anomalous crowd behavior is its high computational cost which limits its applicability for real-time processing (i.e. The execution time per frame with resolution 158 x 234 is 25s [6], 3.8s [4], 0.19s [2] and 0.96s [21] per frame with 240 x 320 resolution). In addition, most of the existing methods for global abnormal detection rely on analysis of optical flow. Optical flow is the most common approach used to describe movement in a scene. However, optical flow is sensitive to noise and unstable [23]. On the other hand, our method adopts space-time auto-correlation of gradients (STACOG) descriptor which based on 3D gradients. In this study, we introduce a novel method that operates in a real-time to detect global abnormal events in crowded scenes. Firstly, STACOG feature descriptors are calculated to describe motion properties in the spatio-temporal domain of the input video sequence. Secondly, the K-medoids clustering algorithm is used to partition the STACOG descriptors of training frames into a set of clusters, and are then used to determine whether a frame is normal or not using a distance metric.

Mainly, the paper's contributions are threefold: First, we introduce STACOG feature descriptor to represent video event for detection of abnormality in crowded scenes. Second, we evaluate our method on publicly available surveillance datasets: UMN dataset [7] and PETS2009 dataset [8]. Experimental results show that the proposed method is comparable to competing methods in terms of accuracy. Third, our method is computationally inexpensive which make it applicable for real-time detection of anomalies in surveillance videos.

The structure of the paper is arranged as follows. Section 2 conducts a literature review concerning anomaly crowd detection. In Section 3 the proposed algorithm is discussed. Experiments and comparison are shown in Section 4. Lastly, the paper is summed up in Section 5.

2. RELATED WORK

Different approaches have been proposed to detect abnormal events in a scene. Those approaches can be fallen into two classes: (i) tracking approaches, which depend on the detection and tracking of interest objects from image sequences; (ii) non-tracking approaches, which depend essentially on the extraction of low-level features, like motion (optical flow or gradient), or texture [2], [9].

For Tracking-based approaches [10], [11], objects of interest are tracked and spatial location of their motion is recorded to produce a trajectory. Any trajectory that deviates significantly from learnt trajectories is treated as anomalies. Tracking-based approaches are suitable for the uncrowded scene, i.e., scene with few objects, but are impractical to deal with a crowded scene, where precise tracking of a target is unattainable due to occlusions. Besides, they take into account abnormalities that result from spatial deviations only , thus we cannot detect the abnormality in object appearance or its motion that pursues a normal trajectory [12].

Non-Tracking approaches are more appropriate for crowded scenes by analyzing the spatio-temporal information rather than separating foreground objects [2], [6], [12], [13]. Depending on the application, the anomalous events are categorized into two class: global anomalous event and local anomalous event.

Concerning the LAE, mixtures of dynamic textures are utilized in [6] to model appearance and dynamics of the normal crowd scene simultaneously and this model assigns outliers to anomaly class. Kratz et al. [14] model the distribution of extracted spatio-temporal gradient with 3D Gaussian distributions, and then HMM is used to detect abnormal events in densely crowded scene. Boiman and Irani [15] propose an algorithm that uses spatio-temporal bricks taken from former samples for composing the new monitored visual data. Regions in the monitored data is considered normal if they can be formed by means of great adjacent bricks from stored frames. In contrast, regions in the monitored data are deemed as suspicious, since they cannot be formed from the database. A limitation of their inference by composition algorithm is that the execution time and space growths linearly with the number of stored samples. Consequently, Mehrsan et al. [2] propose an online unsupervised method for learning the normal patterns and finding local anomalous ones occurring within a scene. By utilizing a probabilistic scheme which models the space-time arrangements of video volumes, anomalies are defined as those video volumes arrangement possessing the very low likelihood of occurrence. Ryan et al. [23] present a novel video event representa- tion called textures of optical flow to detect anomalous behavior in crowded scenes. Based on the observation that the anomalous objects usually produce abnormal flow patterns over their whole surface, they extend greyscale textural features to dense optical flow fields that adopted from [24] to measure the uniformity of a flow field. The Gaussian mixture model (GMM) is utilized to learn motion patterns of the normal class from training image sequences based on spatio-temporal patches, and then the patches of low probability under this model are classified as an anomaly. In [16] HOFM is used to characterize the patterns of moving objects from cells that are not overlapping on the scene. HOFM captures information concerning the orientation of the moving objects and their velocities. During testing, the input patterns of each cell are compared with stored ones at this cell location using nearest neighbor search. In [25] a novel visual representation called motion influence map is proposed that characterizes the motion of moving objects by jointly taking into consideration their movement speeds, directions, sizes, and their interactions. In order to localize the regions of anomalous motion patterns, the K-means clustering is performed for each region of a scene, then the score of the anomaly in this region is determined based on the Euclidean distance between the center of clusters and extracted space-time motion influence feature.

For the GAE, The social force model is introduced by [5] to detect the abnormal behaviors in crowd scenes. By placing a grid of particles on the frame and moving them with the average of optical flow to estimate the interaction forces. And after that, the LDA is used to determine the frame abnormality. Also, Wang et al. [17] capture the frame motion information using a histogram of optical flow orientation feature descriptor. After that, OC-SVM is exploited to label current frame as normal or anomalous. In [21] the crowd escape detection is addressed by modeling the motion of the crowd within the non-escape and escape situation using the Bayesian scheme. They use the optical flow estimation method [22] to describe the motion in crowded scenes. Based on the foreground patch that is determined through the optical flow magnitude, the class-conditional probability density functions of flow attributes: position, magnitude and direction are estimated. A concept of potential destinations is introduced to estimate the class-conditional PDFs of flow directions in nonescape situation, while the divergent centers' notion used to estimate class-conditional PDFs of flow directions in the case of escape. Our method belongs to this class. It is inspired by the method presented in [25] in which the presented visual representation is local and depends on spatio-temporal motion influence feature. Rather, our technique relies on frame-based STACOG feature descriptor which is a global feature descriptor. Another difference regarding [25] is the use of K-medoids rather than K-means clustering technique. A reconstruction cost is used by a number of authors as an anomaly criterion. [3], [4], [18]. Since, the intuition behind the reconstruction cost is that a normal event is probable to get a small reconstruction cost which results from sparse reconstruction coefficients, whereas the anomalous event is different from any of the normal basis, and then gets a large reconstruction cost which results from dense representation. In [4] a multiscale histogram of optical flow is used as feature descriptor for several spatial or temporal structures for sparse representation. While Li et al. [18] detect the global abnormal event by employing histogram of maximal optical flow projection feature descriptor which is extracted from salience map of the optical flow field.

3. PROPOSED METHOD

In this section, the main stages of our algorithm are outlined in Fig. 2. The underlying assumption of the method presented here is that the abnormal event differs from normal ones in their space-time motion pattern. So, STACOG features that considered as spatio-temporal representation is extracted from video sequences. We then perform K-medoids clustering using training features. During the test phase, the anomaly score of each frame is determined from distances between frame-based feature STACOG and center of the clusters.

3.1 Feature Extraction

STACOG descriptor [19] is an effective tool for extracting shift-invariant Motion features in the spatio-temporal domain, i.e., velocity and accelerations. STACOG descriptor [19] captures the geometric characteristics of a motion shape by exploiting the local relationships among the space-time gradients from image sequences. Consider the training image volume $I(x,y,t)$, the space-time gradient vector can get from derivatives (I_x, I_y, I_t) at each space-time point in the video sequence as demonstrated in Figure. 3 (a). The magnitudes of

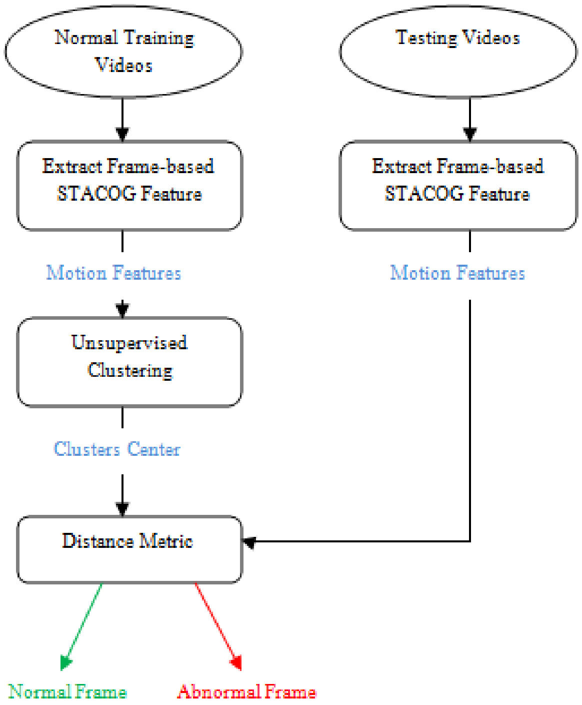


Figure. 2: The flowchart of the proposed abnormal events detection algorithm in crowded scenes

the gradient vectors are computed as $m = \sqrt{I_x^2 + I_y^2 + I_t^2}$. The two angles that represent the gradient vector alongside the magnitude are The spatial orientation $\theta = arctan(I_x, I_y)$ in x-y plane and the temporal elevation $\phi = arcsin(I_t/m)$ over t. The space-time gradient orientation which determined by the spatial orientation and temporal elevation is positioned into B orientation bins on a unit sphere by voting weights to the nearest bins to set up a B-dimensional vector v as indicated in Figure. 3(b). The gradient magnitude m and its vector v determine The Nth order auto-correlation function for the spacetime gradients as follows [19]:

$$R_N(u_1, \dots, u_N) = \int min[m(p), \dots, m(p + u_N)]v(p) \otimes \dots \otimes (p + u_N)\, dp, \qquad (1)$$

where $[u_1, u_2, \dots, u_N]$ are displacement vectors from the reference point p = (x,y,t), and tensor product is indicated by $\otimes$.

In this paper, we consider $N \in \{0,1\}, u_{1x,y} \in \{\triangle p, 0\}, and\ u_{1t} \in \{\triangle t, 0\}$ as indicated in [19], where 4p represent spatial interval while $\triangle$t represent the temporal interval. For $N \in \{0,1\}$, we get the STACOG features as follows:

$$0^{th} order{:}\ F_0 = \sum m(p)h(p), \qquad (2)$$

$$1^{st} order: F_1(u_1) = \sum min[m(p),\ m(p + u_1)\ ] v(p) v(p + u_1)^T. \quad (3)$$

Figure. 4 illustrates thirteen various composition patterns of (p, p + u_1). Therefore, B + 13B^2 is dimensional of the STACOG features descriptor (F_0 and F_1). it is notable that $\triangle r/\triangle t$ is very related to motion velocity. The frame-based STACOG features comprise of the zeroth-order features F_0 in Eq. 2 and the first-order features F_1 in Eq. 3. The former is used to describe gradients , while the last is employed to characterize curvatures.

3.2. Abnormal Events Detection

The space-time motion features extracted by STACOG descriptor from N normal training frames are denoted as F = { f_1,f_2,…,f_N }. We utilize the K-medoids clustering technique to partition F to k clusters and their centers are acquired as follows : C = [c_1,c_2,…,c_k]. Given a test frame, the Euclidean distance between a motion features extracted by STACOG and each center is computed and the smallest one is used as the anomaly score of this frame as indicated in the following equation:

$$S_i = \min_k \|f_i - c_k\|^2 \quad (4)$$

The smaller the value of an anomaly score (S_i), the less likely an abnormal event is to occur in the respective frame. Therefore, the current frame is classified as abnormal If the value of the anomaly score is larger than the predefined threshold.

4. EXPERIMENTAL RESULTS

In order to appraise the accuracy and performance of proposed technique, we performed experiments on a laptop with Intel Core i5-3210M CPU 2.50GHz, 6G RAM and publicly available dataset: UMN dataset [7] and PETS2009 dataset [8].

4.1 UMN Dataset

The UMN dataset [7] has videos of eleven various scenarios that represent an escape event captured at three different scenes. Each video begins with a group of people walking normally, Then series of anomalous events occur like individuals are scattering in all directions or running in one direction. The videos are captured with resolution 320 * 240. The two outdoor scene are Lawn scene which comprises two scenarios with a total of 1453 frames and the Plaza scene that consists of three scenarios with a total of 2142 frames. The indoor scene contains six scenarios with a total of 4144 frames. For the parameter setting in this experiment, the value of k in K-medoids is set to 5, and first 400 normal frames of each scene except for indoor scene, first 300 normal frames are used for training. The remaining frames of each scene are used for testing. For the frame-based STACOG feature descriptor, We use six layers along the hemisphere latitude and each one contains four orientation bins except the pole layer contains one bin as indicated in [19]. Thus, orientation bins count becomes 21 (B= 21). The temporal interval t.

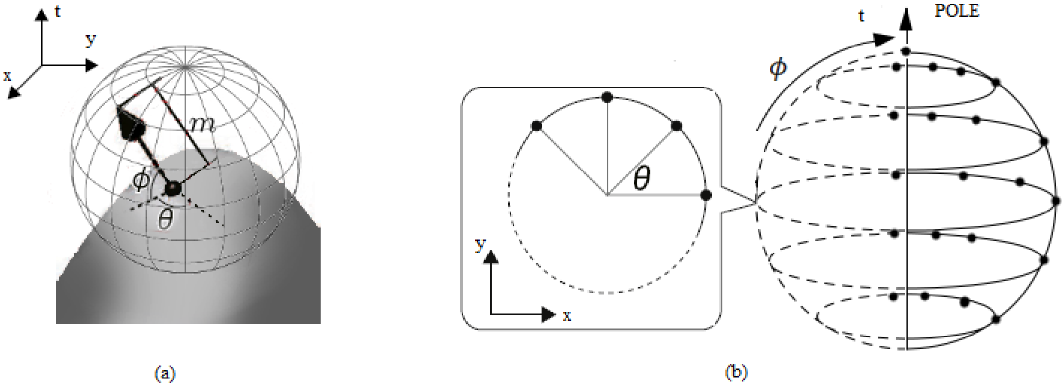


Figure. 3: (a) The 3D gradients. (b) The orientation bins along latitude where the opposite directions ignored and longitude on a hemisphere(from [19])

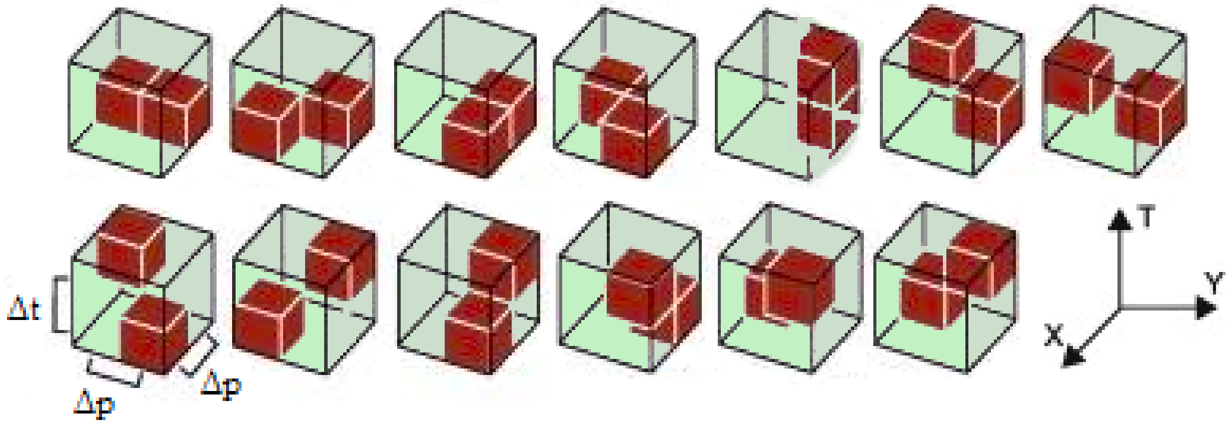


Figure. 4: Patterns arrangement of local pairs.

is set to 1 and The spatial displacement interval r is selected automatically according to [19]. Figure. 5 shows the qualitative results of the proposed algorithm where the detection result is represented by green color in the respective bar if crowd behavior is normal. In contrast, the anomalous crowd behavior represented by red color. The false positive detections in Figure. 5 are a result of the abrupt transition from a scenario to another within the same scene. Since the extraction of STACOG features of a test frame requires the three previous frames to compute the first derivative ([-1 0 1]). Therefore, the STACOG descriptor of those frames is dissimilar to the STACOG descriptors of normal training frames. To analyze results of presented algorithm quantitatively, two evaluation criteria commonly used are the area under the ROC curve (AUC) and the equal error rate (EER). Table 1, Table 2 shows The AUC and EER comparisons of the competing methods. The competitors comprise methods based on Optical Flow [5] , social force model(SF) [5], NN [4], sparse reconstruction cost (SRC) [4], histogram of optical flow orientation (HOFO) [17], and histogram of maximal optical flow projection (HMOFP) [18]. From Table 1, we can notice that the AUC of the presented method Which falls between 0.915 and 0.991, outperforms Optical Flow [5], Social Force [5], and NN [4] and is comparable to the others. At the EER, smaller EER indicates a better performance of the method. The EER of our method is 4.6 % which is 1.64 higher than 2.8% of Sparse[4], the state-of-the-art method as shown in Table 3. The reason for high EER value for the presented method is returned to the synchronization issue at frame labels. For Plaza scene, we have noted that the ground truth label is delayed 37; 45and 42 frames[1]. The value of AUC raised to 97.27% by ground truth error correction. In addition, The EER of the presented method is 3.1 which is comparable with competing methods. This demonstrates the robustness of our method for a global anomalous crowd behavior detection. The implementation of the presented algorithm in Matlab yielded 25 fps which fulfill the requirement of the real-time processing.

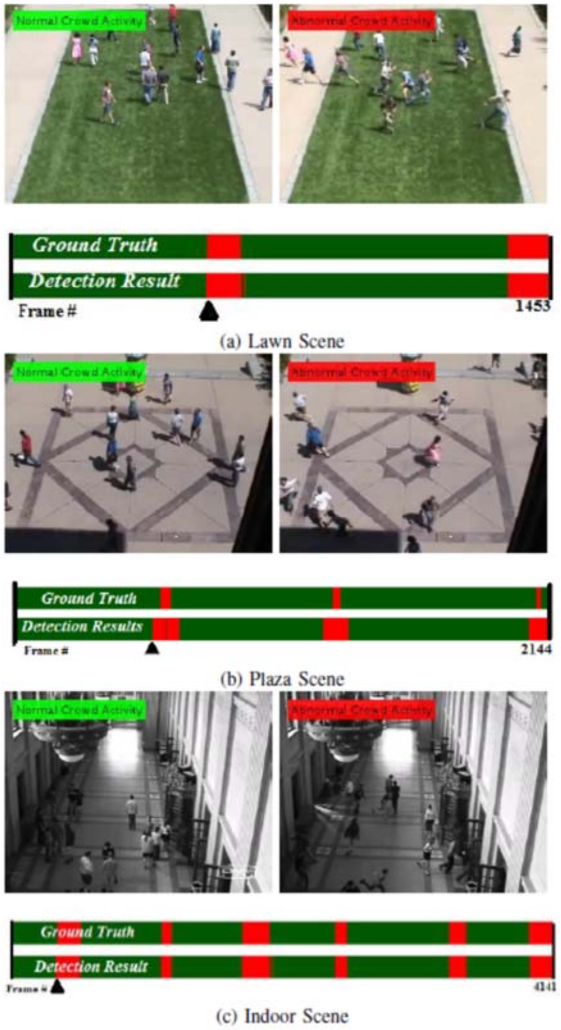
Table 1: AUC comparison of the competing methods on UMN dataset.

Method	AUC		
	Lawn	Indoor	Plaza
Optical Flow [5]	84%		
Social Force [5]	96%		
Nearest neighbor [4]	93%		
Sparse [4]	**99.5%**	97.5%	96.4%
HOFO [17]	98.45%	90.37%	**98.15%**
HMOFP [18]	98.69%	94.07%	97.68%
Ours	98.28%	**99.17%**	91.54%

[1] In Plaza Scene, we noted that a group of people begun running from the 6160th frame of first scenario, 6840th frame of second scenario and 7660th frame of third scenario. But the anomalous behavior in ground truth started from the 6197th,6885th and 7702th frame of first, second and third scenario respectively.

Table 2: EER comparison of the presented method with others on UMN dataset

Method	EER
Social Force [5]	12.6%
H-MDT CRF [20]	3.7%
Sparse [4]	**2.8%**
Ours	4.6%



(a) Lawn Scene

(b) Plaza Scene

(c) Indoor Scene

Figure. 5: The qualitative results of the anomalous crowd event detection in the Lawn, Plaza, and Indoor scene of UMN dataset, respectively. The frames labels in each scene are represented by the ground truth and the detection bar, respectively. The normal frame is represented by green color while the abnormal one is represented by red color. the first frame of the scene is shown in the left column and the right column present the first detected abnormal frame (black triangles).

4.2. PETS2009 Dataset

PETS2009 dataset [8] includes three different subsets, each one of them has a growth in scene complexity of crowd scenarios. The three subsets are S1 which is related to counting of people and density estimation , S2 that concerns individuals tracking, and S3 which involve event detection and its identification. In this experiments, we use two crowd scenarios of S3 subset to do testing. Every one of these scenarios comprise four sequences which captured from various cameras views. Identifying the escape events from those sequences is a challenging task. Because of differences in the lighting and field of view that appear significantly. The first scenario characterizes the movement of a group of people from the right direction to left one where they are walking at the first then begin running from 41th frame. The number of frames in each sequence is 107 frames. The second scenario describes three set of persons moving normally towards the crossroads and remaining at this crossroads for a short period. Then, from 335th frame, they began scattering. The number of frames in each sequence is 378 frames.

To make a comparison between the performance of our algorithm and [21], we do the evaluation procedure as theirs. for each view, we use thirty or one hundred arbitrarily chosen normal frames for training depending on the view in the first or second scenario. Table 3 and 4 demonstrate the accuracy comparison of our method with chaotic invariants (CI) [21], the social force model (SF) [21], the force field (FF) method [21], and the Bayesian model (BM) [21] for escape event detection in these two scenarios. From Table 3, we can note that the average accuracy of our method in the first scenario is 90.95% which is 0.98% higher than 88.92% of the BM, the best performing method. Meanwhile, the processing time of the presented method is 0.04 second per frame under the Matlab environment while the processing time of the BM approach is 0.96 second per frame. The reason behind the low accuracy of FF approach is that FF employed solely motion direction to represent crowd behavior. So, it fails in the detection of escape behavior in case of individuals running without changing the direction of their movement in the first scenario. Whilst BM takes into consideration the motion direction and its velocity to model the motion of crowd, and hence the average accuracy is comparable with us. For the view 1 and view 4 of the second scenario, the proposed method obtains less result comparing with others except for SF method as shown in Table 4. This essentially due to the captured videos of those views stopped awhile then resumed. Consequently, computing of STACOG feature descriptor is affected. Regardless of this, Our method in the second scenario obtains a better average accuracy of 92.13% in comparison with the competing methods. Figure. 6, illustrates our method detection results on samples from the first view of the first scenario and second one of the second scenario in the S3 subset.

Table 3: Accuracy comparison for several methods including the proposed method, Bayesian model , Force field, chaotic invariants and social force on the first scenario from S3 subset

	Ours	BM	FF	CI	SF
First View	**94.29%**	92.45%	37.74%	56.60%	63.21%
Second View	**89.52%**	83.02%	37.74%	83.02%	70.76%
Third View	**90.48%**	89.62%	37.74%	81.13%	52.83%
Fourth View	89.52%	**90.57%**	37.74%	52.83%	48.11%
average accuracy	**90.95%**	88.92%	37.74%	68.40%	58.73%

Table 4: Comparison of accuracy for different methods on the second scenario from S3 subset

	Ours	BM	FF	CI	SF
First View	93.07%	**96.01%**	94.50%	94.95%	91.22%
Second View	**95.47%**	94.15%	63.83%	92.02%	89.36%
Third View	**97.33%**	95.21%	95.48%	94.15%	94.68%
Fourth View	82.67%	91.49%	**96.81%**	89.36%	64.63%
average accuracy	92.13%	**94.22%**	87.66%	92.62%	84.97%

5. CONCLUSION

Through this paper, we proposed an unsupervised method for global abnormal crowd event detection. First, the method is computing STACOG descriptor of the input video sequence. Second, the K-medoids clustering algorithm is used to partition the STACOG descriptors of training frames into a set of clusters, and are then used to determine whether a frame is normal or not using a distance metric. The proposed anomaly detection method was tested on benchmark dataset: UMN dataset, PETS2009. Experiments show that the proposed method achieves comparable results with the state-of-the-art methods in terms of accuracy while requiring a low computational cost than alternative approaches. On the basis of the experimental results, we conclude the following :

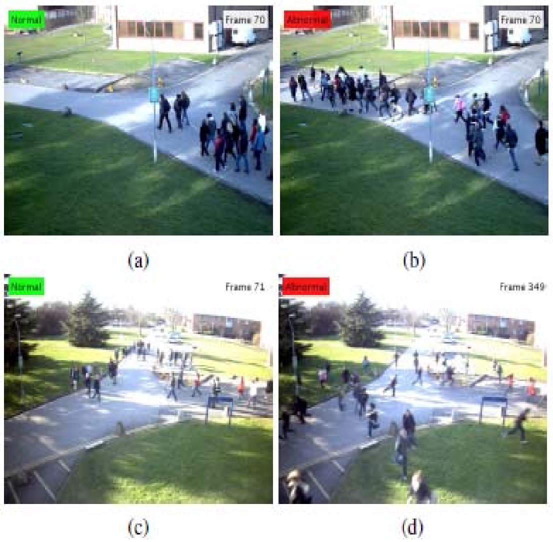


Fig. 6: Detection results of our method on samples from first view of the first scenario (a-b), second view of the second scenario (c-d). For detection results on full image sequences (see: Annexure 1)

- STACOG feature descriptor is effective to represent video event for anomaly detection.
- The proposed method satisfies the demand of real-time performance. In comparison with the best competing method BM [21], its processing time is faster than the ones of BM by 26%.

The potential future research is to investigate how to detect and localize the anomalous regions of a scene (i.e. the regions where the abnormal events occur). This can be achieved by dividing the video frame into overlapping regions and STACOG descriptor is computed for each region separately. Also, we need to study how to improve the accuracy of the proposed method through the use of features of auto-correlation of streak flow [26]. Streak flow captures motions of the crowd accurately by encapsulating the velocity field for a period of time.

REFERENCES:

[1] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 34, no. 3, pp. 334–352, Aug 2004.

[2] M. J. Roshtkhari and M. D. Levine, "An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal com-positions," Computer vision and image understanding, vol. 117, no. 10,pp.1436–1452, 2013.

[3] 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.

[4] 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.

[5] 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.

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

[7] "Univ. minnesota. unusual crowd activity dataset of university of min-nesota." [Online]. Available: http://mha.cs.umn.edu/movies/ crowdactivity-all.avi

[8] J. Ferryman and A. Shahrokni, "Pets2009: Dataset and challenge," in 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Dec 2009, pp. 1–6.

[9] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," IEEE transactions on pattern analysis and machine intelligence, vol. 30, no. 3, pp. 555–560, 2008.

[10] S. Zhou, W. Shen, D. Zeng, and Z. Zhang, "Unusual event detection in crowded scenes by trajectory analysis," in Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on. IEEE, 2015, pp. 1300–1304.

[11] X. Mo, V. Monga, R. Bala, and Z. Fan, "Adaptive sparse representations for video anomaly detection," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 4, pp. 631–645, 2014.

[12] 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.

[13] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on. IEEE, 2011, pp.55–61.

[14] 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.

[15] O. Boiman and M. Irani, "Detecting irregularities in images and in video," International journal of computer vision, vol. 74, no. 1, pp. 17–31, 2007.

[16] R. V. H. M. Colque, C. A. C. Junior,´ and W. R. Schwartz, "Histograms of optical flow orientation and magnitude to detect anomalous events in videos," in Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on. IEEE, 2015, pp. 126–133.

[17] T. Wang and H. Snoussi, "Detection of abnormal visual events via global optical flow orientation histogram," IEEE Transactions on Information Forensics and Security, vol. 9, no. 6, pp. 988–998, 2014.

[18] A. Li, Z. Miao, and Y. Cen, "Global anomaly detection in crowded scenes based on optical flow saliency," in Multimedia Signal Processing (MMSP), 2016 IEEE 18th International Workshop on. IEEE, 2016, pp. 1–5.

[19] T. Kobayashi and N. Otsu, "Motion recognition using local auto-correlation of space–time gradients," Pattern Recognition Letters, vol. 33, no. 9, pp. 1188–1195, 2012.

[20] W. Li, V. Mahadevan, and N. Vasconcelos, “Anomaly detection and localization in crowded scenes,” IEEE transactions on pattern analysis and machine intelligence, vol. 36, no. 1, pp. 18–32, 2014.

[21] S. Wu, H. S. Wong, and Z. Yu, “A bayesian model for crowd escape behavior detection,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 1, pp. 85–98, Jan 2014.

[22] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, “High accuracy optical flow estimation based on a theory for warping,” in Proc. Eur.Conf. Comput. Vision, pp. 25–36, 2004.

[23] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, “Textures of opticalflow for real-time anomaly detection in crowds,”2011 8th IEEE Inter-national Conference on Advanced Video and Signal Based Surveillance(AVSS), 2011,pp. 230–235.

[24] M. J. Black and P. Anandan, “The robust estimation of multiple motions:Parametric and piecewise-smooth flow fields,” Computer Vision and Image Understanding, vol. 63, 1996, pp. 75–104.

[25] D.-G. Lee, H.-I. Suk, S.-K. Park, and S.-W. Lee, “Motion influencemap for unusual human activity detection and localization in crowdedscenes”, IEEE transactions on circuits and systems for video technology,vol. 25, no. 10, 2015, pp. 1612–1623.

[26] R. Mehran, B. E. Moore, and M. Shah, “A streakline representation of flow in crowded scenes”, in European conference on computer vision. Springer, 2010, pp. 439–452

Annexure 1:

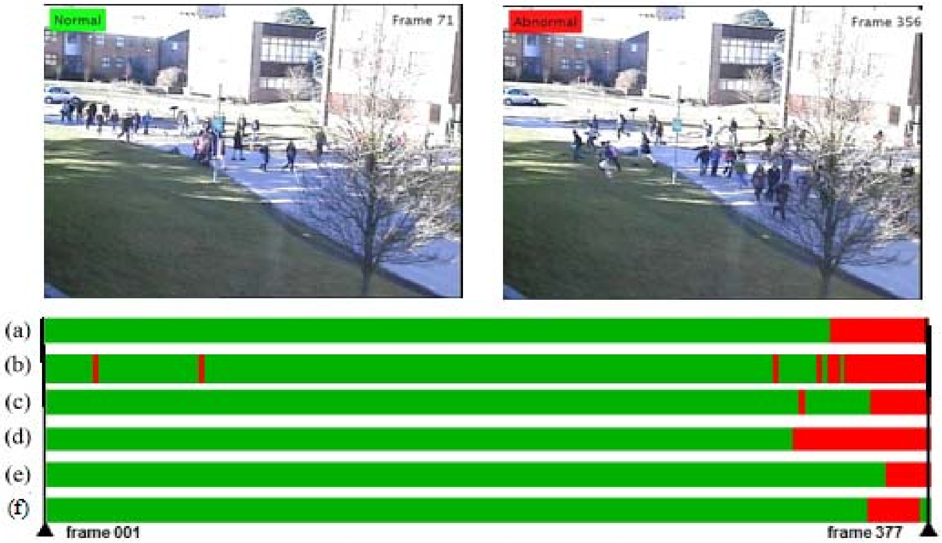


Fig. 7: Detection results of the competing methods on image sequence of third view for the second scenario. (a) Ground truth. (b) Result of the proposed method. (c) Result of BM. (d) Result of FF. (e) Result of CI. (f) Result of SF.

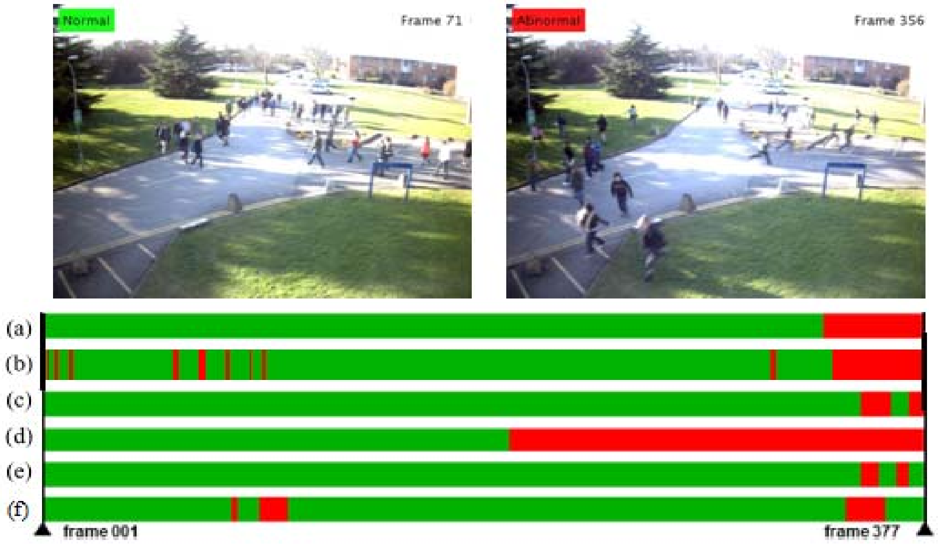


Fig. 8: Detection results of the competing methods on image sequence of second view for the second scenario. (a) Ground truth. (b) Result of the proposed method. (c) Result of BM. (d) Result of FF. (e) Result of CI. (f) Result of SF.