Panic Detection in Crowded Scenes

Ahmed B. Altamimi
College of Computer Science and Engineering
University of Hail
Hail, Saudi Arabia
altamimi.a@uoh.edu.sa

Habib Ullah
College of Computer Science and Engineering
University of Hail
Hail, Saudi Arabia
h.ullah@uoh.edu.sa

Abstract—**A crowd is a gathering of a huge number of individuals in a confined area. Early identification and detection of unusual behaviors in terms of panic occurring in crowded scenes are very important. Panic detection comprises of formulating normal scene behaviors and detecting and identifying non-matching behaviors. However, panic detection and recognition is a very difficult problem, especially when considering diverse scenes. Many methods proposed to cope with these problems have limited robustness as the density of the crowd varies. In order to handle this challenge, this paper proposes the integration of different features into a unified model. Discriminant binary patterns and neighborhood information are used to model complex and unique motion patterns in order to characterize different levels of features for diverse types of crowd scenes, focusing in particular on the detection of panic and non-pedestrian entities. The proposed method was evaluated considering two benchmark datasets and outperformed five existing methods.**

Keywords—crowd analysis; anomaly detection; congestion; bottleneck identification

I. INTRODUCTION

Crowd scene analysis is major problem in computer vision. The challenges of crowd scene analysis are found in gatherings which attract many people, such as sports, festivals, social, political and religious gatherings. Important problems and security concerns are raised in huge crowds. Therefore, dealing with the crowd scenes is a very challenging problem in the subfield of video surveillance, due to the complex behaviors of individuals. A method developed for this purpose could be deployed for detecting critical crowd levels, people counting, and anomalies in such environments. Detecting panic situations in such scenes becomes extremely important for crowd control and safety. Furthermore, in public gatherings, it is important to know the ongoing situation and the behavior of people attending an event, as it can provide useful information for future event planning and public space design. This study aims at detecting panic and non-pedestrian entities representing abnormal behaviors in crowd scenes, integrating robust features into a unified model. The structure of the proposed method consists of the Discriminant Binary Pattern (DBP) [1] and the neighborhood orientation information [2]. These features are combined and classified using a multi-class support vector machine [3]. In the crowd scene context, different feature integration could represent individuals, and patterns that reflect the collective behavior of crowd arise from the interaction of individuals. This kind of modeling deals with the challenges of crowded scenes in terms of various distributions, sparseness, consistency and inconsistency of the scene.

Many computer vision based methods have been proposed to detect panic and other anomalies in crowded scenes. However, most of them are modeled to address a specific scene, while different representations of motion and appearance are analyzed with different techniques. The proposed method explores the properties of abnormal situations, considering both the anomalies in terms of panic and non-pedestrian entities. This method computes dense motion trajectories from input videos. Considering the computed trajectories, different features are combined into a unified hybrid model. Furthermore, the underlying complex motion information is explored, which can formulate complex events in the crowd. This information consists of mid-level characteristics that model the distance between low and high-level features for representing anomalous situations. The proposed hybrid features do not limit the type of features or scenes, enabling the extension of the technique to broader research fields. For experimental evaluation, the method is applied on two widely used benchmark datasets. Additionally, the proposed technique is compared with five existing methods, namely the Spatio-Temporal Texture Model (STTM) [4], the Abnormal Crowd Behavior Model (ACBM) [5], and the spatio-temporal volume based methods: Mixture of Dynamic Texture (MDT) [6], Data Mining for Anomaly (DMA) [7], and Optical Flow and Texture for Anomaly (OFTA) [8].

II. LITERATURE REVIEW

Recently, deep learning methods have shown promising results in different applications in the field of computer vision. Therefore, different computer vision and deep learning based methods have been successfully employed to solve problems related to crowded scenes in real time [9-11]. Crowded scenes can be categorized on gathering locations, e.g., congested urban areas, historical events organized in museums, music and fun concerts, arrival and departure zones in airports, and different sports in stadiums. Various proposed methods can generally help safety, security staff and administrators in identifying security methods. Different crowd analysis methods consider the input data in different ways. For example, some methods process videos from stored archives. The methods

Corresponding author: Habib Ullah

introduced in [12, 13] extract frames from video CCTV cameras. Some researchers introduced unified frameworks for crowd congestion analysis, individual counting, crowd flow analysis on pedestrian pathways, crowd fight analysis, panic and bottleneck detection during exit and entrance [14-16]. In [17], a hybrid method was used to analyze crowd from different angles, such as crowd moving in different directions, computing statistics related to a static crowd, calculation of areas covered by potential crowd, and crowd's social behavior analysis in the form of groups or as a whole. This technique could be deployed to take into account different aspects in public locations, considering a unified framework. Moreover, the techniques introduced in [18-20] could be used to direct crowd flow to avoid stampede or other adverse events. In addition, crowd analysis approaches driven by tracking techniques [21-23] can be exploited to track suspicious individuals in a congested area, in order to ensure security. The tracking driven technique in [24] could be used to understand how individuals are behaving and moving in crowded scenes. These techniques can be used to divert crowd flows in public places to avoid congestion and stampede. In general, crowd analysis methods can be divided into two categories. In the first, each entity in the crowd is considered individually and they are combined to model crowd behavior. However, these methods can only be considered when the scene is not occupied by a very dense crowd. In the second category, methods treat the whole crowd as a single entity. These techniques perform very well in dense crowded scenes. The techniques in [25-27] fall in the first category, and they can be utilized to track individual people in the crowd, calculate the waiting time of a static crowd, compute the length of individuals in a queue, count the number of people in crowded scenes, and analyze people's engagement in crowd scenes. The techniques presented in [28-30] fall in the second category, and they can help security staff and administrators to have in insight into a crowd scene and get the essence of the behavior as a whole.

Different methods have been proposed in dealing with the problems of crowd analysis considering different factors such as individual tracking, abnormal situation detection, flow analysis, and situation awareness. The method proposed in [31] combines different data from different sensors to get more accurate localization of the target in the crowded scene, introducing a modified ensemble of Kalman filters to deal with a high degree of nonlinearity. The method in [6] detects and recognizes anomalies in crowded scenes using features related to the appearance and the dynamics of the crowd, considering both temporal and spatial information to compute mixtures of dynamic textures. The model presented in [32] uses a multi-scale deep convolutional neural network to count the number of individuals in a single image with arbitrary crowd density and perspective. Authors in [33] presented an aggregation of ensembles considering pre-trained ConvNets and a pool of classifiers in order to detect anomalies in crowded scenes. Their approach was inspired by the concept that a set of different fine-tuned CNNs represents various levels of semantic characterization, and therefore encodes a very rich set of robust features. The method in [34] was an agent-based crowd simulation model that exploited a path planning strategy. Different elements, including traveling distance and

turning angle, can be used efficiently for path planning in crowded scenes. An inspiration of momentum from Physics was exploited in [35], integrating foreground information with an object's motion, based on background subtraction, feature computation, behavior identification, and anomaly detection. Authors in [36] discussed various approaches for the analysis of both high and low density crowds to facilitate personal mobility, safety, and security, enabling assistive robotics in crowded scenes. This study demonstrated the main challenges and solutions for the analysis of unified behaviors, in order to explore interpersonal relations and social interaction of people in crowd scenes. Authors in [37] investigated a confined set of socio-cognitive crowd behaviors, discovering the interrelated connection between the movements of individuals to analyze different crowd behaviors. This was a layered approach segmenting visual analysis and semantic crowd behaviors.

TABLE I.   METHODS, FEATURE MODELS AND DATASETS USED

Ref.	Publication date	Features/Model	Dataset
[9]	2017	Statistical	UCD
[11]	2018	Nested motion	PETS2009UMN
[12]	2019	Spatio-temporal	Collected data
[13]	2019	Intra-frame classification	WWW crowd
[19]	2019	Social model	CUHK crowd
[20]	2019	Depth information	UCF
[23]	2019	Texture features	PNNL parking
[24]	2019	Spatio-temporal	UCF crowd
[26]	2016	Simulation features	Mall dataset
[27]	2019	Entropy features	Crowd dataset
[29]	2019	Probabilistic model	UMN crowd
[6]	2010	Dynamic texture	UCSD
[33]	2020	Aggregation of ensembles	Collected data
[35]	2019	Histogram features	UCSD, PETS
[7]	2018	Visual features	WWW crowd
[8]	2019	Optical flow features	CUHK crowd
[38]	2016	Holistic features	PNNL parking
[39]	2011	Low-level features	Collected data
[40]	2011	Adaptive features	UMN
[4]	2019	Spatio-temporal	CUHK crowd

A method based on artificial bacteria colony was investigated in [7], where the optical flow of frames was exploited to get the foreground segments with entity motions as layers. Surveillance problems using image processing and machine learning techniques were studied in [8], exploring algorithms of identifying abnormal crowd behaviors, proposing a unified method for anomaly detection, considering both crowd motion and texture-based analysis. A deep learning method was proposed in [41], calculating crowd density from individual images representing different crowd densities, exploiting both deep and shallow fully convolutional networks to estimate the density map of a crowd image. This method is good in encoding both high level semantic information and low level features. Authors in [42] analyzed stationary crowd by computing the duration of a static foreground pixel. For this purpose, they used dynamic constraints, spatial and temporal features, and mixed partials. Authors in [38] used holistic features for crowd anomaly detection, fusing together different features including crowd collectiveness, conflict, density and mean motion speed. Authors in [43] mapped a given crowd scene to its density, avoiding people occlusion in dense crowd, foreground and background similarities, and huge changes in camera viewpoints. In [44] a combined framework was

presented to deal with the different elements of a crowd scene, including people counting, abnormal behavior detection, and different crowd density. A contextual pyramid technique was presented in [45] to produce a robust map of crowd density and people count, by using both global and local contextual features. Recently, some advanced deep learning methods [46, 47] were presented, which could improve the performance of crowd analysis techniques. The modeling of complex crowd scenes using deep Elman neural network architecture in [46], could extract in a better way the complex structure of a crowd, and emulate perfectly such dynamic systems.

III.   PROPOSED METHOD

Panic detection and non-pedestrian entity localization is a challenging task due to its associated complexities. For this purpose, discriminative features are explored [1]. There is a lack in researches that efficiently assess the underlying structure of crowd scenes, investigating the most discriminant pattern representation of abnormal situations. This study's purpose is to exploit the connection between the motion feature extraction and their discriminability, proposing an integrated model characterized by the discriminative power of combining different features. This framework renders a deep insight into the optimal feature extraction for anomaly detection. This study avoids the modeling of traditional features that use single-type representation and features unable to encode the motion and crowd structure information. Moreover, this study is exploiting local neighborhood features. A refined feature [2] combines more information of the intrinsic structure and is effective for crowd scene analysis. The method's flow diagram is shown in Figure 1.
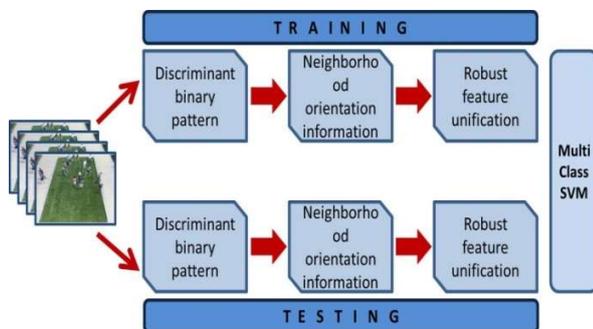

Fig. 1.   Proposed method's flow diagram

The effectiveness of the selected features is not limited to abnormal situation detection, they could also be widely used in many other applications in the field of computer vision and machine learning. The proposed method computes repetitions of different orientations in localized areas of a crowd video. This framework gets some inspiration from edge orientation histograms [48], scale-invariant feature transform descriptors [49], and shape contexts. A significant difference lies in the unification of completely different features, as the model is formulated on a dense grid of all pixels in each frame. The intuition behind the feature unification was that localized crowd pattern appearance and layout within a video frame could be formulated by the distribution of intensity and orientation information. Previous studies [50-52] extracted

different types of orientation features. However, this study investigated the important discriminability of feature unification for generalized purposes, exploring a model to characterize the discriminative power of various types of orientations, so that more discriminative features can be fused together. The proposed method is effective and compact for abnormal situation detection. Hence, it demonstrates complex and unique motion patterns, and it could be considered as a mid-level hierarchy to integrate the space between low and high-level information in order to capture complex crowd events. Features represent motion information in the neighborhood of the region under observation. The most significant ability of this method is the unification of features in the local neighborhood. Crowd scene layout (i.e. appearance structure of the scene) captures reliable information in a reasonable neighborhood, discovering important hints to recognize various events of interest in crowds.

The conventional techniques extract gradient locations and orientation information for feature modeling. The proposed method differs from these as it considers the difference between each pixel and the center one, within local windows of its eight neighbors to produce different orientation features, whereas conventional techniques are confined to only horizontal and vertical. Therefore this kind of conventional modeling cannot be effectively applied without any distinction in the significance and the influence of different orientations. The proposed refined features improve the local scene description ability, enhancing the intrinsic structure of the crowd scenes. In this method, each frame is extracted from the crowd video. It is worth noticing that each class label for each event in the scene is considered. Abnormal situations in each crowd video consist of multiple abnormal events. Inspired by [1], the abnormal situation is extracted through the Discriminant Binary Pattern (DBP). DBP can model the scene changes and implicitly formulate the multiple dominant features, as:

$$\zeta = \frac{1}{2\pi\sigma^2}\exp\left\{-\pi\left(\frac{x^2}{\epsilon^2} + \frac{y^2}{\rho^2}\right)\right\}\cos\left(2\pi\mu(x\cos\theta + y\sin\theta)\right) \quad (1)$$

where ζ represents the function to extract direction information. The potential feature will maximize this response, and it can be considered as the prominent direction feature of the abnormal situation. In (1), ϵ and ρ represent the standard deviations of the data surrounding the position x and y. The convolution between ζ and a video frame I is represented by:

$$\xi_j(x,y) = \zeta * \left(255 - I(x,y)\right) \quad (2)$$

where * represents the convolution operation, and functions ζ and ξ obtain convolved results for each pixel of the video frame. In order to identify individual pixels related to abnormal situations in the crowded scene, the following modeling is performed:

$$\theta\left(I(x,y)\right) = argmax\xi_j(x,y)_{\theta_j} \quad (3)$$

where θ represents the orientation of the pixel under observation. Equation (3) characterizes the connection between the direction feature extraction and the discriminability of the abnormal situation. To enhance the method's encoding ability, it is consolidated as [2]:

$$\beta = \{\alpha_{11} - \alpha_c, \alpha_{12} - \alpha_c, \alpha_{13} - \alpha_c, \alpha_{14} - \alpha_c, .., \alpha_{18} - \alpha_c\}$$
$$= \{\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7, \gamma_8\} \quad (4)$$

where α_c represents the central pixel and $\alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14}, .., \alpha_{18}$ represent the neighboring pixels. The intensities of non-unit neighboring pixels are obtained by arithmetic operations, using their adjacent pixel values. Next, the intensity difference between each neighboring pixel and the central pixel is computed. Differences are concatenated to a difference vector that represents the scene. In fact, for each frame, this information is extracted densely for all 9×9 neighborhoods. Features in (3) and (4) are combined into a unified framework as shown in:

$$\Lambda = \sum_{i=1}^{8} \sum_{j=1}^{4} \left(\exp\left(-\frac{j^2}{2\sigma^2}\right)(\theta_{ij} - \beta_{ij}) \right) \quad (5)$$

This equation considers the entire features of the scene and gathers the statistics of local regions to embed more local information and strengthen the robustness of representing complicated crowd motion patterns for abnormal situation detection. The classification of a crowded scene into normal and abnormal situations, involves exploiting a multi-class support vector machine [3]. Two benchmark datasets were used and the training parts of these datasets consist of panic situation and non-pedestrian entity detection. Therefore, a function that shows at most ε deviation from the actually chosen labels for the training data of both datasets is required. This function was intended to be very flat according to the training data of both datasets. In order to elaborate the theoretical background, errors are not considered as long as they are less than ε. The importance of this theory arises from the fact that it is desired to ensure that error lies in some acceptable margin during the classification process of different abnormal events. To start the classification procedure according to the multi-class support vector machine, a linear function is considered as:

$$f(x) = \langle w, x \rangle + b \; with \; w \epsilon \chi, b \epsilon \mathbb{R} \quad (6)$$

where $\langle .,. \rangle$ represents the dot product. Smoothness in this equation is related to the convergence of the smaller *w*, and this is only possible by minimizing its norm. This process belongs to convex optimization formulation according to:

$$minimize \; \frac{1}{2}\|\omega\|^2 \; subject \; to \; \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon \\ \langle w, x_i \rangle + b - y_i \leq \epsilon \end{cases} \quad (7)$$

According to (7), convex optimization is possible if enough training data are available. For this reason, two benchmark datasets were used to train the multi class support vector machine, taking full advantage of the convex optimization process. Standalone features without integration are not very effective, since the physical structure of crowd scene changes over a temporal window. In addition, single type of features is very sensitive to background and illumination changes, scale variations, and crowd flow direction. To handle these problems, the integration of different robust and reliable features should be explored. For instance, entities in the crowd scattered in the scene with a still background, consist a very challenging environment in terms of modeling different events. Therefore, a method based on single type of features suffers if the uncertainty in crowd flow increases over a large scale, such

as a uniform crowd flow versus a random one. The outcome of the traditional and single type of features would be unreliable in such situations. Moreover, if the flow of the crowd is in one direction and it changes randomly due to the occurrence of either panic situation or a non-pedestrian entity, the same features would behave differently. Therefore a unified model for different features is needed, modeling the random flow of crowd and being robust to different variations. The proposed method can effectively cope with these issues.

IV. RESULTS

The detailed experimental analysis procedure utilized the widely used benchmark datasets UCSD [53] and UMN [54]. These datasets are properly annotated benchmarks for the analysis of abnormal detection and localization in crowded scenes. UCSD consists of anomalous entities represented by non-pedestrian entities in scenes. The videos of this dataset were captured with a CCTV camera fixed at an elevation, at a resolution of 238×158 at 10fps, overlooking people in pedestrian pathways. Non-person objects in pathways and anomalous pedestrian motion patterns are treated as abnormal situations. In this dataset, the abnormal entities are bikers, skaters, small carts, and people walking across a pathway or in the park. The video clips in the dataset are divided into 2 subsets: Ped1 and Ped2, and each one is associated with a different crowd scene. Videos recorded from each scene were categorized into different clips, each of them consisting of about 200 frames. Ped1 consists of 34 training and 36 testing videos. Ped2 consists of 16 training and 14 testing videos. In each video, the ground truth annotation includes a binary flag per frame, showing whether an anomalous entity is present. Moreover, there is a subset consisting of 10 videos with manually produced pixel-level binary masks, which recognize the parts or sub-parts consisting of abnormal events with clear boundaries. This was generated for performance analysis, with respect to the ability of anomaly localization and segmentation. Sample images from the UCSD dataset are depicted in Figure 2. In the top row, non-pedestrian entities are shown, which are not allowed in these pedestrian pathways.

Fig. 2.　Pedestrians walking with other entities present. Images from UCSD © SVCL

Qualitative and quantitative experimental analyses were performed. The qualitative results are shown in Figure 3, where

the top row shows sample frames taken from four video sequences. The bottom row shows the anomalous entities annotated and highlighted in light blue. The bicycle riders in the bottom row are detected as anomalous entities, since they are not allowed to use the pedestrian pathways. Similarly, a vehicle is detected in the bottom left frame.
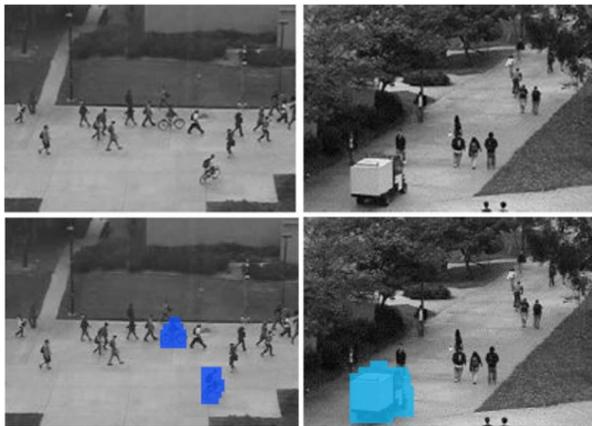

Fig. 3.   The top row shows sample frames taken from four video sequences. The bottom row shows anomalous entities annotated and highlighted in different colors. Images from UCSD © SVCL

For the quantitative analysis, Table II presents the area as a ROC curve (AUC) of the tested methods, in which a larger AUC score shows improved classification results. As it can be noted, the proposed method outputs competitive results against the five reference methods.

TABLE II.   UCSD: AUC PERFORMANCE COMPARISON

Method	STTM [4]	ACBM [5]	MDT [6]	DMA [7]	OFTA [8]	Proposed
AUC	0.77	0.74	0.77	0.73	0.69	0.81

Moreover, experimental analysis in the form of Equal Error Rate (EER) was performed. Figure 4 shows the results in graphs for the five references and the proposed method. Blue and red graphs represent the results for Ped1 and Ped2 subsets respectively. We can see that the proposed method has smaller EER errors compared to the five reference methods.
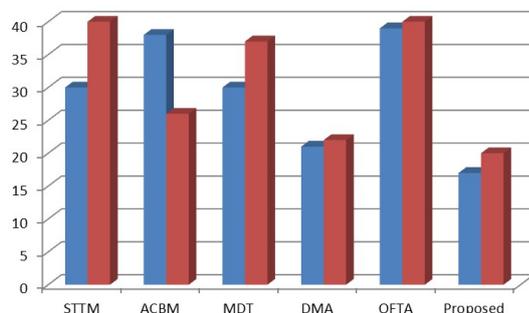

Fig. 4.   UCSD Dataset: Frame level equal error rate comparison

The UMN dataset presents both normal and abnormal crowd video sequences. This dataset comprises of three indoor and outdoor scenes, showing 11 different scenarios of panic events. The dataset consists of 7739 frames in total, with resolution of 320x240 pixels. Each video starts with normal human behaviors, such as walking or standing. Figure 5 depicts sample images from the UMN dataset.

Fig. 5.   People walking and standing. Images from [54], © University of Minnesota

Qualitative results for the UMN dataset are depicted in Figure 6, where the top row shows sample frames taken from four video sequences. The bottom row shows anomalous entities annotated and highlighted in light blue. Panic is detected and highlighted when pedestrians start running in different directions. Table III shows the area under a ROC curve (AUC) of the five references and the proposed method. Again, the proposed method has larger AUC score than the referenced methods.

Fig. 6.   UMN Dataset: Normal frames and panic detection on them. Images from [54], © University of Minnesota

TABLE III. UMN: AUC PERFORMANCE COMPARISON

Method	STTM	ACBM	MDT	DMA	OFTA	Proposed
AUC	0.91	0.75	0.84	0.87	0.79	0.93

Experimental analysis in the form of EER was performed, considering the UMN dataset. Figure 7 shows the results provided for the five references and the proposed method. Blue

and red graphs represent the results for the first two and the following two video sequences of the UMN respectively. As can be noted, the proposed method has smaller EER errors compared to the five reference methods.
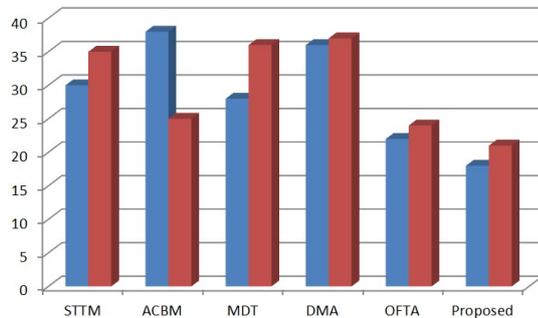

Fig. 7.   UMN Dataset: Frame level equal error rate comparison

V.   CONCLUSION

Many diverse approaches have been proposed to solve the problem of panic detection and anomaly identification. However, most of them are designed to work on specific scenes, where different representations of motion and appearance are analyzed with different models. In this study, the properties of abnormal situations were considered specifically as anomalies in terms of panic, and more general, as anomalies in terms of non-pedestrian entities. Abnormal entities are rare things with unexpected appearance or motion patterns. For anomalous situation detection including panic and non-pedestrian entities, a novel technique was proposed, where dense motion trajectories were computed from the crowd input videos. Considering the computed trajectories, a set of robust features was designed considering HOG, HOF, and MBH. Motion atoms were explored for compact encoding of motion patterns in crowded scenes, representing distinguished motion patterns of crowds. In fact, motion atoms are mid-level characteristics to fuse the distance between low and high-level features for capturing anomalous situations. Since an anomalous situation is described from the view of a feature set, the proposed method can be utilized in different surveillance scenes. Moreover, hybrid features and motion atoms do not limit the type of features or the type of scenes, which helps in extending the proposed technique to broader research fields. The experimental results demonstrated that the proposed approach is effective for real crowd videos containing various types of normal and abnormal activities, in terms of panic situation and the existence of non-pedestrian entities. The proposed method is independent of crowd flow variability and density over temporal windows. Furthermore, the method is not sensitive to crowd concentration in different locations of different scenes. Experimental evaluation was performed considering two benchmark datasets and the proposed method outperformed five known methods both quantitatively and qualitatively.

REFERENCES

[1]   L. Fei, B. Zhang, Y. Xu, D. Huang, W. Jia, J. Wen, "Local discriminant direction binary pattern for palmprint representation and recognition", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 30, No. 2, pp. 468-481, 2020

[2]   W. Zhang, W. Zhang, K. Liu, J. Gu, "A feature descriptor based on local normalized difference for real-world texture classification", IEEE Transactions on Multimedia, Vol. 20, No. 4, pp. 880-888, 2017

[3]   J. T. Zhou, I. W. Tsang, S. S. Ho, K. R. Muller, "N-ary decomposition for multi-class classification", Machine Learning, Vol. 108, No. 5, pp. 809-830, 2019

[4]   Y. Hao, Z. J. Xu, Y. Liu, J. Wang, J. L. Fan, "Effective crowd anomaly detection through spatio-temporal texture analysis", International Journal of Automation and Computing, Vol. 16, No. 1, pp. 27-39, 2019

[5]   Y. Liu, K. Hao, X. Tang, T. Wang, "Abnormal crowd behavior detection based on predictive neural network", 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, October 17, 2019

[6]   V. Mahadevan, W. Li, V. Bhalodia, N. Vasconcelos, "Anomaly detection in crowded scenes", 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Fransisco, USA, August 5, 2010

[7]   J. Ramos, N. Nedjah, L. de Macedo Mourelle, B. B. Gupta, "Visual data mining for crowd anomaly detection using artificial bacteria colony", Multimedia Tools and Applications, Vol. 77, No. 14, pp. 17755-17777, 2018

[8]   P. Ingole, V. Vyas, "Anomaly detection in crowd using optical flow and textural feature", in: Soft Computing and Signal Processing, Advances in Intelligent Systems and Computing, Vol. 900, pp. 723-732, Springer, 2019

[9]   S. D. Khan, M. Tayyab, M. K. Amin, A. Nour, A. Basalamah, S. Basalamah, S. A. Khan, "Towards a crowd analytic framework for crowd management in Majid-al-Haram", 17th Scientific Meeting on Hajj & Umrah Research, 2017

[10] K. Ahmad, N. Conci, F. G. De Natale, "A saliency-based approach to event recognition", Signal Processing: Image Communication, Vol. 60, pp. 42-51, 2018

[11] H. Ullah, A. B. Altamimi, M. Uzair, M. Ullah, "Anomalous entities detection and localization in pedestrian flows", Neurocomputing, Vol. 290, pp. 74-86, 2018

[12] R. Nawaratne, D. Alahakoon, D. De Silva, X. Yu, "Spatiotemporal anomaly detection using deep learning for real-time video surveillance", IEEE Transactions on Industrial Informatics, Vol. 16, No. 1, pp. 393-402, 2019

[13] K. Xu, T. Sun, X. Jiang, "Video anomaly detection and localization based on an adaptive intra-frame classification network", IEEE Transactions on Multimedia, Vol. 22, No. 2, pp. 394-406, 2019

[14] J. Wan, A. Chan, "Adaptive density map generation for crowd counting", IEEE International Conference on Computer Vision, Seoul, South Korea, October 27 – November 2, 2019

[15] X. Jiang, Z. Xiao, B. Zhang, X. Zhen, X. Cao, D. Doermann, L. Shao, "Crowd counting and density estimation by trellis encoder-decoder networks", IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, June 15-20, 2019

[16] H. Yu, G. Pan, L. Zhang, Z. Li, M. Pan, "Translation domain segmentation model based on improved cosine similarity for crowd motion segmentation", Journal of Electronic Imaging, Vol. 28, No. 2 2019

[17] N. Bisagno, B. Zhang, N. Conci, "Group LSTM: Group grajectory prediction in crowded scenarios", in: Proceedings of the European conference on computer vision (ECCV), pp. 213-225, Springer, 2018

[18] R. Trabelsi, I. Jabri, F. Melgani, F. Smach, N. Conci, A. Bouallegue, "Complex-valued representation for RGB-D object recognition", in: Pacific-Rim Symposium on Image and Video Technology, pp. 17-27. Springer, Cham, 2017

[19] L. Pan, H. Zhou, Y. Liu, M. Wang, "Global event influence model: integrating crowd motion and social psychology for global anomaly

detection in dense crowds", Journal of Electronic Imaging, Vol. 28, No. 2, 2019

[20] M. Xu, Z. Ge, X. Jiang, G. Cui, P. Lv, B. Zhou, C. Xu, "Depth information guided crowd counting for complex crowd scenes", Pattern Recognition Letters, Vol. 125, pp. 563-569, 2019

[21] X. Alameda-Pineda, E. Ricci, N. Sebe, "Multimodal behavior analysis in the wild: an introduction", in: Multimodal Behavior Analysis in the Wild, pp. 1-8. Academic Press, 2019

[22] M. Ullah, H. Ullah, N. Conci, F. G. B. De Natale, "Crowd behavior identification", 2016 IEEE International Conference on Image Processing, Phoenix, USA, September 25-28, 2016

[23] H. Kim, J. Han, S. Han, "Analysis of evacuation simulation considering crowd density and the effect of a fallen person", Journal of Ambient Intelligence and Humanized Computing, Vol. 10, No. 12, pp. 4869-4879, 2019

[24] Y. Hao, Z. Xu, Y. Liu, J. Wang, J. Lun Fan, "Effective crowd anomaly detection through spatio-temporal texture analysis", International Journal of Automation and Computing, Vol. 16, No. 1, pp. 27-39, 2019

[25] J. Li, L. Wei, F. Zhang, T. Yang, Z. Lu, "Joint deep and depth for object-level segmentation and stereo tracking in crowds", IEEE Transactions on Multimedia, Vol. 21, No. 10, pp. 2531-2544, 2019

[26] K. Shimura, S. D. Khan, S. Bandini, K. Nishinari, "Simulation and evaluation of spiral movement of pedestrians: towards the tawaf simulator", Journal of Cellular Automata, Vol. 11, No. 4, 2016

[27] X. Zhang, X. Shu, Z. He, "Crowd panic state detection using entropy of the distribution of enthalpy", Physica A: Statistical Mechanics and its Applications, Vol. 525, pp. 935-945, 2019

[28] D. Kang, Z. Ma, A. B. Chan, "Beyond counting: comparisons of density maps for crowd analysis tasks—counting, detection, and tracking", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 29, No. 5, pp. 1408-1422, 2018

[29] M. Dimitrievski, P. Veelaert, W. Philips, "Behavioral pedestrian tracking using a camera and lidar sensors on a moving vehicle", Sensors, Vol. 19, No. 2, 2019

[30] T. Figueiredo, R. Castro, "Passengers perceptions of airport branding strategies: the case of Tom Jobim International Airport–RIOgaleoo, Brazil", Journal of Air Transport Management, Vol. 74, pp. 13-19, 2019

[31] Z. Zhang, K. Fu, X. Sun, W. Ren, "Multiple target tracking based on multiple hypotheses tracking and modified ensemble Kalman filter in multi-sensor fusion", Sensors, Vol. 19, No. 14, 2019

[32] D. Ji, H. Lu, T. Zhang, "End to end multi-scale convolutional neural network for crowd counting", in: Eleventh international conference on machine vision (ICMV 2018), Vol. 11041, International Society for Optics and Photonics, 2019

[33] K. Singh, S. Rajora, D. K. Vishwakarma, G. Tripathi, S. Kumar, G. S. Walia, "Crowd anomaly detection using aggregation of ensembles of fine-tuned ConvNets", Neurocomputing, Vol. 371, pp. 188-198, 2020

[34] S. K. Tan, N. Hu, W. Cai, "A data-driven path planning model for crowd capacity analysis", Journal of Computational Science, Vol. 34, pp. 66-79, 2019

[35] S. D. Bansod, A. V. Nandedkar, "Crowd anomaly detection and localization using histogram of magnitude and momentum", The Visual Computer, Vol. 36, pp. 609-620, 2020

[36] N. Conci, N. Bisagno, A. Cavallaro, "On modeling and analyzing crowds from videos", in: Computer Vision for Assistive Healthcare, pp. 319-336, Academic Press, 2018

[37] M. S. Zitouni, A. Sluzek, H. Bhaskar, "Towards understanding socio-cognitive behaviors of crowds from visual surveillance data", Multimedia Tools and Applications, 2019

[38] M. Marsden, K. Mc Guinness, S. Little, N. E. O' Connor, "Holistic features for real-time crowd behaviour anomaly detection", IEEE International Conference on Image Processing, Phoenix, USA, September 25-28, 2016

[39] A. B. Chan, N. Vasconcelos, "Counting people with low-level features and bayesian regression", IEEE Transactions on Image Processing, Vol. 21, No. 4, pp. 2160-2177, 2011

[40] M. D. Zeiler, G. W. Taylor, R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning", 2011 International Conference on Computer Vision, Barcelona, Spain, November 6-13, 2011

[41] L. Boominathan, S. S. S. Kruthiventi, R. V. Babu, "Crowdnet: a deep convolutional network for dense crowd counting", in: Proceedings of the 24th ACM international conference on ,ultimedia, pp. 640-644, ACM, 2016

[42] S. Yi, X. Wang, C. Lu, J. Jia, H. Li, "L_0 regularized stationary-time estimation for crowd analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 5, pp. 981-994, 2017

[43] D. B. Sam, S. Surya, R. V. Babu, "Switching convolutional neural network for crowd counting", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, July 21-26, 2017

[44] M. Marsden, K. Mc Guinness, S. Little, N. E. O' Connor, "Resnetcrowd: a residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification", 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, August 29–September 1, 2017

[45] V. A. Sandagi, V. M. Patel, "Generating high-quality crowd density maps using contextual pyramid CNNs", 2017 IEEE International Conference on Computer Vision, Venice, Italy, October 22-29, 2017

[46] L. B. Sallah, F. Fourati, "Systems modeling using deep Elman Neural Network", Engineering, Technology & Applied Science Research, Vol. 9, No. 2, pp. 3881-3886, 2019

[47] S. R. Basha, J. K. Rani, "A comparative approach of dimensionality reduction techniques in text classification", Engineering, Technology & Applied Science Research, Vol. 9, No. 6, pp. 4974-4979, 2019

[48] B. Alefs, G. Eschemann, H. Ramoser, C. Beleznai, "Road sign detection from edge orientation histograms", 2007 IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, June 13-15, 2007

[49] J. Clemons, "SIFT: scale invariant feature transform", available at: https://pdfs.semanticscholar.org/19d1/c9a4546d840269ef534f6c1c8e379 8ce81ac.pdf

[50] P. H. Gosselin, N. Murray, H. Jegou, F. Perronnin, "Revisiting the fisher vector for fine-grained classification", Pattern Recognition Letters, Vol. 49, pp. 92-98, 2014

[51] S. Dasgupta, "Experiments with random projection", in: Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, pp. 143-151, ACM, 2013

[52] G. V. de Lima, P. T. Saito, F. M. Lopes, P. H. Bugatti, "Classification of texture based on bag-of-visual-words through complex networks", Expert Systems with Applications, Vol. 133, pp. 215-224, 2019

[53] Statistical Visual Computing Lab–UC San Diego, "UCSD anomaly detection dataset", available at: http://www.svcl.ucsd.edu/projects/ anomaly/dataset.htm

[54] University of Minnesota, "Monitoring Human Activity", available at: http://mha.cs.umn.edu/