Received: 15 March 2021, Accepted: 5 October 2021, Published: 30 November 2021, Publisher: UTP Press, Creative Commons: CC BY-NC-ND 4.0

A COMPARATIVE ANALYSIS OF TECHNIQUES FOR CROWD BEHAVIOUR DETECTION IN DENSE SCENES

Aliyu Nuhu Shuaibu[1*], Aamir Saeed Malik[2], Ibrahima Faye[3]

[1]Department of Electrical and Electronics Engineering, University of Jos, Nigeria
[2]Department of Computer Systems, Faculty of Information Technology, Brno University of Technology, Czech Republic
[3]Department of Fundamental & Applied Science, Universiti Teknologi PETRONAS, Malaysia

[*]Email: nuhua@unijos.edu.ng

ABSTRACT

Behaviour analysis is considered a critical area of research in the computer vision research community. Recently, visual monitoring systems of human gathering have found application in different areas such as safety, security, entertainment, and personal archives. Although many approaches have been proposed, certain limitations exist and many unresolved issues remain open. The objective of the paper is to present recent advances on abnormal human behaviour analysis and hierarchical crowd behaviour classification based on the level of complexity. The paper also provides a clear perspective with a broad and in-depth review of the research conducted in this area. In addition, it points out unresolved problems that demand improvement. In general, researchers can use this paper as a starting point for further advancement of behavioural analytic methods to propose novel approaches as well as an exploration of approaches that have received meager attention. This study also investigates the performance comparison of state-of-the-art techniques on anomaly detection: hand-crafted feature approach, and deep learning approach. Finally, limitations of the current methods and promising future research directions are presented.

Keywords: *Human detection, crowd anomaly, deep learning, crowd analysis, artificial intelligence*

INTRODUCTION

The word crowd is regarded as a phenomenon that comprises large sparse individuals or objects sharing a common goal in a given environment as presented in Figure 1. A large number of pedestrians and objects in the crowded scene resulted in a variant of challenges. The necessity for safety and security in crowded places is an additional reason for research in human behaviour analytic. The area visual surveillance in recent times has been given global support and attention due to the importance of surveillance of public places: shopping centers, banks, airports, train stations, subway stations, sports areas, traffic control, and congestion prediction. This helps authorities avoid dangerous scenarios and detection of unusual activity in the scene. Crowd analysis is classified into the following area of applications: crowd behaviour, crowd segmentation, crowd tracking, crowd motion detection, and crowd density estimation. More intensive reviews on the existing works [1]-[2] reported a wider view of the application of computer vision and artificial intelligence in the efficient representation of unusual activity in pedestrian scenes. Therefore, in this paper, we focus primarily on unusual human behaviour recognition based on hand-crafted engineering features and deep learning approaches.

There are different contextual terms used in the definition of human behaviour including atomic actions, action sequence, and activity [3]. Atomic actions correspond to instantaneous entities upon which an action is formed. Actions correspond to

Figure 1 Frames of a video sequence showing pedestrians and crowded scenes obtained from the following datasets: (a) Pets2014, (b) Subway, (c) CUHK, (d) Pets2010, (e) UCSD, and (f) Data-driven

a sequence of atomic actions that fulfill a task or purpose; while activities consist of a sequence of actions over space and time. The overall objective of any behaviour recognition system is to detect, analyse and interpret the contextual activity. This may be classified as "suspicious event" [4], "irregular behaviour" [5], "uncommon behaviour" [6], "abnormal behaviour" [7]-[11], and "anomaly" [12]-[15]. The interpretation of human behaviour is subject to variations of context, for instance: ambiguous, rare, and unusual. To bring about intelligent aid to human capability, the crowd analysis aims to create an efficient platform with the capability of mimicking the human intelligence by learning and classifying behaviours in normal and abnormal contexts.

Issues regarding human behaviour identification

It is evident that the interaction among individuals in a crowd scene is usually unnoticeable, thus abnormal behaviour representation poses a great challenge. Furthermore, abnormal behaviour spread-out very fast within a high-density crowded setup and hence results in difficulty to understand individual behaviour. In high-density scenarios, object detection becomes more challenging since the proportion of object pixels decreases with rising crowd density [16]. In addition, ambiguous appearance information delivered from only a few pixels on the target object poses a great challenge in continuous object tracking. Interpretation of the activity of an object is sometimes harder when only partial features of the target object are available. The physical nature of the visual scene may act as the main contributing factor to the occlusion problem, which may lead to the inaccurate detection of objects.

Comparisons with previous reviews

In the last decade, several reviews have been published on human behaviour recognition. A recent survey by Turaga et al. [17], covered human behaviour recognition techniques with classification based on the level of complexity and recognition approaches from simple actions classes to complex activities. A study in [18], focuses on vision-based human action recognition and discusses the following; action representation and action classification as part of every action recognition framework. In like manner, a study by Aggarwal & Ryoo [19], the authors presented a review on behaviour recognition techniques in the following sequence (atomic action, people interaction, human-object interaction, and group behaviour). In their study, they categorised behaviour recognition techniques into two classes: single-layered method and hierarchical method. The single-layered method recognises human activity directly based on video frame sequences whereas the hierarchical method recognises high-level human activity by utilising simple sub-event activities features. Correspondingly, surveys by Chandola [20], Patcha & Park [21] presented general reviews on the area of abnormal behaviour detection. In broader and recent reviews [22]-[23], both extensively analysed human behaviour detection techniques in the following categories: initialisations, tracking, pose estimation, and recognition. The survey in [22], performed a literature survey from 2000 to 2015 which extended the previous survey in [23], which analysed human behaviour surveysfrom 2000 to 2006.

Additionally, another closely related review on behaviour recognition has been presented recently in [24]. They conducted a detailed survey on activity recognition using semantic space features such as pose, poselet, attributes, related object, and scene context. They extensively exploited the aforementioned features to recognise behaviour in video data and still images. According to Moeslund et al. [23], an action primitive is composed of simple joint movement that is translated at the functional level of the human body. Actions are simple motion levels that are commonly executed by a single person such as in cycling, walking, running, etc. Activities are coordinated actions among a few individuals which involved relatively more complex semantics. Behaviours involved more complex and high-level representation of activities carried out among few to a large number of individuals.

However, despite the fast-growing of crowd behaviour analysis, yet the majority of the state-of-the-art approaches focused more on non-crowded scenes. Thus, there are few surveys on crowd behaviour analysis to date. According to Zhan et al. [25], conventional computer vision techniques fail when large crowded video scenes are analysed. The author presents a comprehensive survey on current crowd analysis work in computer vision from sociological, psychological, and computer graphics perspectives. In another recent work in [1], the authors presented a survey on crowd analysis based on people counting/density estimation, tracking in crowded scenes, and behaviour understanding. Recently, an attempt to bridge the gap between physics and biology in the area of crowd analysis was presented in [26]. The survey provides an in-depth analysis and discussions on crowd analysis from the physics and biologically inspired perspectives.

The work we present is different in several approaches from the previous surveys. For instance, in [1],[25], the authors presented an extensive review on crowd analysis from computer vision and computer graphics perspectives. An analysis of vital features needed for a particular crowd analysis application was also reported. Moderate training computation was generally observed in the surveys presented in Table 1. However, it was observed that qualitative and quantitative comparisons between performances of several multiple object behaviour (sparse crowd) are absent in previous studies. Therefore, this paper seeks answers to the following questions: (1) Could group detectors and descriptors properties are used to improve the performance of abnormal behaviour recognition? (2) What are the modifications and improvements made for the abnormal behaviour recognition system in the last decade? (3) What are the optimal parameter settings and datasets for the application of abnormal behaviour analysis models? Answers to these questions could be beneficial to the research community. Moreover, these answers could provide a benchmark for expert researchers as well as starting point for novice researchers. A summary of published reviews on crowd analysis is presented in Table 1.

MATERIALS AND METHOD

Crowd behaviour recognition approaches

As presented in Junior, Musse & Jung [1] and Kuan et al. [32], there are two main classification approaches for crowd behaviour analysis based on Hand-crafted Feature Approach (HFA): *object-based* and *holistic-based* approach. The overall Schematic flow pattern of crowd behaviour analysis based on recent advances is presented in Figure 2.

Object-based

The object-based approach analyses crowd activities based on a collection of individual motion, where tracking of individual trajectories or regions of

Table 1 Summary of the survey papers published on crowd analysis

Publication	Year	Title	Methodology	Shortcomings
Zhan et al. [25]	2008	Crowd analysis: A survey	Describe crowd motion as thinking fluids	There is no quantitative evaluation provided for the state of the methods
Junior et al. [1]	2010	Crowd analysis using computer vision techniques	The review presented crowd analysis techniques in computer vision and other integrated models for modeling and event inference	There is no quantitative evaluation provided for the state of the methods
Moore et al. [27]	2011	Visual crowd surveillance through a hydrodynamics lens	Present a comprehensive review on pedestrian tracking, density estimation, event detection, validation, and simulation.	Limited evaluation; tested on a few datasets with few video sequences
Sjarif et al. [28]	2011	Detection of abnormal behaviour in crowd scene: A review	Presented a survey on abnormal behaviour detection in crowded scenes from 2000 to 2010	There is no quantitative and quantitative evaluation provided for the state of the methods
Jo et al. [29]	2013	A review on physics-based methods for group and crowd analysis in computer vision	Presented the advances in Physics-based methods for crowd analysis and group formation in computer vision	There is no qualitative and quantitative evaluation provided for the state of the methods, no discussions regarding crowd datasets
Loy et al. [30]	2013	Crowd counting and profiling: Methodology and evaluation	Present comprehensive and recent advances in crowd analysis from 2010 to 2014. The publicly available dataset used for performance evaluation are highlighted	Limited evaluation; tested different datasets with no common evaluation framework
Li et al. [31]	2014	Crowded scene analysis: A survey	Focuses on computer vision techniques (Feature representation and model learning) in Crowd analysis	The algorithms were not tested on a common platform

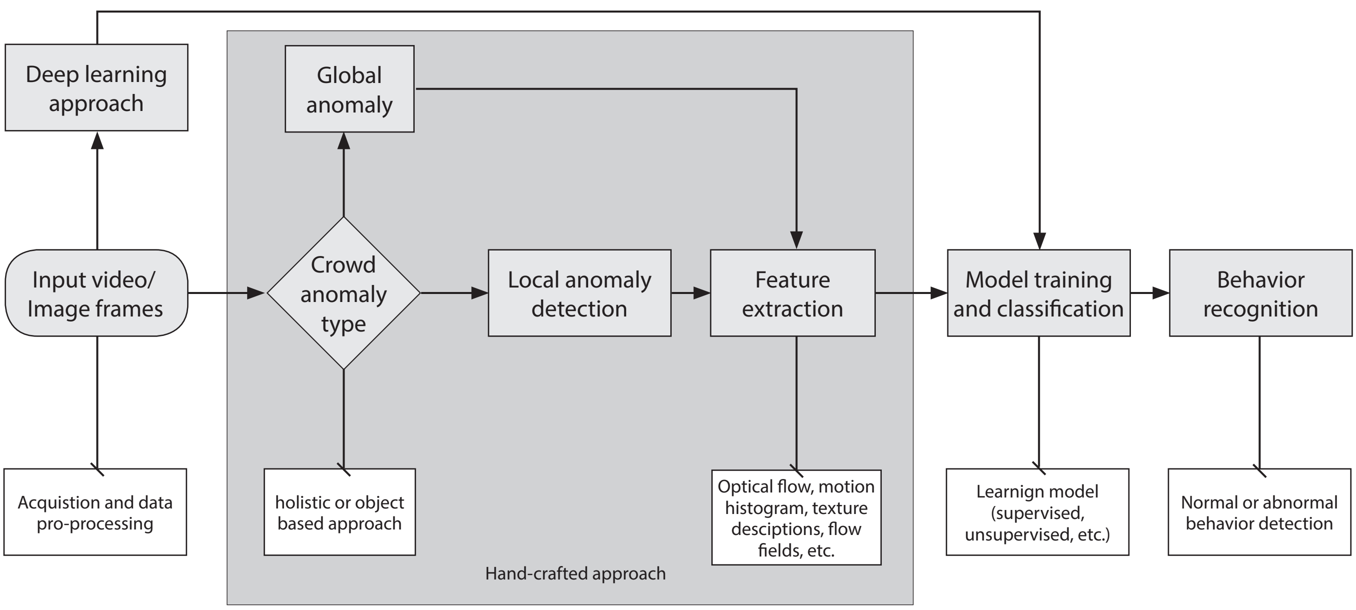

Figure 2 Schematic flow pattern of crowd behaviour analysis based on recent advances

interest is needed [33]. To understand crowd activity, segmentation, tracking or detection should be performed in advance. In the case of a simple scene, high precision is achievable using such techniques; however, in crowded situations, object occlusion affects the precision of pedestrian tracking or segmentation. Consequently, this results in low detection performance. Similarly, the computational cost will be enormous in high-density scenes. The object-based approach according to Junior, Musse & Jung [34], tries to identify individual behaviour in the crowd scene. This is possible by utilising the conventional methods, such as detection and segmentation in low-density crowds. For instance, the identification of a person moving against the stream of a crowd could be associated with a possibly dangerous situation. In a snatch theft, commonly the theft comes closer to the individual from behind, and this detection needs the person identification for tracking purposes. As earlier mentioned, tracking of trajectories is required in an object-based approach. This is because it is essential to preserve individual pattern records for recognition. Then, the learned patterns are compared to the query trajectories for unusual behaviour detection. Therefore, the object-based approach applies to moderately crowded scenarios.

Holistic-based

In a holistic-based approach, crowds are treated as a single entity. In this approach, top-down taxonomy is applied, which handles the challenge associated with the tracking of people in the occluded scene. The holistic approach mostly utilised global information such as the main flow direction. Also, local information is not necessarily needed in a holistic approach such as a single individual movement in an opposite pattern to the main crowd flow. Recently, Lagrangian Particle Dynamic (LPD) approaches for crowd behaviour analysis have been proposed [35]. The LPD was used to model the dynamics of the crowded scene by tracking the region of the instability of the crowd flow field. In their work, an optical flow-based method is applied in detecting the movements of a crowd, which are used to generate a velocity field. The algorithm in Weina, Collins & Ruback [35], is capable of detecting flow field regions of instability. Correspondingly, another study in [36] presented a framework that detects a region of interest using low-level information extracted from the optical flow. Based on the aforementioned techniques, only direction and velocity are regarded as the motion information, the limitation of their findings is focused based on motion information only. Similarly, these techniques cannot withstand categories of unique motion features such as congestion, sources, and sinks.

Anomaly types

Abnormal Behaviour Detection (ABD) is an essential aspect of crowded scenarios, which has attracted numerous researchers' attention. Nevertheless, the issue of ABD is quite an open challenge, and research efforts are scattered not only in methods but also in the interpretation of the problem, hypothesis, and goals [37].

Abnormal behaviour detection in the crowded scene can be categorised into two classes: Global Anomaly Detection (GAD) and Local Anomaly Detection (LAD) as presented in Yang, Junsong & Ji [38]. A detailed description of the GAD and LAD is presented below.

Global anomaly detection (GAD)

Global anomaly detection aims to distinguish the normal condition of crowd motion from abnormal ones. The related algorithms for GAD generally tend to detect changes or events based on the estimated motion observed on the whole frame sequence. It is essential for the GAD to accurately detect the presence of abnormality and subsequently, indicate the beginning and the end of the event, as well as the transition between them. It should be noticed that the holistic approach for crowd analysis [39]-[42] could be applied for global anomalies detection.

Local anomaly detection (LAD)

In addition to global anomaly detection, we are often curious to know where exactly the anomaly took place. To achieve this, many LAD methods have been proposed. According to a recent review, the common methods can be classified into vision-based approaches (HMM, dynamic texture, bag-of-words, sparse representation, and manifold learning); and physics-inspired approaches (flow field, social force model, and crowd energy model) [31]. The crowd density estimation technique presented by Moore et al. [43] was also utilised to detect unusual behaviour. According to their findings, the density variations could mean specific dangerous situations in a scene and applied support vector machine to classify the extracted features based on congestion and emptiness in captured scene sequence. Another recent approach for LAD explored an integrated framework using a Mixture of Dynamic Texture models for the appearance and dynamic of the complex scene presented by Weixin, Mahadevan & Vasconcelos [44] which proposed a joint detector for spatial and temporal anomalies in a crowd scene. Based on their algorithms spatial and temporal unusual behaviours are interpreted based on different scales. Temporal normalcy is modeled with dynamic texture and allows behaviour detections based on deviation from pre-defined behaviours. Spatial normalcy is modeled with discriminate saliency detectors that permit the detection of behaviours that deviate from the pre-defined behaviour of a given crowd scene.

Deep learning approach for spatio-temporal features

As earlier mentioned, the traditional method named HFA involved: feature extraction, then feature representations which involved converting all frames into the fixed-sized description. The most commonly used approach involved quantisation of all extracted features using a learned k-means dictionary and assigning a visual codeword over a given duration of the sequence into a histogram of the spatio-temporal pattern. Finally, a classifier such as Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Ensemble subspace discriminant, or Support Vector Machine (SVM) is used for training and testing and thereby distinguished between the behaviour classes.

However, in Deep Learning models, the aforementioned three stages are replaced by a single neural network layer that is trained from the row data (pixel values or motion channels) to a classifier (recognition). Recently, the work presented in Karpathy et al. [45], treated a sample of the video sequence as a bag of short and fixed-sized clips. The connectivity pattern in the time dimension is described using: early fusion, late fusion, or slow fusion.

Time information fusion in convolution neural network (CNN)

According to Karpathy [45], data fusion into the temporal domain of CNNs can be carried out at an early stage by varying the convolution filters concerning time, or by the late approach. The late approach involved positioning two different sequences at a distance in time and fusing their outputs to a fully-connected layer as presented in Figure 3. The two network branches are shown (appearance and motion channels) have the same convolutional network architecture. The notations are used to represent the CNNs network parameters as reported in Jing et al. [46].

$$E = -\frac{1}{\mathrm{M}}\sum_{m=1}^{M} t_m \log p_m + (1-t_m)\log(1-p_m) \qquad (1)$$

where M denotes the total number of output neurons, t_m $(m = 1,..., M)$ is the target classes, and p_m $(m = 1,..., M)$. Equation 1 represents the output probability prediction of the output class based on the Softmax classifier at the fully connected layer.

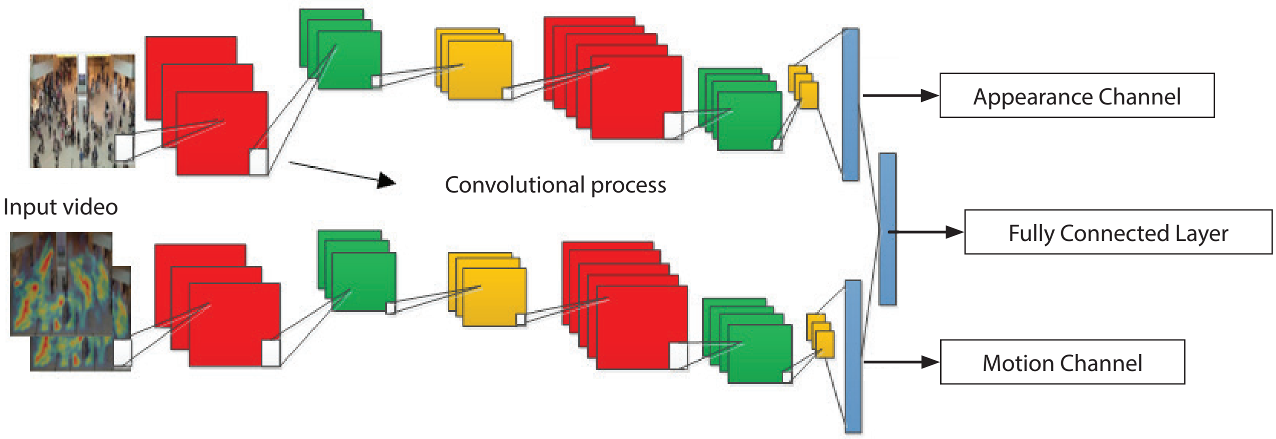

Figure 3 Deep Convolutional Neural Network Architecture: appearance and motion channels

Fusion models

The early fusion combines details of the entire frame time window on the pixel level. An implementation in Ouyang et al. [47] was done by modifying the filter on the initial layer. The direct connectivity with the frame pixels allows the framework to detect motion directions and speed. The late fusion combined two branches of a single network frame as earlier presented at convolution layer C(256,3,1) with common parameter and varying frame distance of 15 frames apart before the fully-connected layer fusion of the branch stream. The next type of fusion is slow fusion which is considered as the intermediate or balanced combination between the aforementioned approaches, which slowly fused the spatial and temporal details of the input data to the network. The activation is computed by connecting all convolution layers in time together with the spatial and temporal convolution as presented in Adam et al. [48].

Appearance and motion channels

The conventional input of the deep learning model is the motion map of the individual frame (RGB channels) or multiple frames as presented in Karpathy et al. [45]. Recently, three different scene-independent motion attribute (collectiveness, stability, and conflict) presented in [46] were used as the complement of the appearance channel.

DATASETS

Characteristics summary of publicly available datasets for anomaly detection are as follows:

University of Minnesota (UMN) dataset [49] contains videos of 11 different sequences of escape scenarios. The scene was captured in both indoor and outdoor. Each clip starts with normal behaviours scenario and ends with abnormal behaviours. Recorded behaviour includes walking and sudden running. The main limitations of this dataset are that: (1) it is relatively small (scenes 1, 2, and 3 contain two, six, and three anomaly instances), (2) it has no pixel-level ground truth, (3) the anomalies are staged, and (4) it produces very salient changes in the average motion intensity of the scene. As a result, several methods achieve near-perfect detection.

U-turn dataset presented consists of one video sequence (roughly 6,000 frames of size 360 240) recorded by a static camera overlooking the traffic at a road intersection. The video is split into two clips of equal length for cross-validation and anomalies consist of illegal vehicle motion at the intersection. The main limitations of this dataset are 1) the limited size, 2) the absence of pixel-level ground truth, and 3) the sparseness of the scenes [50].

University of California San Diego (UCSD) Anomaly Detection Dataset [51] was acquired with a stationary camera mounted at an elevation, overlooking pedestrian walkways. The crowd density in the walkways was variable, ranging from sparse to very crowded scenes. In the normal setting, the video contains only pedestrians. Abnormal events are due to either: the circulation of non-pedestrian entities in the walkways and anomalous pedestrian motion patterns. Commonly occurring anomalies include bikers, skaters, small carts, and people walking across a walkway.

SUBWAY dataset is proposed by Adam et al. [48] including two videos: "entrance gate" (1 hour 36 minutes long with 144,249 frames) and "exit gate" (43 minutes long with 64,900 frames). Normal behaviours include people entering and exiting the station; abnormal consists of people moving in the wrong direction (exiting the entrance or entering the exit) or avoiding payment. The main limitations of this dataset are 1) reduced number of anomalies, and 2) predictable spatial localisation (entrance and exit regions).

WWW means Who does What at somewhere crowd dataset [52]. It comprises about 10,000 videos from 8,257 crowded scenes. The videos in WWW dataset are from real-world scenarios collected from different sources and obtained from different cameras. It has a resolution of 640 × 360.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This paper presents wider review coverage (from 2006 to 2015) on abnormal crowd behaviour analysis and comparative analysis based on computer vision techniques. Perfectives from hand-crafted features to deep learning approaches are presented, as this categorisation has been applied intensively in the field of action recognition and complex behaviour analytics. A comprehensive search was carried out on the published works, from high-quality journals and conference proceedings related to the keywords: "Human Behaviour", "Crowd Analysis", "Artificial Intelligence", "Anomaly Detection", and "Deep Learning" resulting in the selection of 69 journal and conference papers. Publications in this coverage were clustered into the following classes: object-based, holistic-based, local anomaly detection, global anomaly detection; publicly available datasets, and performance evaluation. In addition, the review lists some of the strengths and weaknesses of the state-of-the-art methods. Despite the quantity of the literature reviewed, there exists a possibility of inaccessible academic databases or papers written in other languages.

Finally, with rapid development in machine learning and artificial intelligence, a more intelligent framework is needed for anomaly detection and localisation of human behaviour. Additionally, research work should focus more on improving crowd attribute recognition. Major progress is required for behaviour representation in complex dynamic scenes, high-level semantic reasoning, and recognition. From the Industrial point of view, certain advances may be required: view-invariant human motion capture for a different application; reliable tracking and behaviour recognition in clutter scenes. As technology and computational techniques advance, it is expected that more commercial enthusiasm on this topic and real-world applications should emerge. Quite a several modifications and algorithms based on sparse group detectors and descriptors properties have been used for optimal behaviour recognition in crowd analysis tasks. Challenging datasets for the application of abnormal behaviour analysis models have been proposed as presented in section 3. Furthermore, researchers could explore metaheuristic algorithms such as flower pollination algorithm, hybrid cuckoo search algorithm, and particle swarm optimisation in the abnormal behaviour detection in crowded scenes. This is because reviews in the aforementioned areas have clearly shown that these metaheuristic algorithms received little attention from researchers in this domain.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Teknologi PETRONAS, Malaysia for the opportunity to carry out this research work.

REFERENCES

[1] J.C.S.J. Junior, S. Raupp Musse, & C.R. Jung, "Crowd Analysis Using Computer Vision Techniques," *Signal Processing Magazine, IEEE,* 27, pp. 66-77, 2010.

[2] V.J. Kok, M.K. Lim, & C.S. Chan, "Crowd Behaviour Analysis: A Review where Physics meets Biology," *Neurocomputing,* 2015.

[3] S. Gong & T. Xiang, *Visual analysis of behaviour: from pixels to semantics*: Springer Science & Business Media, 2011.

[4] G. Lavee, L. Khan, & B. Thuraisingham, "A framework for a video analysis tool for suspicious event detection," *Multimedia Tools and Applications,* 35, pp. 109-123, 2007.

[5] Y. Zhang & Z.-J. Liu, "Irregular behaviour recognition based on treading track," in *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on*, pp. 1322-1326, 2007.

[6] A. Wiliem, V. Madasu, W. Boles, & P. Yarlagadda, "Detecting uncommon trajectories," in *Digital Image Computing: Techniques and Applications (DICTA), 2008*, pp. 398-404, 2008.

[7] Y.-X. Hung, C.-Y. Chiang, S. J. Hsu, & C.-T. Chan, "Abnormality detection for improving elder's daily life independent," in *Aging Friendly Technology for Health and Independence*, ed: Springer, pp. 186-194, 2010.

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

[9] J. Varadarajan & J.-M. Odobez, "Topic models for scene analysis and abnormality detection," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pp. 1338-1345, 2009.

[10] Y. Wang, K. Huang, & T. Tan, "Abnormal activity recognition in office based on R transform," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, pp. I-341-I-344, 2007.

[11] C.-K. Lee, M.-F. Ho, W.-S. Wen, & C.-L. Huang, "Abnormal event detection in video using n-cut clustering," in *Intelligent Information Hiding and Multimedia Signal*

Processing, 2006. IIH-MSP'06. International Conference on, pp. 407-410, 2006.

[12] L. Kratz & 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*, pp. 1446-1453, 2009.

[13] C. Brax, L. Niklasson, & R. Laxhammar, "An ensemble approach for increased anomaly detection performance in video surveillance data," in *Information Fusion, 2009. FUSION'09. 12th International Conference on*, pp. 694-701, 2009.

[14] T. Xiang & S. Gong, "Video behaviour profiling for anomaly detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 30, pp. 893-908, 2008.

[15] F. Pan & W. Wang, "Anomaly detection based on the regularity of normal behaviours," *1st International Symposium on Systems and Control in Aerospace and Astronautics,* pp. 6 pp.-1046, 2006.

[16] S. Ali & M. Shah, "Floor Fields for Tracking in High Density Crowd Scenes," in *Computer Vision – ECCV 2008*. vol. 5303, D. Forsyth, P. Torr, and A. Zisserman, Eds., ed: Springer Berlin Heidelberg, 2008, pp. 1-14.

[17] P. Turaga, R. Chellappa, V. S. Subrahmanian, & O. Udrea, "Machine recognition of human activities: A survey," *Circuits and Systems for Video Technology, IEEE Transactions on*, 18, pp. 1473-1488, 2008.

[18] R. Poppe, "A survey on vision-based human action recognition," *Image and vision computing,* 28, pp. 976-990, 2010.

[19] J.K. Aggarwal & M.S. Ryoo, "Human activity analysis: A review," *ACM Computing Surveys (CSUR),* 43, p. 16, 2011.

[20] V. Chandola, A. Banerjee, & V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.,* 41, pp. 1-58, 2009.

[21] A. Patcha & J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer networks,* 51, pp. 3448-3470, 2007.

[22] P. Afsar, P. Cortez, & H. Santos, "Automatic visual detection of human behaviour: A review from 2000 to 2014," *Expert Systems with Applications,* 42, pp. 6935-6956, 11/15/ 2015.

[23] T.B. Moeslund, A. Hilton, & V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer vision and image understanding,* 104, pp. 90-126, 2006.

[24] M. Ziaeefard & R. Bergevin, "Semantic human activity recognition: A literature review," *Pattern Recognition,* 48, pp. 2329-2345, 2015.

[25] B. Zhan, D. Monekosso, P. Remagnino, S. Velastin, & L.-Q. Xu, "Crowd analysis: a survey," *Machine Vision and Applications,* 19, pp. 345-357, 2008.

[26] V.J. Kok, M.K. Lim, & C.S. Chan, "Crowd behaviour analysis: A review where physics meets biology," *Neurocomputing,* 177, pp. 342-362, 2016.

[27] B.E. Moore, S. Ali, R. Mehran, & M. Shah, "Visual crowd surveillance through a hydrodynamics lens," *Communications of the ACM,* 54, pp. 64-73, 2011.

[28] N. Sjarif, S. Shamsuddin, & S. Hashim, "Detection of abnormal behaviours in crowd scene: a review," *Int. J. Advance. Soft Comput. Appl,* 4, 2012.

[29] H. Jo, K. Chug, & R.J. Sethi, "A review of physics-based methods for group and crowd analysis in computer vision," *Journal of Postdoctoral Research,* 1, pp. 4-7, 2013.

[30] C.C. Loy, K. Chen, S. Gong, & T. Xiang, "Crowd counting and profiling: Methodology and evaluation," in *Modeling, Simulation and Visual Analysis of Crowds*, ed: Springer, 2013, pp. 347-382.

[31] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, & S. Yan, "Crowded scene analysis: A survey," *IEEE Transactions on Circuits and Systems for Video Technology,* 25, pp. 367-386, 2015.

[32] L. Mei Kuan, K. Ven Jyn, L. Chen Change, & C. Chee Seng, "Crowd Saliency Detection via Global Similarity Structure," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pp. 3957-3962, 2014.

[33] A.M. Cheriyadat & R.J. Radke, "Detecting dominant motions in dense crowds," *IEEE Journal of Selected Topics in Signal Processing,* 2, pp. 568-581, 2008.

[34] J.S.J. Junior, S. Musse, & C. Jung, "Crowd analysis using computer vision techniques," *IEEE Signal Processing Magazine,* 5, pp. 66-77, 2010.

[35] G. Weina, R.T. Collins, & R.B. Ruback, "Vision-Based Analysis of Small Groups in Pedestrian Crowds," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 34, pp. 1003-1016, 2012.

[36] M. Mancas, N. Riche, J. Leroy, & B. Gosselin, "Abnormal motion selection in crowds using bottom-up saliency," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pp. 229-232, 2011.

[37] A. Sodemann, M.P. Ross, & B.J. Borghetti, "A review of anomaly detection in automated surveillance," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* 42, pp. 1257-1272, 2012.

[38] C. Yang, Y. Junsong, & L. Ji, "Sparse reconstruction cost for abnormal event detection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 3449-3456, 2011.

[39] S. Ali & M. Shah, "A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1-6, 2007.

[40] B. Solmaz, B.E. Moore, & M. Shah, "Identifying Behaviours in Crowd Scenes Using Stability Analysis for Dynamical Systems," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 34, pp. 2064-2070, 2012.

[41] W. Shandong, B.E. Moore, & M. Shah, "Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2054-2060, 2010.

[42] L. Kratz & K. Nishino, "Tracking with local spatio-temporal motion patterns in extremely crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 693-700, 2010.

[43] B.E. Moore, S. Ali, R. Mehran, & M. Shah, "Visual crowd surveillance through a hydrodynamics lens," *Communications of the ACM,* 54, pp. 64-73, 2011.

[44] L. Weixin, V. Mahadevan, & N. Vasconcelos, "Anomaly Detection and Localization in Crowded Scenes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 36, pp. 18-32, 2014.

[45] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, & F.-F. Li, "Large-Scale Video Classification with Convolutional Neural Networks," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 1725-1732, 2014.

[46] S. Jing, K. Kai, L. Chen Change, & W. Xiaogang, "Deeply learned attributes for crowded scene understanding," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pp. 4657-4666, 2015.

[47] W. Ouyang, P. Luo, X. Zeng, S. Qiu, Y. Tian, H. Li, et al., "Deepid-net: multi-stage and deformable deep convolutional neural networks for object detection," *arXiv preprint arXiv:1409.3505,* 2014.

[48] A. Adam, E. Rivlin, I. Shimshoni, & D. Reinitz, "Robust Real-Time Unusual Event Detection using Multiple Fixed-Location Monitors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 30, pp. 555-560, 2008.

[49] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, & A. Baskurt, "Sequential Deep Learning for Human Action Recognition," in *Human Behaviour Understanding*. vol. 7065, A. Salah and B. Lepri, Eds., ed: Springer Berlin Heidelberg, pp. 29-39, 2011.

[50] Unusual crowd activity dataset of University of Minnesota. [Online]. Available: http://mha.cs.umn.edu/movies/crowdactivity-all.avi.

[51] V. Mahadevan, L. Weixin, V. Bhalodia, & N. Vasconcelos, "Anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 1975-1981, 2010.

[52] S. Jing, K. Kai, L. Chen Change, & W. Xiaogang, "Deeply learned attributes for crowded scene understanding," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pp. 4657-4666, 2015.