Social Modeling Meets Virtual Reality: An Immersive Implication

Habib Ullah[1], Sultan Daud Khan[2], Mohib Ullah[3(✉)],
and Faouzi Alaya Cheikh[3]

[1] College of Computer Science and Engineering, University of Ha'il,
Hail, Saudi Arabia
[2] National University of Technology, Islamabad, Pakistan
[3] Department of Computer Science, Norwegian University of Science and
Technology, Gjøvik, Norway
`mohib.ullah@ntnu.no`

Abstract. The development of novel techniques for social modeling in the context of surveillance applications has significantly reduced manual processing of large and continuous video data. These techniques for social modeling widely cover crowd motion analysis since the impact of social modeling on crowd is significant. However, existing crowd motion analysis methods face a number of problems including limited availability of crowd data representing a specific behavior and weaknesses of proposed models to explore the underlying patterns of crowd behavior. To cope with these problems, we propose a novel method based on energy modeling and social interaction of individual particles in crowd to detect unusual behavior. Our method describes collective dissipative interactions among particles in a crowd scene. We reveal the changing patterns about the crowd behavior states, to support the conversion between different social behaviors during evolution. To further improve the performance of our method, virtual reality can be considered to consolidate the acquisition of data associated with a particular behavior. Therefore, we provide theoretical background of immersive implication considering virtual reality that can expose individuals to virtual crowds and acquire useful data on human motion and behaviors in crowds. The experimental evaluation of our energy and social interaction driven method shows convincing results.

Keywords: Virtual reality · Crowd analysis · Social modeling · Anomaly detection

1 Introduction

Video analysis of crowded scenes in different places such as public parks, universities, stadiums and stations has received significant interest in surveillance applications that aim to detect different unusual behavior [25,28]. An unusual

behavior occurs less frequently and is aberrant from those usual ones that frequently take place in crowd scenes. Manual analysis of surveillance videos to identify such behaviors is a highly time consuming process. In situations where a stampede or panic is reported, often many days of recordings need to be checked. It is, therefore, highly possible that due to manual error such abnormalities might cost a lot, and so, automatic systems are particularly beneficial in such situations. The description of unusual activities in complex crowd scenes like pedestrian walkways and stadiums is an arduous task, as an unusual behavior can be anything ranging from a panic situation to a stampede. In fact, it is difficult to develop an exhaustive list of all unusual behaviors. Therefore, a technique developed for any unusual behavior detection must be significantly driven by generalization capabilities to handle any such situation.

In the literature, different methods [4,24,26] are proposed for unusual crowd behavior detection. For instance, Kasudiya et al. [15] proposed the use of a wireless sensor network for monitoring the crowd and their behavior. Singh et al. [23] proposed aggregation of ensembles for detecting an anomaly in video data showing crowded scenes, which leverage the existing capability of pre-trained ConvNets and a pool of classifiers. Their method uses an ensemble of different fine-tuned convolutional neural networks based on the hypothesis that different network architectures learn different levels of semantic representation from crowd videos and thus an ensemble will enable enriched feature sets to be extracted. Behera et al. [3] investigated that the behavior of a dense crowd can be approximated using well-known physics-based models. They proposed a computer vision guided expert system with the help of a Langevin equation-based force model to represent the linear flow of the crowd, particularly in situations when the density is high. Hatirnaz et al. [12] introduced optical flow based features for abnormal crowd behaviour detection. They annotated surveillance videos using a new semantic metadata model based on multimedia standards using semantic web technologies. In this way, globally inter-operable metadata about abnormal crowd behaviours are generated. Then based on crowd behaviours a novel concept-based semantic search interface is proposed. Ullah et al. [27] used a modified variant of the social force model to highlight potential particles of interest for crowd behavior analysis. Hu et al. [13] proposed a parallel spatial-temporal convolution neural networks model to detect and localize the abnormal behavior in video surveillance. For crowd behavior analysis, Cheung et al. [5] proposed a method composed of two components: a procedural simulation framework to generate crowd movements and behaviors, and a procedural rendering framework to generate different videos or images.

Unusual crowd behavior detection is a challenging problem due to complex social relationship among individuals in the form of groups. The methods we discussed in the literature do not fully exploit social modeling to develop robust model for unusual crowd behavior detection due to the associated issues. For instance, interaction modeling is a problem when huge number of objects are present in a crowded scene. Also, within the same scene, the behavior of individuals may vary greatly, due to which, distinguishing between anomalies

and a normal event is a difficult task. To address these issues, we propose a novel method considering the energy and social interaction based modeling. Our method is characterized by unique social interaction representation since we take into account the characteristics of neighboring particles. Our method can be extended to any type of unusual crowd behavior detection. Additionally, we present theoretical background related to virtual reality to improve the performance of our proposed method. In fact, virtual reality is a powerful platform to understand the hidden patterns in crowd scenes.

The rest of the paper is organized as follows: Sect. 2 illustrates the proposed methodology and describes the working of each of the components. In Sect. 3, we present the theoretical background about the use of virtual reality to improve the performance of our proposed method. In Sect. 4, we describe the experiments performed for the validation of the proposed model and compare its performance with other reference methods. Section 5 concludes our proposed method.

2   Crowd Analysis

We present an energy model inspired by [18] to calculate the energy of the interacting particles. Our interaction model takes into account the collisions that occur in unusual crowd behavior. This model is based on the concept to describe collective dissipative interactions among particles in a crowd scene. Our modeling of particle interactions in the regime of crowd analysis redefined the entropy through information theory. This broadened the picture and allows the concept of entropy to permeate the area of crowd motion analysis.

We initialize a set of particles over the video frame. The particle initialization represents the localization of pixels positions uniformly spread over the video. We then track the initialized particles using the Lucas-Kanade optical flow technique. Since we are interested only in the particles associated with the moving entities instead of static regions, we propose an energy model in Eq. 1 for retaining the moving particles.

$$\eta = \frac{1}{|p-1|}(|1 - \sum_{i=1}^{N}(p)^{q_i}|) + \sum_{i=1}^{N}\frac{sin(q_i)}{2sin|q_i - p|} \tag{1}$$

In the equation, p represents the velocity of central particle and q_i represents the velocity of a neighboring particle in a localized region. In fact, our model moves away from the traditional way of handling interaction and energy dissipation related to unusual crowd behavior. Our model is well-defined for particle interactions and it gives physical meaning to the associated parameters. We assume that the crowd scene is associated with N particles. An individual particle moves with a velocity p. According to Eq. 1 we filter out static particles and retain particles associated with motion in crowd scene. Our model resembles with the model of kinetic theory of gases; the collisions of particles with each other are elastic. If we consider crowd as a system, then this crowd system with unusual behavior would be totally out of equilibrium system. The initial

state and all subsequently intermediate states of the crowd system are a non-equilibrium thermodynamic system. The particles interact with each other and we will simplify that interaction by assuming that they only do it by collisions since we are targeting unusual crowd behavior. In general, in this system the particles lose and gain energy just by the collisions. Considering the forced state to the equilibrium state, these particles will collide with each other multiple times in such a way that they can lose energy. Then, those particles dissipate energy to the surroundings due to the shocks, but we will assume that the interaction is made in a peculiar way; when they collide with different energies there is no loss of energy, this is when the difference exceeds a certain amount.

The benefits of interaction can drive the evolution of a crowd structure that represents different behaviors. The social interaction of individuals in a group is an emerging methodology to research the theory of crowd behavior. In light of this, we mimic group theory, and then develop our method to solve the complex problems associated with unusual crowd behaviors. Group theory is based on the studies of collective phenomena such as flocks of birds, colonies of ants, and swarms of bees to model biological swarms. These models have been applied as nature-inspired algorithms for solving complex problems in the real world. In fact, the influence of biological analogies is attested by subfields of computer science, such as artificial neural networks, genetic algorithms, and evolutionary computation. For example, the particle swarm optimization (PSO) model [16] is inspired by the social behavior of bird flocking, the ant colony model (ACO) [8] is based on the division and cooperation of ants foraging, the artificial bee colony (ABC) model [14] is constructed by mimicking the cooperative behavior of bee colonies, and the social spider optimization (SSO) model [7] is based on the simulation of cooperative behavior of social-spiders. Each of these algorithms has its own characteristics. Their modeling processes and effective mechanisms inspire us when developing our own method for crowd behavior detection.

We consider the retained particles according to Eq. 1 to find out particles representing unusual crowd behavior. For this purpose, our proposed method is based on social behavior modeling [10] to detect unusual crowd behavior. We model the mechanism between Leaders and Followers particles to infer local dynamics. In this way, our method reveals the changing patterns about the crowd behavior states, to support the conversion between different social behaviors during evolution, to demonstrate crowd behaviors from social grouping perspective, and to avoid the integration of one crowd behavior with another. In fact, our model analyzes the behavior of social groups, the changing patterns of different behaviors, and to set the changing patterns as behavior's criterion of a crowd state. In addition, the mathematical model of our method is deduced from the group theory, crowd dynamics, and the crowd motion pattern theory.

According to the social model [10], the position and velocity of a central particle p_i are $xp_i(t)$ and $vp_i(t)$, respectively. The position and velocity of a surrounding particle q_j is $xq_j(t)$ and $vq_j(t)$. In the $(t+1)th$ iteration, the position and velocity of particle i will be updated as,

$$vp_i(t + 1) = w \times vp_i(t) + (\frac{1}{N}\sum_{j=1}^{N}vq_j(t) - vp_i(t)) \tag{2}$$

$$xp_i(t + 1) = xp_i(t) + (vp_i(t + 1) \times \delta(t)) \tag{3}$$

In the equation, w is a parameter which is set equal to a fixed value during the experiments. The above equations do not take into account the orientation information of particles. The particle status process belongs to the exploration behavior of the central and surrounding particles. This process reveals the change rule of social population in term of localized crowd, and supports the conversion between different social behaviors during evolution. Therefore, such equations should be adjusted to include velocity, position, and angle. To consider all the information into a unified equation $(M_{p,q})$, the modeling is performed as presented in Eq. 4,

$$M_{p,q} = \prod_{j=1}^{N} \frac{\int\limits_{p,q_j}^{p,q_j} 2wvp_i + (\varphi + sin\theta)d\theta}{4d^2} \tag{4}$$

where N is the total number of surrounding particles q_j and vp_i is the velocity of the central particle p_i. φ is the prior average orientation information of all the surrounding particles q_j. θ is the angle between the central particle p_i and a surrounding particle q_j. d is Eucleadien distance between particle p_i and q_j. Equation 4 classifies each particle p_i as belonging to unusual crowd behavior or normal crowd behavior. This equation renders the classification characteristics of each particle. Therefore, it could be treated as features to identify and detect crowd behavior as a whole.

3   Crowd and Virtual Reality

Mathematical modeling of social behaviors of crowd present significant success in the field. However, some limitations still exist to fully characterize crowd representing different behaviors. These limitations include limited availability of crowd data representing a specific behavior, weaknesses of proposed models to explore the underlying patterns of crowd, the perception and understanding of an unusual crowd behavior from first person point of view to ensure people safety, and the impact of varying densities of crowd on individuals. To improve the performance of proposed models and to cope with the key challenges, virtual reality platforms can be exploited since virtual reality has been successfully used for wide range of applications including medical education [2], rehabilitation [11], and data visualization [9,17] to name a few. In fact virtual reality is a powerful platform to acquire useful data on human motion and behaviors in crowds. It allows exposing individuals to virtual crowds: only one individual is required to observe holistic behavior in crowded situations. In the virtual reality, stimuli can be properly controlled and repeated over several individuals. These characteristics have made virtual reality a significant tool to perform experiments

in socio-psychology, spatial-cognition, and motion control. In fact, this platform is an effective tool to study how we navigate in crowds. To effectively use virtual reality platform, we have to ensure the validity of obtained data in various crowd situations, perform proper modeling so that locomotion trajectories in virtual environment are similar to reality trajectories, minimize the affect of interaction loop in virtual reality on user behavior, and ensure that the visual feedback in virtual reality enables participants to make realistic navigation decisions.

To address the aforementioned limitations considering virtual reality, two different approaches can be considered to deal with unusual crowd behaviors. The macroscopic approach considers the crowd as a single entity and the microscopic approach considers that global crowd pattern emerge from local interactions between individuals. The use of virtual reality in analyzing crowd motion have a wide range of applications, from training personnel to ensure people safety in crowd, to architecture in building analysis and emergency evacuation studies. The significant impact of virtual reality can be used to extract more and more realistic crowd patterns to enlarge the available patterns associated with a particular crowd behavior. In this context, we define realism as a match between extracted crowd patterns from virtual environment and real patterns extracted from real crowd data. To provide realism to crowd motion analysis for unusual behavior identification, there is a need to understand and model how humans move and behave during local interactions with their neighborhood in virtual conditions. To extract patterns from virtual environment, we have to consider the underlying modeling of human motion and interactions in various situations (several kinds of motion or interactions) and take into account multiple factors such as sociological or psychological ones.

Data can be extracted from virtual environment if we consider uncertainties on people states and motivations and other uncontrolled factors. The virtual environment [21] is based on local interactions between individuals. Many interactions occur between walkers, with many factors of influence. There is a need for observations of individuals facing interactions in crowds to better understand them and improve the level of realism in virtual environment. Considering the presented information about virtual reality in the context of crowd behavior analysis, virtual reality can be used as an experimental tool to perform such observations with an accurate control of experimental conditions.

4   Experiemental Evaluation

We assess the performance of our social interaction based method for unusual crowd behavior detection on a standard dataset available publicly that is UMN [1]. The UMN dataset (as shown in Fig. 1) from the university of Minnesota consists of videos representing unusual crowd behavior. There are three different indoor and outdoor scenes showing 11 different scenarios of unusual crowd behavior. In total, there are 7739 frames and the resolution of each frame is 320×240 pixels. The initial part of each video consists of normal behaviors of pedestrian walking and standing.

We compare the results with four closely related reference methods: spatio-temporal anomaly model (STAM) [20], abnormal behavior model (ABM) [6] and crowd influence model (CIM) [22]. For quantitative evaluation of unusual crowd behavior detection, the equal error rate (EER) for frame-level and the detection rate (DR) for pixel-level analysis are calculated to measure the overall performance. In the literature, the frame-level criterion is mostly used by researchers. The frame-level criterion only measures temporal localization accuracy. It could cause errors due to lucky detection of unusual crowd behavior. Therefore, it assigns a perfect score to a model that detects unusual behavior at a random location of a frame. By considering this fact, it seems that the pixel-level criterion is much better evaluation criterion. Therefore, we use both the temporal and spatial accuracies to rule out random detection. Both criteria are based on true-positive rates (TPR) and false positive rates (FPR). The presence and absence of unusual behaviors are represented by a positive and a negative, respectively. This is compared to the frame level ground-truth, to determine the number of true and false-positive frames. Similarly, pixels related to the unusual crowd behavior are compared to the pixel-level ground-truth to determine the number of true-positive and false-positive.

Fig. 1. UMN dataset. Four different scenes are shown representing unusual crowd behavior where people are running in different directions randomly.

Table 1. UMN dataset. Equal error rate (EER) and detection rate (DR) for the reference methods and our proposed method are presented in the first row and the second row, respectively. Lower EER and higher DR represent better performance.

Dataset	STAM	ABM	CIM	Prop.
UMN	21	34	19	17
	55	61	68	71

For quantitative performance, we calculated the average EER and average DR for UMN dataset reported in Table. For better performance, the EER rate should be lower and the DR rate should be higher. As can be seen in the table, our proposed method outperforms all the reference methods: STAM [20], ABM [6], and CIM [22]. These results show that there is a significant advantage of our proposed method that uses energy and interaction based modeling bringing forth strong generalization capabilities. In fact, the reference methods based on shallow features cannot cope with the adaptively changing sparse movement of the people flows where dynamic motion and occlusions exist. Also the reference methods finds raw features without taking into account the appearance information. Furthermore, STAM [20], ABM [6], and CIM [22] fail to encode unique motion patterns because informative movements only occur in specific regions of the videos. Our proposed method represents high quality description of unusual crowd behavior with the energy and social interaction components. Therefore, we outperform the reference methods in both frame-level and pixel-level analysis. Presenting results based on both criteria reveal the effectiveness of our method.

To use virtual reality for crowd behavior analysis, Unity [19] cross-platform engine can be considered as a powerful tool for crowd behavior research. It can be explored for the ability to create three-dimensional visual scenes and to measure responses to the visual stimuli for testing the hypotheses in a manner and scale that were previously unfeasible (Table 1).

5   Conclusion

We proposed a novel method for unusual crowd behavior detection. Our proposed method represents high quality description of unusual behavior in term of energy and interaction modeling. We also provided the background of using virtual reality platform to improve the performance by extracting of crowd patters from virtual environment. The performance of our proposed method is tested on a standard dataset and compared to three closely related reference methods. The performance metrics EER and DR show that our method outperforms all the reference methods in both frame-level and pixel-level analysis.

As a future work, we would also extend our proposed method to take into account the benefits of virtual reality platforms.

References

1. Unusual crowd activity dataset of university of minnesota. http://mha.cs.umn.edu/movies/crowd-activity-all.avi
2. Ammanuel, S., Brown, I., Uribe, J., Rehani, B.: Creating 3D models from radiologic images for virtual reality medical education modules. J. Med. Syst. **43**(6), 166 (2019)
3. Behera, S., Dogra, D.P., Bandyopadhyay, M.K., Roy, P.P.: Estimation of linear motion in dense crowd videos using langevin model. Expert Systems with Applications, p. 113333 (2020)
4. Cao, H., Sankaranarayanan, J., Feng, J., Li, Y., Samet, H.: Understanding metropolitan crowd mobility via mobile cellular accessing data. ACM Trans. Spatial Algorithms Syst. (TSAS) **5**(2), 1–18 (2019)
5. Cheung, E., Wong, A., Bera, A., Wang, X., Manocha, D.: LCrowdV: generating labeled videos for pedestrian detectors training and crowd behavior learning. Neurocomputing **337**, 1–14 (2019)
6. Chibloun, A., El Fkihi, S., Mliki, H., Hammami, M., Thami, R.O.H.: Abnormal crowd behavior detection using speed and direction models. In: 2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC), pp. 197–202. IEEE (2018)
7. Cuevas, E., Cienfuegos, M., ZaldíVar, D., Pérez-Cisneros, M.: A swarm optimization algorithm inspired in the behavior of the social-spider. Expert Syst. Appl. **40**(16), 6374–6384 (2013)
8. Dorigo, M., Maniezzo, V., Colorni, A.: Ant system: optimization by a colony of cooperating agents. IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) **26**(1), 29–41 (1996)
9. El Beheiry, M., Doutreligne, S., Caporal, C., Ostertag, C., Dahan, M., Masson, J.B.: Virtual reality: beyond visualization. J. Mol. Biol. **431**(7), 1315–1321 (2019)
10. Feng, X., Wang, Y., Yu, H., Luo, F.: A novel intelligence algorithm based on the social group optimization behaviors. IEEE Trans. Syst. Man Cybern. Syst. **48**(1), 65–76 (2016)
11. Hartney, J.H., Rosenthal, S.N., Kirkpatrick, A.M., Skinner, J.M., Hughes, J., Orlosky, J.: Revisiting virtual reality for practical use in therapy: patient satisfaction in outpatient rehabilitation. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 960–961. IEEE (2019)
12. Hatirnaz, E., Sah, M., Direkoglu, C.: A novel framework and concept-based semantic search interface for abnormal crowd behaviour analysis in surveillance videos. Multimedia Tools and Applications, pp. 1–39 (2020)
13. Hu, Z.P., Zhang, L., Li, S.F., Sun, D.G.: Parallel spatial-temporal convolutional neural networks for anomaly detection and location in crowded scenes. J. Visual Commun. Image Representation **67**, 102765 (2020)
14. Karaboga, D.: An idea based on honey bee swarm for numerical optimization. Technical report, Technical report-tr06, Erciyes university, engineering faculty, computer (2005)
15. Kasudiya, J., Bhavsar, A., Arolkar, H.: Wireless sensor network: a possible solution for crowd management. In: Somani, A.K., Shekhawat, R.S., Mundra, A., Srivastava, S., Verma, V.K. (eds.) Smart Systems and IoT: Innovations in Computing. SIST, vol. 141, pp. 23–31. Springer, Singapore (2020). https://doi.org/10.1007/978-981-13-8406-6_3

16. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN'95-International Conference on Neural Networks, vol. 4, pp. 1942–1948. IEEE (1995)

17. Kvinge, H., Kirby, M., Peterson, C., Eitel, C., Clapp, T.: A walk through spectral bands: using virtual reality to better visualize hyperspectral data. In: Vellido, A., Gibert, K., Angulo, C., Martín Guerrero, J.D. (eds.) WSOM 2019. AISC, vol. 976, pp. 160–165. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-19642-4_16

18. López-Carrera, B., Yáñez-Márquez, C.: A simple model for the entropy of a system with interacting particles. IEEE Access **7**, 108969–108979 (2019)

19. Messaoudi, F., Simon, G., Ksentini, A.: Dissecting games engines: the case of unity3d. In: 2015 International Workshop on Network and Systems Support for Games (NetGames), pp. 1–6. IEEE (2015)

20. Ojha, N., Vaish, A.: Spatio-temporal anomaly detection in crowd movement using sift. In: 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 646–654. IEEE (2018)

21. Olivier, A.H., Bruneau, J., Cirio, G., Pettré, J.: A virtual reality platform to study crowd behaviors. Transp. Res. Procedia **2**, 114–122 (2014)

22. Pan, L., Zhou, H., Liu, Y., Wang, M.: Global event influence model: integrating crowd motion and social psychology for global anomaly detection in dense crowds. J. Electron. Imaging **28**(2), 023033 (2019)

23. Singh, K., Rajora, S., Vishwakarma, D.K., Tripathi, G., Kumar, S., Walia, G.S.: Crowd anomaly detection using aggregation of ensembles of fine-tuned convnets. Neurocomputing **371**, 188–198 (2020)

24. Ullah, H.: Crowd Motion Analysis: Segmentation, Anomaly Detection, and Behavior Classification. Ph.D. thesis, University of Trento (2015)

25. Ullah, H., Altamimi, A.B., Uzair, M., Ullah, M.: Anomalous entities detection and localization in pedestrian flows. Neurocomputing **290**, 74–86 (2018)

26. Ullah, H., Ullah, M., Conci, N.: Dominant motion analysis in regular and irregular crowd scenes. In: Park, H.S., Salah, A.A., Lee, Y.J., Morency, L.-P., Sheikh, Y., Cucchiara, R. (eds.) HBU 2014. LNCS, vol. 8749, pp. 62–72. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11839-0_6

27. Ullah, M., Ullah, H., Conci, N., De Natale, F.G.: Crowd behavior identification. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 1195–1199. IEEE (2016)

28. Zitouni, M.S., Sluzek, A., Bhaskar, H.: Visual analysis of socio-cognitive crowd behaviors for surveillance: a survey and categorization of trends and methods. Eng. Appl. Artif. Intell. **82**, 294–312 (2019)