



HAL
open science

Characterization of Orderly Behavior of Human Crowd in Videos Using Deep Learning

Shreetam Behera, Shaily Preetham Kurra, Debi Prosad Dogra

► **To cite this version:**

Shreetam Behera, Shaily Preetham Kurra, Debi Prosad Dogra. Characterization of Orderly Behavior of Human Crowd in Videos Using Deep Learning. 4th International Conference on Intelligence Science (ICIS), Feb 2021, Durgapur, India. pp.217-226, 10.1007/978-3-030-74826-5_19 . hal-03741725

HAL Id: hal-03741725

<https://inria.hal.science/hal-03741725>

Submitted on 1 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Characterization of Orderly Behavior of Human Crowd in Videos Using Deep Learning

Shreetam Behera, Shaily Preetham Kurra, and Debi Prosad Dogra

School of Electrical Sciences, IIT Bhubaneswar, Odisha, PIN 752050, India
{sb46,ksp10,dpdogra}@iitbbs.ac.in

Abstract. In the last few decades, understanding crowd behavior in videos has attracted researchers from various domains. Understanding human crowd motion can help to develop monitoring and management strategies to avoid anomalies such as stampedes or accidents. Human crowd movements can be classified as structured or unstructured. In this work, we have proposed a method using deep learning technique to characterize crowd behavior in terms of order parameter. The proposed method computes features comprised of motion histogram, entropy, and order parameter of the frames of a given crowd video. The features are fed to a Long Short Term Memory (LSTM) model for characterization. We have tested the proposed method on a dataset comprising of structured and unstructured crowd videos collected from publicly available datasets and our recorded video datasets. Accuracy as high as 91% has been recorded and the method has been compared with some of the recent machine learning algorithms. The proposed method can be used for real-time applications focusing on crowd monitoring and management.

Keywords: Crowd characterization · LSTM · supervised learning · crowd anomalies · motion histogram · crowd behaviour · crowd monitoring

1 Introduction

Crowd motion provides important cues in understanding crowd behavior. Understanding the dynamics of the crowd motion can be used to understand different crowd anomalies such as stampede. Computer-vision algorithms have been popular among the researchers as it handles high volumes of data, reduces human intervention, and are often accurate and less error-prone. These algorithms can be used for crowd flow segmentation, behavioral analysis, and anomaly detection. Existing crowd analysis algorithms usually employ a combination of physics-based approaches and computer-vision approaches resulting in effective analysis of the crowd as mentioned in [10, 12, 30]. A Lagrangian particle dynamics-based framework has been proposed in [1] to segment crowd flows and to find flow instabilities in the crowd. However, this technique is highly complex and computationally intensive. The approach mentioned in [11] can detect similar regions in the crowd. A force-field based approach has been used to understand the flow regions in a crowd in [2]. Social force models mentioned in [5, 9, 18] are also popular in defining crowd in terms of flow, interactions, and abnormalities. Wu et

al. in [28] have described the crowd behavior in terms of bi-linear interactions of crowd flows by analyzing the curl and divergence of these flows. The crowd has been modeled as dynamical systems for behavioral analysis in [13, 15, 26, 27] for behavioral analysis. Zhou et al. in [32] have analyzed crowd collective motion segments and analyzed these segments' collective behavior. In [4, 25, 33], Convolutional Neural Networks (CNN) have been used to perform crowd behavior analysis. The authors in [29] have developed a crowd anomaly algorithm based upon sparse representation-based scheme. There are a few similar approaches proposed in the areas of crowd monitoring systems and detecting abnormal behaviors [14, 17, 19, 22]. Hao et al. in [7] have proposed a crowd anomaly detection scheme using spatio-temporal texture analysis. The texture analysis was performed based upon the grey-level co-occurrence matrix model. Ryan et al. in [20] have proposed an algorithm for detecting anomalous behaviors in crowds by analyzing the optical flow map as a texture. In [6], the authors have proposed a nearest-neighbor search approach for detecting anomalous behavior in crowd videos. Sabokrou in [21] has proposed a real-time anomaly detection method that performs scene localization in large crowds. The authors in [31] have proposed a model based on the social-force and entropy to analyze macrostates in the crowd. Their analysis shows that a disordered crowd motion has higher entropy as compared to orderly crowd motion. However, a few parameters are needed for a complete and accurate analysis of crowd behavior. In [3], fluctuations in energy and entropy curves have been used to characterize crowd in terms of speed and randomness. However, none of those mentioned above papers address the crowd characterization task in terms of ordered behavior, and none of them use a supervised learning algorithm for crowd characterization. This has motivated us to develop a method that can characterize crowd videos based on orderly behavior of the particles using supervised learning.

The rest of the paper is organized as follows. The proposed methodology is presented in Section 2. The results obtained from the proposed method is discussed in Section 3. The work has been concluded along with the future scope in Section 4.

2 Proposed Methodology

The human crowd can be characterized based on physical quantities such as energy, randomness, and orderliness. These physical quantities defined the typical behavior of the crowd. For example, in a marathon, as presented in Figure 1a, the crowd can be seen moving in a single direction, which can be considered as structured crowd behavior with less random movements, characterized by lower entropy and higher order parameter. In another example, people moving inside a shopping mall as presented in Figure 1e can be considered as highly random and is less structured in nature. This type of movement is considered as unstructured movement, and such movements are characterized by higher entropy and lower order parameter. These behaviors are illustrated in Figures 2 and 3, respectively. Identifying these types of movements will certainly help to understand the cause

of human movements leading to normal and abnormal activities. Now, the crowd characterization task can be mapped to a typical classification problem.



Fig. 1: (a-c) Samples of structured crowd, and (d-f) unstructured crowd videos. (Best viewed in color)

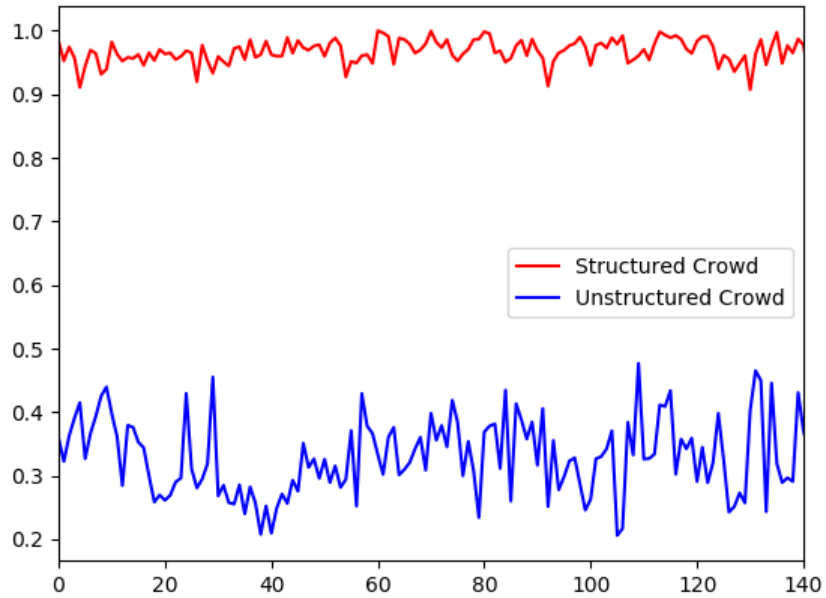


Fig. 2: Order parameter for structured and unstructured crowd, where y-axis represents ϕ or the order parameter and x-axis represents time of frame numbers. (Best viewed in color)

2.1 Feature Extraction

Feature extraction is performed for each video segment consisting of 20 frames. The feature vector consists of motion histogram, entropy, and order parameter.

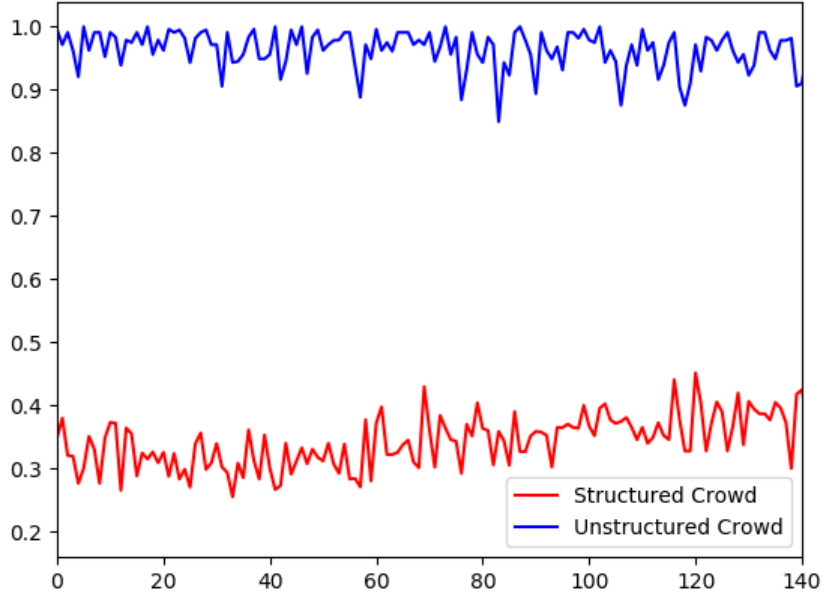


Fig. 3: Entropy for structured and unstructured crowd, where y-axis represents entropy and x-axis represents time of frame numbers. (Best viewed in color)

For a sequence of 20 frames of a video, keypoints have been extracted using Shi-Thomas corner detection [24]. Next, these keypoints (N) are fed to the optical flow process, as mentioned in [16] for the flow computation. The magnitudes and orientation of the keypoints (with v_x, v_y velocities across x and y axis, respectively) are calculated using (1) and (2), respectively.

$$m = \sqrt{(v_x)^2 + (v_y)^2} \quad (1)$$

$$\theta = \arctan((v_y)/(v_x)) \quad (2)$$

The obtained magnitude values are normalized within a range of $[0, 1]$. A 2D histogram is then constructed using the normalized magnitude and orientation values with m and n number of bins. Thus, $1/m$ and $360/n$ ranges are obtained for each keypoint. For example, if $m = 4$ and $n = 4$, the number of bins formed are 4 with the ranges of $[0, 0.25)$, $[0.25, 0.5)$, $[0.5, 0.75)$, and $[0.75, 1]$ and the orientation fall in the ranges of $[0, 90)$, $[90, 180)$, $[180, 270)$, and $[270, 360]$. Thus, we obtain a 4×4 matrix with 16 bins of the above-specified magnitude and orientation ranges. The count for each bin is increased when the number of keypoints with magnitude and orientation has been found to be within the aforementioned range. This is performed for all keypoints leading to a 2D histogram. Then the

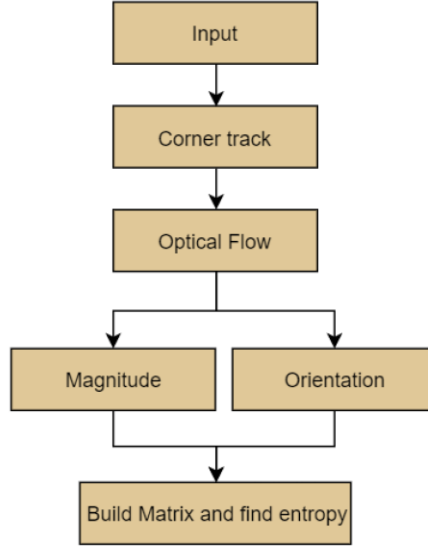


Fig. 4: Block diagram of the feature extraction technique.

entropy and the order parameter are calculated using (3) and (4), respectively,

$$S = \sum_{bin=1}^{m*n} -P_{bin} \ln(P_{bin}) \quad (3)$$

where m and n are the size of the 2D Histogram such that $v_k = v_x \hat{i} + v_y \hat{j}$ is the velocity vector for each keypoint k . The final feature vector is represented as $[Hist_{m*n}, S, \phi]$. These features are then fed to the classifier for characterization.

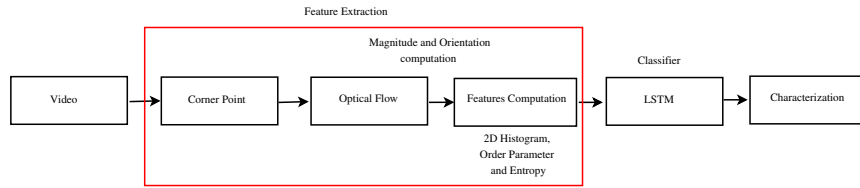


Fig. 5: Block diagram of the proposed characterization scheme.

$$\phi = \frac{1}{N} \sum_{k=1}^N |v_k(t)| \quad (4)$$

2.2 Crowd Classification

The crowd characterization is considered as a binary classification problem, where a crowd’s behavior is classified as structured or unstructured. In this work, Long Short Term Memory (LSTM)-based classifier [8] has been used for characterization as LSTM can deal with temporal data, and the features extracted for each sequence in this work are temporally related. The architecture of the LSTM characterization model is presented in Table 1. The proposed model consists of 2 layers. The first layer consists of 32 memory units of LSTM. The equations involved with a single memory unit are presented in (5)-(10).

$$ig_t = \sigma(W_{ig}x_t + Ub_{ig}hg_{t-1} + Vb_i cg_{t-1}) \quad (5)$$

$$fg_t = \sigma(W_{fg}x_t + Ub_{fg}hg_{t-1} + Vb_f cg_{t-1}) \quad (6)$$

$$og_t = \sigma(W_{og}x_t + Ub_{og}hs_{t-1} + Vb_o cg_{t-1}) \quad (7)$$

$$\tilde{c}g_t = \tanh(W_{cg}x_t + Ub_{cg}hs_{t-1}) \quad (8)$$

$$cg_t = fg_t^i \odot cg_{t-1} + ig_t \odot \tilde{c}g_t \quad (9)$$

$$hs_t = og_t \odot \tanh(cg_t) \quad (10)$$

The second layer is a fully connected neural layer with a sigmoid activation function. Since, characterization is a binary classification problem, binary cross-entropy presented in (11) is considered as objective function for loss calculation.

$$L_{\text{entropy}} = - \sum_{i=1}^N [yTarget_i \ln(yPredict_i) + (1 - yTarget_i) \ln(1 - yPredict_i)] \quad (11)$$

The model is trained with Adam’s optimizer for a batch size of 16 and 100 epochs. In the above formulations, at each time step t , the output of the LSTM is controlled by input gate ig_t , a forget gate fg_t , an output gate og_t , a memory cell cg_t , and a hidden state hs_t . x_t is the input at the current time step, σ denotes the logistic sigmoid function, \tanh , and \odot stands for hyperbolic tangent function and element-wise multiplication. W , Ub , and Vb represent weight and bias matrices that are updated during the training stage, where N is the number of output scalar values in the model, $yTarget_i$ is the expected output, and $yPredict_i$ is the predicted output of the i^{th} scalar value of the model.

Table 1: Architecture of the proposed LSTM-based model for characterization

Layers	Output Shape	Number of parameters
LSTM	32	4352
Fully Connected Layer with sigmoid activation	1	33

3 Results

This section discusses about the results obtained from the proposed LSTM classification scheme.

3.1 Video Dataset

In this work, 251 videos have been selected from a publicly available video dataset [23] and our videos recorded during Puri Rath Yatra that happens every year in Odisha, India. There are 106 structured crowd videos consisting of linearly and non-linearly ordered motion flows, and 145 unstructured videos in this selection consist of crowd mixing and random pedestrian movements. The structured and unstructured videos have been labeled. The training and testing sets comprise of 80% and 20% videos of the dataset.

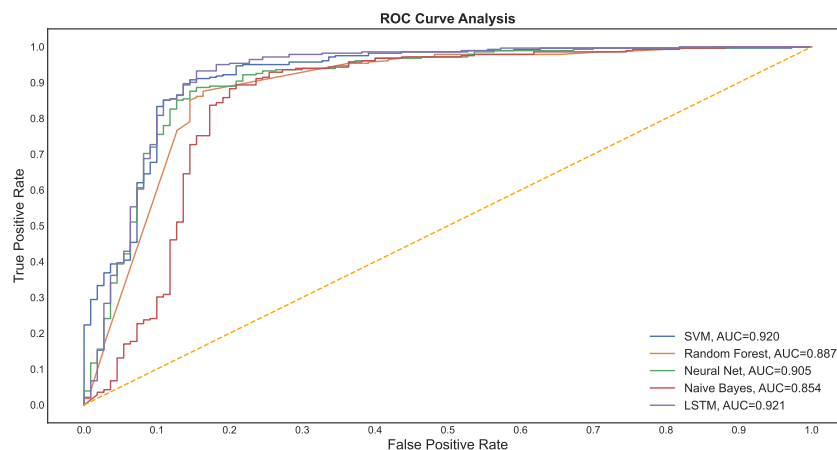


Fig. 6: ROC curves (receiver operating characteristic) for the classifiers used for crowd characterization. (Best viewed in color)

3.2 Classification Results

The dataset is fed to the LSTM-based classifier that has been trained with 100 epochs and the outputs are evaluated in terms of the area under ROC curve (AUC). The proposed characterization scheme has also been compared with the existing popular machine learning algorithms like Support Vector Machine (SVM), Random Forest (RF), Multilayer Perceptron (MLP), Naive Bayes. The proposed algorithm outperforms all the above mentioned classification algorithms.

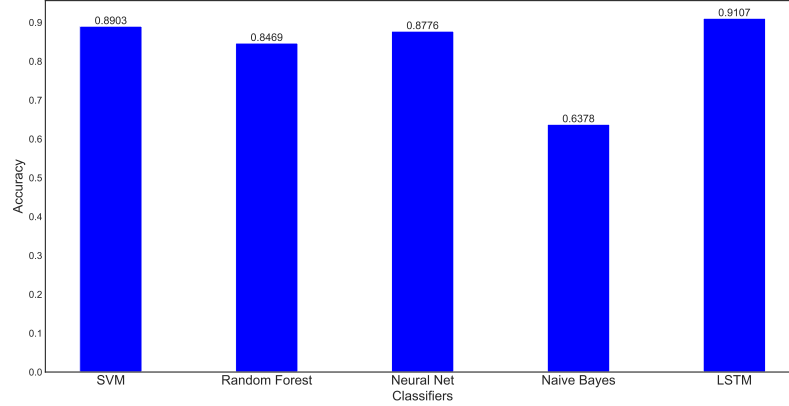


Fig. 7: Accuracy bar graph for the classifiers used for crowd characterization.

Our proposed classification achieves accuracy and AUC values as high as 91% and 0.9218, respectively. The AUC and accuracy comparisons are illustrated in Figures 6 and 7, respectively. The table depicting the values of each classifier is illustrated in Table 2.

Table 2: Accuracy and area under ROC values for the classifiers used for crowd characterization.

Classifier	Accuracy	Area Under ROC
SVM	0.8903	0.920
Random Forest	0.8469	0.887
Neural Net	0.8776	0.905
Naive Bayes	0.6378	0.854
LSTM	0.9107	0.921

4 Conclusion and Future Scopes

Crowd characterization is a classification problem of identifying the crowd movements based on ordered behavior at a global level. The proposed method can successfully characterize the video sequences as structured and unstructured compared to the existing machine learning-based algorithms. The method can reduce the human load on crowd behavior analysis, segregating typical behaviors in the crowd more distinctively. Furthermore, the proposed method is robust and straightforward to implement. Thus, real-time applications are possible. In

future, the method can be improved by integrating motion dynamics-based features, and deep learning-based features can be used for describing the crowd behavior. Furthermore, the proposed algorithm can further be extended to detect and localize the structured and unstructured behavior at local levels.

References

1. Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–6 (2007)
2. Ali, S., Shah, M.: Floor fields for tracking in high density crowd scenes. In: European Conference on Computer Vision. pp. 1–14. Springer (2008)
3. Behera, S., Dogra, D.P., Roy, P.P.: Characterization of dense crowd using gibbs entropy. In: Proceedings of 2nd International Conference on Computer Vision & Image Processing. pp. 289–300. Springer (2018)
4. Cao, L., Zhang, X., Ren, W., Huang, K.: Large scale crowd analysis based on convolutional neural network. *Pattern Recognition* **48**(10), 3016–3024 (2015)
5. Chaker, R., Al Aghbari, Z., Junejo, I.N.: Social network model for crowd anomaly detection and localization. *Pattern Recognition* **61**, 266–281 (2017)
6. Colque, R.V.H.M., Caetano, C., de Andrade, M.T.L., Schwartz, W.R.: Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos. *IEEE Transactions on Circuits and Systems for Video Technology* **27**(3), 673–682 (2016)
7. Hao, Y., Xu, Z.J., Liu, Y., Wang, J., Fan, J.L.: Effective crowd anomaly detection through spatio-temporal texture analysis. *International Journal of Automation and Computing* **16**(1), 27–39 (2019)
8. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural computation* **9**(8), 1735–1780 (1997)
9. Ji, Q.G., Chi, R., Lu, Z.M.: Anomaly detection and localisation in the crowd scenes using a block-based social force model. *IET Image Processing* **12**(1), 133–137 (2017)
10. Junior, J.C.S.J., Musse, S.R., Jung, C.R.: Crowd analysis using computer vision techniques. *IEEE Signal Processing Magazine* **27**(5), 66–77 (2010)
11. Khan, S.D., Bandini, S., Basalamah, S., Vizzari, G.: Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows. *Neurocomputing* **177**, 543–563 (2016)
12. Kok, V.J., Lim, M.K., Chan, C.S.: Crowd behavior analysis: A review where physics meets biology. *Neurocomputing* **177**, 342–362 (2016)
13. Kountouriotis, V., Thomopoulos, S.C., Papelis, Y.: An agent-based crowd behaviour model for real time crowd behaviour simulation. *Pattern Recognition Letters* **44**, 30–38 (2014)
14. Kratz, L., Nishino, K.: Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1446–1453. IEEE (2009)
15. Lim, M.K., Chan, C.S., Monekosso, D., Remagnino, P.: Detection of salient regions in crowded scenes. *Electronics Letters* **50**(5), 363–365 (2014)
16. Lucas, B.D., Kanade, T., et al.: An iterative image registration technique with an application to stereo vision (1981)

17. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 1975–1981. IEEE (2010)
18. Mehran, R., Moore, B.E., Shah, M.: A streakline representation of flow in crowded scenes. In: European Conference on Computer Vision. pp. 439–452. Springer (2010)
19. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. pp. 935–942. IEEE (2009)
20. Ryan, D., Denman, S., Fookes, C., Sridharan, S.: Textures of optical flow for real-time anomaly detection in crowds. In: Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on. pp. 230–235. IEEE (2011)
21. Sabokrou, M., Fathy, M., Hoseini, M., Klette, R.: Real-time anomaly detection and localization in crowded scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 56–62 (2015)
22. Saligrama, V., Chen, Z.: Video anomaly detection based on local statistical aggregates. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2112–2119. IEEE (2012)
23. Shao, J., Change Loy, C., Wang, X.: Scene-independent group profiling in crowd. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2219–2226 (2014)
24. Shi, J., et al.: Tomasi. good features to track. In: Computer Vision and Pattern Recognition. pp. 593–600 (1994)
25. 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)
26. Solmaz, B., Moore, B.E., Shah, M.: Identifying behaviors in crowd scenes using stability analysis for dynamical systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(10), 2064–2070 (2012)
27. Wang, X., Yang, X., He, X., Teng, Q., Gao, M.: A high accuracy flow segmentation method in crowded scenes based on streakline. *Optik-International Journal for Light and Electron Optics* **125**(3), 924–929 (2014)
28. Wu, S., Su, H., Yang, H., Zheng, S., Fan, Y., Zhou, Q.: Bilinear dynamics for crowd video analysis. *Journal of Visual Communication and Image Representation* **48**, 461–470 (2017)
29. Yuan, Y., Wan, J., Wang, Q.: Congested scene classification via efficient unsupervised feature learning and density estimation. *Pattern Recognition* **56**, 159–169 (2016)
30. Zhang, X., Yu, Q., Yu, H.: Physics inspired methods for crowd video surveillance and analysis: a survey. *IEEE Access* (2018)
31. Zhao, Y., Yuan, M., Su, G., Chen, T.: Crowd macro state detection using entropy model. *Physica A: Statistical Mechanics and its Applications* **431**, 84–93 (2015)
32. Zhou, B., Tang, X., Zhang, H., Wang, X.: Measuring crowd collectiveness. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(8), 1586–1599 (Aug 2014). <https://doi.org/10.1109/TPAMI.2014.2300484>
33. Zhou, S., Shen, W., Zeng, D., Fang, M., Wei, Y., Zhang, Z.: Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Processing: Image Communication* **47**, 358–368 (2016)