Estimation of a high-dense crowd based on a Balanced Communication-Avoiding Support Vector Machine classifier

Shaya A. Alshaya

Computer Science Department, College of sciences and humanities sciences at alGhat, Majmaah University, Majmaah, 11952, Saudi Arabia

Summary

The crowd estimation presents the main problem of many applications as a pedestrian crowd, traffic flow. The crowd estimation based on computer vision is faced to scale variation, occlusions, and non-uniform distribution challenges. This defy could be seen as a physical problem by considering it as a particle dynamics-model or as a computer-vision model. Seen that importance an accurate estimation provides in-depth the behavior of the crowd. We propose a smart crowd estimation system based on video processing. The proposed uses the k-means clustering and the Balanced Communication-Avoiding Support Vector Machine classifier to estimate the number of pedestrians in crowds. The achieved results prove the accuracy of the system in comparison with previous works. The error is less than one for a low-dense crowd, around two eighty-eight of persons for a medium-dense crowd, and around eighty-nine persons for a high-dense crowd. Findings not only show a low error of the density but also a minimum time of execution (13 ms).

Key words:

Crowd density, count density, Balanced Communication-Avoiding Support Vector Machine, crowd behavior

1. Introduction

Crowd monitoring increasingly enhanced by the researchers due to the recent improvement related to visual technology and the learning machine algorithms. Automated decision-making systems based on computer vision could manage the crowd better than other semi-automated methods. Therefore, the automated system not only decreased the intervention of humans but also computed the crowd with the lowest errors. Expert systems provide a learning method that could detect and analyze unusual behavior. These analyses propose scenarios to avoid dangerous incidents. These objectives are faced with the weakness of existing frameworks that are not supported by the crowd with high density [1].

Crowd estimation system presents a challenging task because the crowd is characterized by a set of groups that moves through similar orientations. This problem has overcome by two strategies: (1) particle dynamics-methods, and (2) computer vision methods [2].

In this paper, a deep comparison between particle dynamics methods and computer vision-model are drawn. Then, we reveal a Balanced Communication-Avoiding Support Vector Machine (BCA-SVM) algorithm as an artificial intelligence technique to estimate the density of a high crowd scene.

2. Related Works

This section constitutes a primordial phase because the crowded system has many research study developed in the literature. Related work to the estimation of the density of a crowd could be classified into two groups according to the applied method: (1) particle dynamics-methods, and (2) computer vision methods. Existing frameworks have focused on the detection of the flow and the segmentation [3].

2.1. Particle dynamics-methods

Existing studies related to these methods are based on physics. The researchers considered the crowd flows the same as the fluid flow model. These methods could not apply directly to the data of the crowd because particles in the crowd move freely in contrary to the fluid flow.

The authors who used these methods had to impose restrictions. Langevin theory, as an example, could not support the dynamic of the crowd without considering the correlation associated with forces in dense [3]. In [4], the authors attempted to segment the crowd flow and to analyze the instability of the flow by applying Lagrange particle dynamics. The same researcher in [5] proposed a method using a structure-based force model to identify pedestrian in a high crowd scene. The authors in [6] proposed an intelligent method to detect suspicious behavior in a crowd. A Markov random algorithm computed the crowd energy to achieve their objective. Unfortunately, this method depends on the background and it is not accurate for video shaking.

The authors in [7][8] analyzed the flow according to the social force graph technique. In [7], the technique is applied to streak lines of the crowd flow, and in [8] pixels and block levels were the targets.

In [9], the authors used bilinear interaction to analyze the crowd. The authors in [10] attempted to manage the variation of the crowd density by performing a method

Manuscript received June 6, 2020
Manuscript revised June 20, 2020

based on density-independent hydrodynamics. The segmentation results did not support the finer-level crowd.
The researchers in [11] used a spatial-temporal domain of the force model to provide group segmentation. However, their proposed framework requested to learn multi-views for each parameter's combinations. The authors in [12] considered the stability to identify the behavior of the crowd. This method is performed without proceeding the training, object detection, or tracking. The main shortcoming of their proposed method is related to the inability to detect the randomness exposed by the crowd.
A method based on an aggregation of the thermal diffusion and time-series clustering is applied in [13]. The authors aimed to analyze the crowd according to coherent regions, but the merge between the two techniques (motion and non-motion area) provided a lack of coherency.
In [14], the authors proposed a model based on agents for crowd scenes. The proposed mixture dynamic pedestrian-agents learned the behavior of crowds. But, their proposition did not achieve accurate results when the scene contains many density variations.
Other studies described in [15] provided an analysis of the crowd behavior based on a density and velocity. The authors in [16] attempted to monitor the crowd according to walking features. Results showed high accuracy when the crowd is similar to a balanced human flow.
A recent study [2] tried to understand crowd behavior based on linear flow representation. A Langevin equation-based force model is used to guide the crowd provided by the computer-vision. Experiments led to an improvement in the segmentation of about 5%.

2.2. Computer vision-methods

This section exposed the different techniques based on computer vision used for crowd monitoring.
In [17] the authors used the translation domain to provide the crowd flow segmentation. The problem of this method is presented in the case of non-local translation flow. A method based tracking is introduced in [18]. The authors applied Lucas-Kanade Tracker to the density associated with each frame of the video. This method detects and tracks the crowd, but it suffers from low accuracy. Besides, the time computation is raising.
The authors in [19] used the optical domain to analyze the movement of the crowd. The segmentation method is based on Delaunay triangulation to estimate the smoothing aspect. But, some motion regions were not detected and the accuracy is decreased rapidly in the case of a high crowd scene.
A method based on trajectory clustering has been described in [20] to learn about the crowd motion patterns. For the closed environments, the method in [21] analyzes the pedestrian walking. A sequence waypoints are computed.
In [22], microscopic level processing is done to detect abnormal events in a crowd. At the same level [23], another study used a combined spatial-temporal method to identify suspicious events in the crowd. The method presented in [24] attempted to track pedestrians at the microscopic layer. A multi-stage tracker is performed to localize targets precisely. The authors in [25] identified the crowd behavior by modeling the emotion of pedestrian. A finite machine method [26] is applied to monitor the crowd in closed environments. Unfortunately, this method requested to know information related to the source and the destination pixels.
Another attempt [27] described the crowd behavior via semantic information. In [28], the authors detected the abnormal actions in the crowd using the global grid motion. The fails happened when a loss in motion occurred. A method based on active contour segmentation is described in [28]. Then a tracking bloc-level method followed by clustering is applied. The main limit is the non-real-time.
The authors in [29] used regression to guide Convolutional Neural Network (CNN) classification. This method needed the availability of proper labeling stored in the database to obtain an accurate estimation of the troop. Zhou et al., [30] applied CNN for the spatial and the temporal domains to discover abnormal events in the crowd.
In [31], a conditional random field is described to analyze the crowd flow. This method was inaccurate in the intersection between flows. Ilg et al., [32] monitored the crowd behavior as an optical flow of frames based on a deep learning algorithm. Dynamical scenarios were not supported in this case.
Wang et al., [33] applied Markovian model to extract features related to the optical flow. The authors aimed to detect anomaly in a crowded scene.
Dong et al. [35] introduced the MMNet as a framework to estimate the density of crowd using a scaleware CNN. The authors attempted to overcome the problem of the scale variation of person’s head by using multi-size filters and multi-scale features. The non-uniform partition is solved by considering a multi-level density associated with the special information.
In light of this brief literature review, we conclude the following points:

- Methods based on particle dynamics achieved, in general, an accurate crowd density but shortcomings could be resume on:

  (1) Methods need powerful calculators
  (2) Methods are time-consuming to provide the crowd density at a high-precision.
  (3) Methods cannot be implemented for the Real-Time case.
- Methods based on computer vision are accurate for low and medium crowd density. For high density, methods suffer from:

(1) Density error is lower when the used algorithms supported the Real-Time
(2) Time-consuming when the used algorithms are with high-precision

During this paper, a method based on computer vision is chosen. A BCA-SVM classifier is performed to ensure the accurate result and overcome limits cited bellow.
The next section describes the proposed architecture. Section 4 presents the principles findings in comparison with previous attempts. Finally, we conclude the study and we highlight the main future work.

3. Proposed system-based BCA-SVM architecture

We attempt to predict the crowd density using BCA-SVM classifier. The proposed architecture follows the steps described in figure 1:

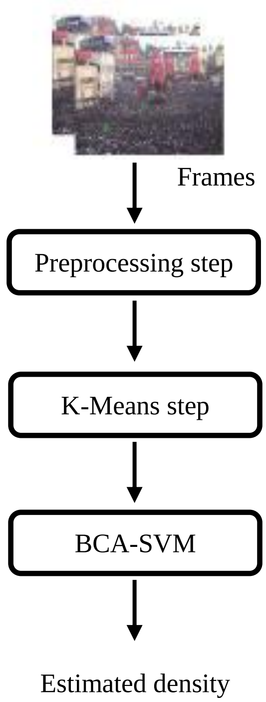

Fig. 1 The proposed architecture based on BCA-SVM

(1) The preprocessing step decreases the noise and provides only useful data in frames.
(2) The K-Means step is applied randomly from selected samples. A balanced condition is considered during the clustering process.
(3) The BCA-SVM is performed to predict the crowd density. This prediction is done under different layers described in figure 2. The balanced value is computed as the result of the decision between the number of zonal with all regions and the number of zonal with specific regions.
The CA-SVM is considered as an improvement of the SVM classifier. It ensures speedups over the SVM algorithm but it causes a small loss on accuracy.
The Balanced CA-SVM aims to keep the speed of the classifier with an enhancement of the accuracy over the traditional SVM.
The SVM classifier is composed of two steps:
(1) Training step: In this step the classifier constructs the model according to the input data. This step speed down the process due to the increased size of the dataset. This is why, the classifer requests more number of processors.
(2) Prediction step: In this step the classifier classify new data. This step is speedy.
The BCA-SVM is focused on training step.

```
Loop
  Loop
    Loop
      Loop
        (1) Check the balanced value
        (2) Compute the centroid of local region
        (3) Perform BCA-SVM process for local region
      Until local region
      (1) Compute the centroid of sub region
      (2) Perform BCA-SVM process for sub region
    Until sub region
    (1) Compute the centroid of region
    (2) Perform BCA-SVM process for region
  Until region
  Perform BCA-SVM process for zonal
Until zonal
```

Fig. 2 The proposed BCA-SVM algorithm

4. Experiment

The experiment section presents the achieved results. These findings are discussed and compared with several related works methods.

The crowd estimation based on BCA-SVM is evaluated using many datasets ad well UCSD, Shanghaitech, UCF_CC_50, and UCF_QNRF [34]. These Datasets are chosen to cover low, medium, and high dense crowds. The number of pedestrians related to datasets moves from 20 persons to 12000 persons. The comparison results are introduced according to findings and datasets highlighted in our previous paper [34].
The crowd estimation is evaluated through mainly two factors: The Mean Absolute Error (MAE), see eq. 1, and the Mean Squared Error (MSE), see eq. 2.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|X_i - X| \quad (1)$$

$$MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - X)^2} \quad (2)$$

The proposed method achieves the following findings presented in table 1.

Table 1: Findings of the BCA-SVM method for the crowd estimation

Datasets	MAE	MSE
Shanghaitech (Part A)	14.2	85
UCSD	0.75	0.83
UCF_CC_50	88.3	137.4
UCF_QNRF	89	145

These results prove the choice of the BCA-SVM method to estimate the crowd. The errors seem to be good. These findings could be more significant in comparison with other attempts. Therefore, in table 2, we attempt to compare our findings and other studies presented in [34] based on the MAE and the MSE through different datasets.

Table 2: Comparative results according to different datasets

Methods	Shanghaitech (Part A)		UCSD		UCF_CC_50		UCF_QNRF	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Zhikang et al. [34]	17.5	103.5	1.01	1.29	291.6	337	-	-
C. Zhang et al. [34]	49.8	336.1	1.6	3.31	467	498.5	-	-
Walach et al. [34]	-	-	1.53	-	474	-	-	-
Sam et al. [34]	21.6	135	1.62	2.1	318.1	439.2	228	445
Liping et al. [34]	65.6	107.2	1.08	1.44	257	343.9	-	-
Y. Zhang et al. [34]	26.4	173.2	1.08	1.35	377.6	509.1	277	426
Lempitsky et al. [34]	-	-	1.7	-	493.4	487.1	-	-
Sindaghi et al. [34]	20	152.4	-	-	322.8	397.9	-	-
Idrees et al. [34]	-	-	-	-	419.5	541.6	315	508
Dong et al. [35]	60.8	99	0.77	0.95	209.7	309.7	104	178
Our method	**14.2**	**85**	**0.75**	**0.83**	**88.3**	**237.4**	**89**	**145**

Based on table 2, we conclude that the proposed method is the best one. It gives an accurate result either for a low dense crowd, medium dense crowd, and even for a high dense crowd. Shown results prove that the computed error using UCF_CC_50 and UCF_QNRF datasets seems to be approximate. This observation demonstrates that the error in the case of the high-dense crowd increases slowly. We can conclude that this method supports high-crowd scenes.
We discuss the achieved results according to the density level (low, medium, high). Figure 3 displays the MAE and the MSE values computed for different methods by using the UCSD dataset. The proposed method gave the density of the crowd with the lowest error. The average error, in this case, is about 4%. For the medium density of the crowd, figure 4 shows the MAE and the MSE values performed using the Shanghaitech (Part A) and UCF_CC_50 datasets. Our method done by the BCA-SVM algorithm predicts the crowd density accurately. The average error is around 4.35%. Figure 5 related to the high-density crowd highlights the importance of the proposed method, especially in this case. The average error is about 1.2%.

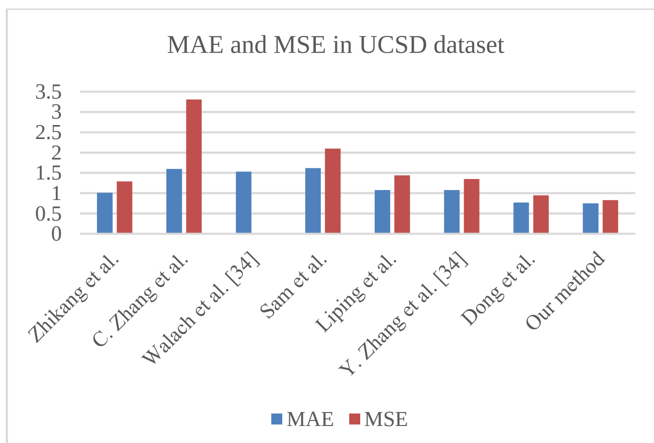

Fig. 3 MAE and MSE in UCSD dataset

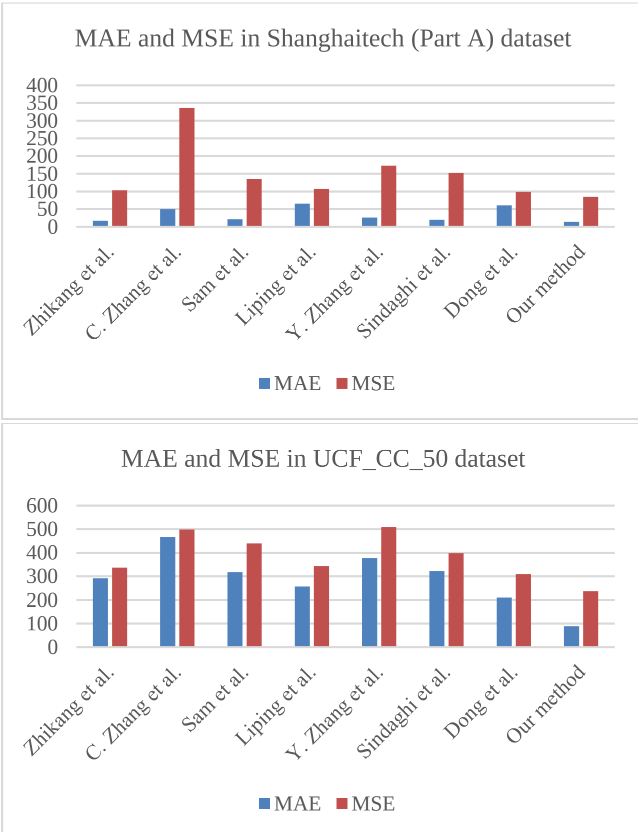

Fig. 4 MAE and MSE in Shanghaitech and UCF_CC-50 datasets

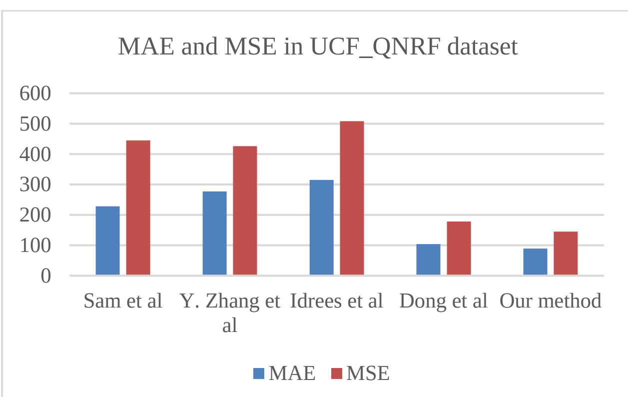

Fig. 5 MAE and MSE in UCF_QNRF dataset

Table 3 presents the variation of the error under the dataset.

Table3. The Average Error (AE) of the proposed method

	Low-density	**Medium-density**	**High-density**
AE	3.75 %	4.35%	**1.2%**

This table proves that BCA-SVM provides a high precision result in the case of the high-density. This method is more suitable for the crowd over 10000 pedestrians,

In addition to that, the presented method is evaluated in terms of runtime and memory requested. The estimation of the crowd belongs to the field of the monitoring and the management of the crowd and the real-time exigence is a critical factor. Table 4 presents the comparison results related to the hardware feature in the case of the Shanghaitech (part A).

Table 4: Hardware feature results

Methods	**Runtime**	**Memory**
Y. Zhang et al. [34]	25 ms	**527.6 KB**
Sam et al. [34]	153 ms	57.6 MB
Dong et al. [35]	86 ms	22.4 MB
Our method	**13 ms**	3.2 MB

Results show that the estimation of the crowd using the BCA-SVM classifier is a Real-Time system seen that the execution time is around 13 ms. The requested memory for the proposed method is 3.2 Mb which is acceptable. But the memory usage is the minimum when using Y.Zhang et al. method.
In summary, The BCA-SVM used for the crowd provides an accurate estimation even for a high dense crowd. Besides, this method could be implemented in standard board microcontroller-based because it did not request a huge memory.

5. Conclusions

The behavior of the crowd is still an interesting field. Provide an accurate estimation of the crowd is a challenging task especially for a high-dense crowd. Our proposed method based on both K-means and BCA-SVM classifiers provides the estimation of the crowd through video streaming. The method is evaluated across many datasets. When using the UCSD dataset as a low-dense crowd (less than 25 persons), we conclude that the estimation error did not exceed 1 pedestrian in the crowd. When using Shanghaitech Part A (less than 3000 persons) dataset, and UCF_CC_50 (less than 4000 persons) dataset, the error is inferior to 88. An error of about 89 persons is achieved for a high-dense crowd when using the UCF_QNRF (less than 12000 persons) dataset.
The proposed method is evaluated also in terms of runtime and memory features. We found that BCA-SVM provides the minimum time of execution (about 13ms) and ensures the Real-Time constraint. This attempt can be extended to embed the proposed system on a smart camera. Further work includes handling multi-view issues from multi-camera.

References

[1] V. A. Sindagi and V. M. Patel, "A survey of recent advances in cnn-based single image crowd counting and density estimation," Pattern Recognit. Lett., vol. 107, pp. 3–16, 2018.

[2] S. Behera, D. P. Dogra, M. K. Bandyopadhyay, and P. P. Roy, "Estimation of linear motion in dense crowd videos using Langevin model," Expert Syst. Appl., vol. 150, p. 113333, 2020.

[3] X. Zhang, Q. Yu, and H. Yu, "Physics inspired methods for crowd video surveillance and analysis: a survey," IEEE Access, vol. 6, pp. 66816–66830, 2018.

[4] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–6.

[5] S. Ali and M. Shah, "Floor fields for tracking in high density crowd scenes," in European conference on computer vision, 2008, pp. 1–14.

[6] Z. Zhong, N. Ding, X. Wu, and Y. Xu, "Crowd surveillance using Markov random fields," in 2008 IEEE International Conference on Automation and Logistics, 2008, pp. 1822–1828.

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

[8] Q.-G. Ji, R. Chi, and Z.-M. Lu, "Anomaly detection and localisation in the crowd scenes using a block-based social force model," IET Image Process., vol. 12, no. 1, pp. 133–137, 2017.

[9] S. Wu, H. Su, H. Yang, S. Zheng, Y. Fan, and Q. Zhou, "Bilinear dynamics for crowd video analysis," J. Vis. Commun. Image Represent., vol. 48, pp. 461–470, 2017.

[10] H. Ullah, M. Uzair, M. Ullah, A. Khan, A. Ahmad, and W. Khan, "Density independent hydrodynamics model for crowd coherency detection," Neurocomputing, vol. 242, pp. 28–39, 2017.

[11] R. Li and R. Chellappa, "Group motion segmentation using a spatio-temporal driving force model," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 2038–2045.

[12] B. Solmaz, B. E. Moore, and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 10, pp. 2064–2070, 2012.

[13] W. Lin, Y. Mi, W. Wang, J. Wu, J. Wang, and T. Mei, "A diffusion and clustering-based approach for finding coherent motions and understanding crowd scenes," IEEE Trans. Image Process., vol. 25, no. 4, pp. 1674–1687, 2016.

[14] B. Zhou, X. Tang, and X. Wang, "Learning collective crowd behaviors with dynamic pedestrian-agents," Int. J. Comput. Vis., vol. 111, no. 1, pp. 50–68, 2015.

[15] A. Nicolas, M. Kuperman, S. Ibañez, S. Bouzat, and C. Appert-Rolland, "Mechanical response of dense pedestrian crowds to the crossing of intruders," Sci. Rep., vol. 9, no. 1, pp. 1–10, 2019.

[16] R. Ye et al., "Experimental study of pedestrian flow through right-angled corridor: uni-and bidirectional scenarios," J. Stat. Mech. Theory Exp., vol. 2019, no. 4, p. 43401, 2019.

[17] S. Wu, Z. Yu, and H.-S. Wong, "Crowd flow segmentation using a novel region growing scheme," in Pacific-Rim Conference on Multimedia, 2009, pp. 898–907.

[18] F. Santoro, S. Pedro, Z.-H. Tan, and T. B. Moeslund, "Crowd analysis by using optical flow and density based clustering," in 2010 18th European Signal Processing Conference, 2010, pp. 269–273.

[19] S. Wu, Z. Yu, and H.-S. Wong, "A shape derivative based approach for crowd flow segmentation," in Asian Conference on Computer Vision, 2009, pp. 93–102.

[20] W. Lu, X. Wei, W. Xing, and W. Liu, "Trajectory-based motion pattern analysis of crowds," Neurocomputing, vol. 247, pp. 213–223, 2017.

[21] M. Nasir, C. P. Lim, S. Nahavandi, and D. Creighton, "Prediction of pedestrians routes within a built environment in normal conditions," Expert Syst. Appl., vol. 41, no. 10, pp. 4975–4988, 2014.

[22] F. Anwar, I. Petrounias, T. Morris, and V. Kodogiannis, "Mining anomalous events against frequent sequences in surveillance videos from commercial environments," Expert Syst. Appl., vol. 39, no. 4, pp. 4511–4531, 2012.

[23] K. Kardas and N. K. Cicekli, "SVAS: surveillance video analysis system," Expert Syst. Appl., vol. 89, pp. 343–361, 2017.

[24] G. S. Walia, S. Raza, A. Gupta, R. Asthana, and K. Singh, "A novel approach of multi-stage tracking for precise localization of target in video sequences," Expert Syst. Appl., vol. 78, pp. 208–224, 2017.

[25] Y. Mao, Z. Li, Y. Li, and W. He, "Emotion-based diversity crowd behavior simulation in public emergency," Vis. Comput., vol. 35, no. 12, pp. 1725–1739, 2019.

[26] A. Fernández-Caballero, J. C. Castillo, and J. M. Rodr\'\iguez-Sánchez, "Human activity monitoring by local and global finite state machines," Expert Syst. Appl., vol. 39, no. 8, pp. 6982–6993, 2012.

[27] H. Fradi, B. Luvison, and Q. C. Pham, "Crowd behavior analysis using local mid-level visual descriptors," IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 3, pp. 589–602, 2016.

[28] S. Li, Y. Yang, and C. Liu, "Anomaly detection based on two global grid motion templates," Signal Process. Image Commun., vol. 60, pp. 6–12, 2018.

[29] L. Cao, X. Zhang, W. Ren, and K. Huang, "Large scale crowd analysis based on convolutional neural network," Pattern Recognit., vol. 48, no. 10, pp. 3016–3024, 2015.

[30] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei, and Z. Zhang, "Spatial--temporal convolutional neural networks for anomaly detection and localization in crowded scenes," Signal Process. Image Commun., vol. 47, pp. 358–368, 2016.

[31] S. S. S. Kruthiventi and R. V. Babu, "Crowd flow segmentation in compressed domain using CRF," in 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 3417–3421.

[32] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2462–2470.

[33] T. Wang et al., "Abnormal event detection based on analysis of movement information of video sequence," Optik (Stuttg)., vol. 152, pp. 50–60, 2018.

[34] S. A. Alshaya, "Open Challenges for Crowd Density Estimation," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 1, 2020, doi: 10.14569/IJACSA.2020.0110123.

[35] L. Dong, H. Zhang, Y. Ji, and Y. Ding, "Crowd Counting by Using Multi-level Density-based Spatial Information: A Multi-scale CNN Framework," Inf. Sci. (Ny)., 2020.