https://doi.org/10.2352/ISSN.2470-1173.2020.6.IRIACV-072

CROWD CONGESTION DETECTION IN VIDEOS

Sultan Daud Khan[1] *Maqsood Mahmud* [2] *Habib Ullah*[1] *Mohib Ullah*[3] *Faouzi Alaya Cheikh*[3]

[1] College of Computer Science and Engineering, University of Ha'il, Saudi Arabia.
[2] Department of Management Information System,
University of Imam Abdulrahman Bin Faisal, Dammam, Saudi Arabia.
[3] Department of Computer Science, Norwegian University of Science and Technology, Norway.

ABSTRACT

Automatic detection of crowd congestion in high density crowds is a challenging problem, with substantial interest for safety and security applications. In this paper, we propose a method that can automatically identify and localize congested regions in crowded videos. Our proposed method is based on the notion that pedestrians in the congested region follow a particular behavior. Pedestrians in the congested areas cannot move freely due to space unavailability and tend to undergo lateral oscillations. In our method, we first extract trajectories by using particle advection technique and then compute oscillatory features for each trajectory. Trajectories with higher oscillation values and with less proximity are clustered, indicating the congested regions. We perform experiments on a diversity of challenging scenarios. From the experimental results, we show that our method provides precise localization of congested regions in crowd videos.

***Index Terms*—** Crowd analysis, crowd congestion, anomaly detection, pedestrian dynamics.

1. INTRODUCTION

Crowd congestion is a frequent phenomenon in large gatherings such as those related to sports, festivals, concerts, and religious events which attract thousands of people in a restricted environment. As the security of computer networks and resources are important for an organization [1, 2], congestion detection and managing the visual scene has paramount importance for the public safety. In order to ensure public safety, it is important to understand crowd dynamics and congestion circumstances. However, despite several strides in crowd management, safety measures and the adoption of technology, crowd disasters still occur with consistent frequency.

The current practice of understanding crowd dynamics is often based on manual analysis of video data. Based on rules of thumb gained from previous experience as well as simulation models [3], limits on the number of people that can be simultaneously present at a venue are decided. For understanding crowd dynamics, several physicals models have been proposed. The popular among them are the dynamic model [4], social force model [5] and cellular automata [6] which model pedestrian dynamics on a microscopic level. In [7], the basic social force model is extended in order to incorporate effects of panic by adding further random forces. Empirical studies are performed in [8,9] to understand human behavior and improve the existing physical models by incorporating more parameters such as crowd density, speed, flow, and crowd pressure. There are many limitations of the physical models. For instance, the results and observations are based on experimental data (captured in a constrained environment) while physical models require proper calibration and validation by means of real-time data.

Surveillance cameras are used as an alternative approach for understanding crowd dynamics. Usually these kinds of analysis involve analysts sitting in a control room and looking for specific activities. However, these kinds of video analysis lead to human errors due to limited capabilities of human operators to analyze and infer critical information from multiple videos over a long period of time [10]. Therefore, as a solution, there is an increased interest in automated crowd analysis [11]. Since there has been an interest in understanding of crowd dynamics [12] in general, detecting crowd congestion has not been studied in depth. Most of the recent work in crowd analysis has focused on tracking [13], segmentation [14], crowd flow understanding [15], and anomaly detection [16]. It is worth noticing that anomaly detection methods do not include congestion detection in general [17]. Most of the methods in the literation consider panic situation and circulation of non-pedestrian entities in the crowd as anomalies. However, a congested situation in a crowded scene is characterized by different dynamics.

Crowd congestion is a condition when individuals in a crowd are prevented from moving smoothly due to overcrowding, and are unable to make progress towards their desired goal at normal speed. Our goal in this paper is to use only easily measurable motion features to reliably localize pedestrian congestion in surveillance videos in a variety of scenarios. It has been reported in crowd dynamics literature that individuals in a crowd tend to undergo lateral oscillations, that is, to and fro motion orthogonal to their desired direction of motion [18]. This is attributed to the shifting of weight from one foot to another, termed swaying, as longitudinal progress is restricted [19]. We exploit this feature to look for groups of trajectories that change from smooth motion to oscillatory harmonics, as this corresponds to a change from unrestricted flow to a substantial restriction in the ability to make progress towards the desired goal. We show that an oscillatory metric derived on this principle can reliably localize and highlight areas of congestion in diverse scenarios.

2. RELATED WORK

Most similar works are reported in [20–23]. In [20], the authors propose an automatic vision system for detecting congestion in real-time videos. In their proposed framework, dense optical flow is computed between two consecutive frames of the input video. After optical flow computation, two-dimensional histograms of motion magnitude and motion direction of flow vectors are computed. Next, k-means clustering algorithm is adopted for clustering two-dimensional histograms. Computed histograms show small motion (low magnitude) along major directions reflecting lateral oscillation of people which is a potential indicator of congestion situation. Similar approach is adopted in [21], where entropy is computed as an indicator to crowd congestion and captures the dispersion of velocity distribution of magnitude and direction simultaneously. To the best of our knowledge, these are the only reported vision based methods that detect congestion in real-time videos. The limitations of these methods are: (1) perform global analysis of video and can not localize congestion (2) experiments are carried out using only one video which limits the effectiveness of their methods when applied to more diverse scenarios.

In contrast to earlier literature, we focus on localizing congestion in a variety of scenarios, such as religious rituals, concerts and train stations. We characterize congestion as a spatio-temporal phenomenon, and model it as an evolving and dynamic characteristic of crowd motion. Therefore, we demonstrate congestion localization in diverse scenarios.

3. CROWD CONGESTION

In this section, we discuss our proposed method for congestion detection in crowded scenes. Our method first divides the input video into multiple overlapping temporal segments. We then extract motion information (trajectories) from each segment by particle advection technique. After extracting trajectories from each segment, we compute oscillation feature for each trajectory and generate corresponding oscillation maps. After quantizing oscillation maps, we select candidate critical locations. We then employ spatial and temporal filtering to the oscillation maps. Locations with high oscillation values and low proximity are identified as congestion locations.

3.1. Computation of oscillatory feature of trajectory

In this section, we discuss how do we compute oscillatory feature of a trajectory. For a given trajectory $S\{\Pi, \beta\}$, where Π represents the spatial (horizontal and vertical) coordinates and β represents orientations, with S containing k points, we perform the following steps,

1. Compute circular mean, i.e, β_μ of β as in [24] for the given trajectory S.
2. Compute circular distance η from the mean for all trajectory points, i.e, $\eta_i = (\beta_\mu - \beta_i)$, with $0 \leq i < k$ and where β_i is i^{th} point of the trajectory.
3. Compute an indicator of oscillation

$$I_i = \begin{cases} 1, & \text{if (} \beta_\mu \text{ - } \beta_i \text{) } < \rho \text{ where } \rho \text{ is set to 0.785} \\ 0 & \text{otherwise} \end{cases}$$

4. Compute the overall trajectory oscillation indicator

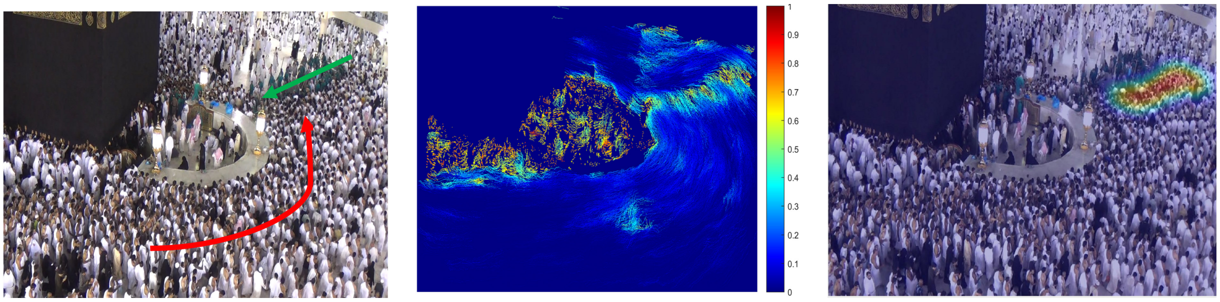
$$I = \frac{\sum_{i=0}^{k-1} I_i}{k}$$

Fig. 1(a) shows a sample frame from a video. Fig. 1(b) shows the corresponding oscillation map. As illustrated in Fig. 1(b), small oscillation in pedestrian trajectories indicates smooth motion of pedestrians and higher values indicate that the motion of pedestrians is blocked, leading to congestion as shown in Fig. 1(c). However, similar higher oscillations are also observed in pedestrian trajectories during loitering behavior of pedestrians. These trajectories do not belong to congested regions and hence treated as false positives.

In order to suppress false positives and detect precise congested regions, we quantize oscillation map into 5 layers (layers). Oscillation map is a spectrum of different oscillation values ranging from 0 to 1, where 0 represents smooth trajectory with no oscillations and 1 represents highest oscillation in the trajectory. We quantize the oscillation map with a step of 0.2, i.e, [0, 0.2], (0.2, 0.4], (0.4, 0.6], (0.6, 0.8], (0.8, 1]. We encode the layers with colors. Layer1, layer2, layer3, layer4, and layer5 are annotated in white, cyan, yellow, blue and red, respectively. We then sample trajectories from the highest layer, i.e, 5th (red) as shown in Fig. 3. Fig. 2 illustrates the step-wise process of achieving critical blobs. As obvious from the Fig. 2(c), some false positive are also detected as potential congested locations. Since these blobs are produced at random in space and time, therefore, they can be suppressed by employing a sptio-temporal filter.

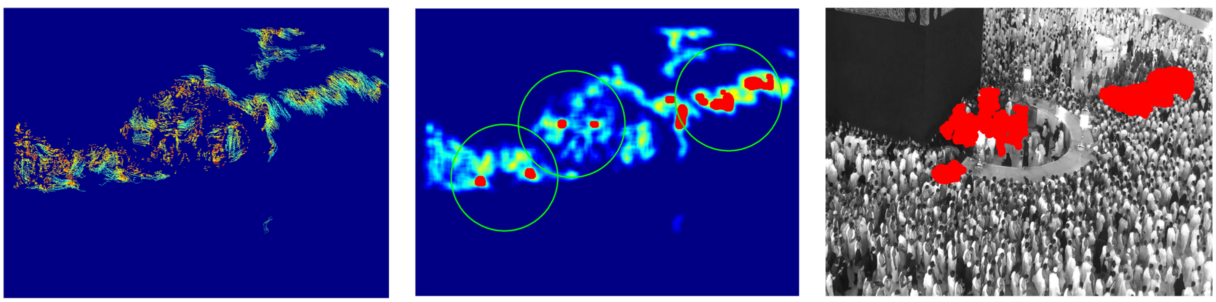
3.1.1. *Spatial and Temporal Filters*

The critical blobs obtained in the previous section need to be refined in order to suppress false positive blobs which are produced due the reasons discussed before. The oscillations in



(a) (b) (c)

Fig. 1: (a) Sample frame, (b) Corresponding oscillation map and (c) Detected congested area.



(a) (b) (c)

Fig. 2: (a) sample trajectories of highest level, (b) gaussian filter and local peaks are clustered (in green circles) (c) detected critical blobs.

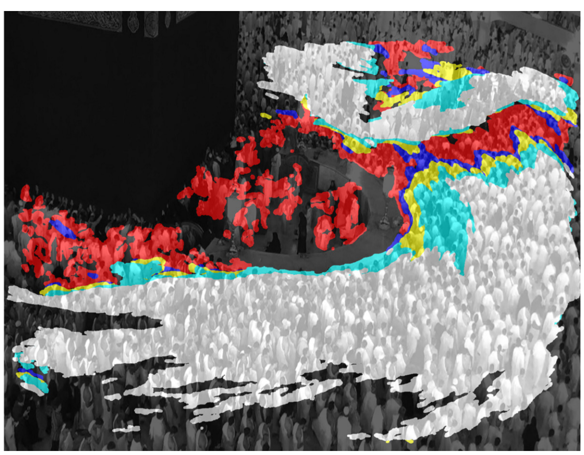


Fig. 3: Quantized oscillation map with color encoded layers.

trajectories of pedestrians gradually decrease, if we go away (in the reverse direction of the flow) from the epicenter of the congested area. This phenomena is understandable by Fig. 3, where the oscillation map is quantized in color encoded layers. Critical blobs (in red) are the actual congested blobs if followed by the subsequent layers, while the false positive blobs do not follow the same phenomena. For each critical blob (in red) B_i, we move in the reverse direction of the flow and sample layer's color. Critical blob B_i will be classified as congested blob, if all the subsequent layer's color are sampled. As obvious from the Fig. 3, critical blobs created by pedestrian's loitering and other behaviors could not sample the subsequent layer colors and are suppressed. We then apply a temporal filter to the remaining set of blobs. We define a temporal window of size 5 video segments. For each blob B_i, we find the number of detections of blob B_i in the given temporal window, i.e . we suppressed the blob, if the following ratio is less than 50 %.

$$\frac{\text{Number of detections of blob}}{\text{window size (5 in this case)}} \times 100$$

4. EXPERIMENTAL RESULTS

This section discusses the qualitative analysis of the results obtained from experiments. We carried out our experiments on a PC of 2.6 GHz (Core i7) with 16.0 GB memory and video sequences obtained from other research groups and acquired through field observations. The videos have different field of view, resolution, frame rates, and duration, yet our method performed well in all cases. We first divide each video into a set of temporal segments with 25% overlap.

First row of Fig. 4 shows pilgrims sequence which was taken in the context of the yearly pilgrimage to Makkah, Saudi Arabia, and it shows a very high density situation. The resolution of video is 1080 x 1440 pixels and composed of 28000 frames (18.66 min) long with frame rate of 25 fps. In this video sequence, a large group of pedestrians are performing annual ritual by circulating around Kaaba. Another small group of pedestrians (the direction of which is orthogonal to large flow) trying to penetrate inside the large flow, blocking the motion of the large flow. As the small group of pedestrians penetrates forward, causes a congestions at different locations at different temporal segments. This video sequence shows the example of dynamic congestion, where the location of congested area changes with time.

Second row of Fig. 4 shows a station sequence which is acquired from other research group. This resolution of this video is 360 x 480 pixels and it covers a low density scenario. The video is relatively short with 250 frames with 25 fps. In this video sequence, pedestrians try to get on and off of a train; flows change in time due to the congestion that arises near one of the entrances, and the density varies with time. As a result of this variation in density which increases with time, the area of congestion also increases. As obvious from the second row, a smaller congested area is detected in the first segment, where the density is low in comparison to fourth segment, where larger congested area is detected. This sequence represents a scenario of growing congestion, where the size of congestion increases with time.

Third row of Fig. 4 shows a concert sequence which was recorded in the context of the concert held in San Siro stadium, Milan, Italy. This video sequence covers very high density crowd, the resolution is 1080 x 1920 with frame rate of 30 and it is composed of 5000 frames. This video sequence is a special case of evacuation scenario, where large number of pedestrians, after attending the concert trying to leave through a narrow exit. Intuitively, such narrow exit creates congestion which is efficiently detected by our method. This is a scenario of a fixed congestion, where congestion of almost the same area is created at the same locations in all temporal segments.

Since our proposed method focuses on congestion detection and localization, therefore, comparison with state-of-the-art methods [20, 21] becomes irrelevant. Their methods only detect congestion in a global fashion and are limited to only one video sequence (LoveParade). LoveParade video sequence covers an evacuation scenario, and the problems of congestion detection and localization in such sequence are already addressed in our concert sequence. Datasets that covers congestion situations are rarely available, therefore, we limited our experimental results to our three but diverse video sequences.

Additionally in order to show the robustness of our proposed method, we use another video sequence which is downloaded from Youtube and recorded at Shibuya Crossing Japan, one of the world busiest crossing and famous for its scramble crossing. We asked all 20 coders to annotate the video.

To quantitatively evaluate the performance, we compute two measures detection accuracy (DA) and localization accuracy (LA). Detection accuracy is calculated by DA = # of frames correctly detected / Total # of frames. Localization accuracy is based on the overlap of detection region and ground truth. Overlap among the detected and ground truth region is calculated by $\frac{C}{max(N,M)}$, where C is the number of common points among detection and ground truth region, N is the number of points in detected region where as M is the number of points in ground truth region. We also compute missed detection by (MD) = # of missed frames / Total # of frames. Table 1 shows the quantitative results. In Table 1. DA, MD and LA are computed for all video sequences. It is obvious from the table that our proposed method detects congestion with 100% detection accuracy in almost all the video sequences. The reduce performance in *Hseq01* videos is due to the fact that relatively small blobs are produced in some of the temporal segments of the videos which are filtered by the threshold λ. λ represents the area of the blob size and its value is fixed to 1000 pixels for all analysed videos. However, the performance is increased to 100% when the value of λ is reduced to 500 pixels.

Table 1: Quantitative results. We present the quantitative results in terms of correct detection, missed detection, and overlapped.

Video	**Correct detect**	**Missed detect**	**Overlap**
Station	100%	0.0%	83.46%
Concert	100%	0.0%	82.42%
Hseq01	61.07%	38.92%	67.09%
Average	**96.95%**	**3.04%**	**83.21%**

All the video sequences in the dataset exhibit congestion and our proposed method successfully models the dynamics of congestion regardless of the scene characteristics. These results further indicate that computing oscillation feature of trajectory is capable of locating congestion in the regions of crowd scene. Our method locates regions of congestion and does not label individuals.

5. CONCLUSION

Using motion features, we generate oscillation map which is exploited in a spatio-temporal fashion to detect and localize congestion in diverse videos. The results of our method indicate that the method is effective in detection and localization of congestion in the crowd.

6. REFERENCES

[1] Muhammad Mudassar Yamin, Basel Katt, Kashif Sattar, and Maaz Bin Ahmad, “Implementation of insider threat detec-

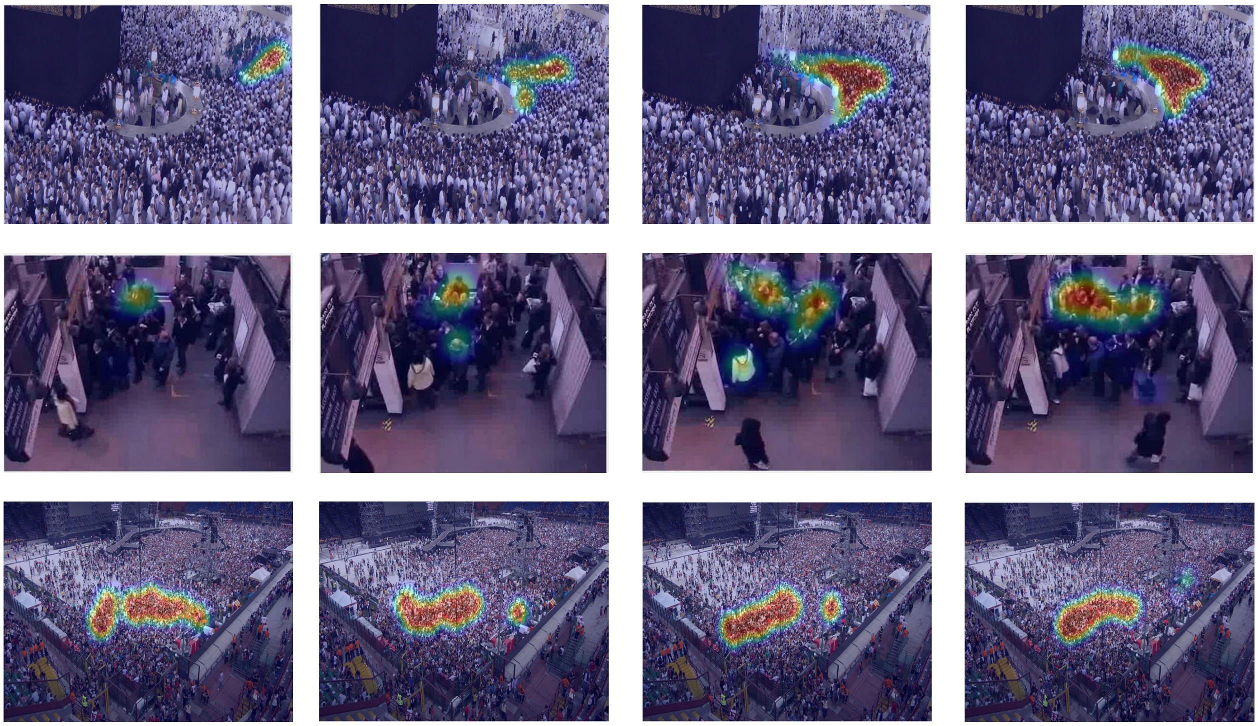


Fig. 4: First row shows detected congestion locations in pilgrims sequence. Second row shows detected congestion locations in station sequence and the third row shows detected congestion locations in the concert sequence. The columns represent different temporal segments of the corresponding video sequences.

tion system using honeypot based sensors and threat analytics," in *Future of Information and Communication Conference*. Springer, 2019, pp. 801–829.

[2] Muhammad Mudassar Yamin and Basel Katt, "Modeling attack and defense scenarios for cyber security exercises," in *5th interdisciPlinary cyber research conference 2019*, 2019, p. 7.

[3] G Keith Still, *Introduction to crowd science*, CRC Press, 2014.

[4] Mohib Ullah, Habib Ullah, Nicola Conci, and Francesco GB De Natale, "Crowd behavior identification," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 1195–1199.

[5] Dirk Helbing and Peter Molnar, "Social force model for pedestrian dynamics," *APS physical review E*, vol. 51, no. 5, pp. 4282, 1995.

[6] C Burstedde, K Klauck, A Schadschneider, and J Zittarz, "Simulation of pedestrian dynamics using a 2-dimensional cellular automaton, 2001," *Physica A*, vol. 295, pp. 507.

[7] Dirk Helbing, Illés Farkas, and Tamas Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487–490, 2000.

[8] Dirk Helbing, Anders Johansson, and Habib Zein Al-Abideen, "Dynamics of crowd disasters: An empirical study," *APS physical review E*, vol. 75, no. 4, pp. 046109, 2007.

[9] Bernhard Steffen and Armin Seyfried, "Methods for measuring pedestrian density, flow, speed and direction with minimal scatter," *Elsevier physica A: statistical mechanics and its applications*, vol. 389, no. 9, pp. 1902–1910, 2010.

[10] Noah Sulman, Thomas Sanocki, Dmitry Goldgof, and Rangachar Kasturi, "How effective is human video surveillance performance?," in *IEEE international conference on pattern recognition*, 2008, pp. 1–3.

[11] Teng Li, Huan Chang, Meng Wang, Bingbing Ni, Richang Hong, and Shuicheng Yan, "Crowded scene analysis: A survey," *IEEE transactions on circuits and systems for video technology*, vol. 25, no. 3, pp. 367–386, 2015.

[12] Habib Ullah, Ahmed B Altamimi, Muhammad Uzair, and Mohib Ullah, "Anomalous entities detection and localization in pedestrian flows," *Neurocomputing*, vol. 290, pp. 74–86, 2018.

[13] Mohib Ullah, Faouzi Alaya Cheikh, and Ali Shariq Imran, "Hog based real-time multi-target tracking in bayesian framework," in *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2016, pp. 416–422.

[14] Habib Ullah, Mohib Ullah, and Muhammad Uzair, "A hybrid social influence model for pedestrian motion segmentation," *Neural Computing and Applications*, pp. 1–17, 2018.

[15] Ning Liu, Yongchao Long, Changqing Zou, Qun Niu, Li Pan, and Hefeng Wu, "Adcrowdnet: An attention-injective deformable convolutional network for crowd understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3225–3234.

Fig. 5: Temporal segments of Shibuya crossing, where no congestion is detected.

[16] Habib Ullah and Nicola Conci, "Crowd motion segmentation and anomaly detection via multi-label optimization," in *ICPR workshop on pattern recognition and crowd analysis*, 2012.

[17] Habib Ullah, *Crowd Motion Analysis: Segmentation, Anomaly Detection, and Behavior Classification*, Ph.D. thesis, University of Trento, 2015.

[18] Xuan Liu, Weiguo Song, and Jun Zhang, "Extraction and quantitative analysis of microscopic evacuation characteristics based on digital image processing," *Elsevier physica A: statistical mechanics and its applications*, vol. 388, no. 13, pp. 2717–2726, 2009.

[19] Serge P Hoogendoorn and Winnie Daamen, "Pedestrian behavior at bottlenecks," *Informs transportation science*, vol. 39, no. 2, pp. 147–159, 2005.

[20] Barbara Krausz and Christian Bauckhage, "Loveparade 2010: Automatic video analysis of a crowd disaster," *Elsevier journal of computer vision and image understanding*, vol. 116, no. 3, pp. 307–319, 2012.

[21] Lida Huang, Tao Chen, Yan Wang, and Hongyong Yuan, "Congestion detection of pedestrians using the velocity entropy: A case study of love parade 2010 disaster," *Elsevier physica A: statistical mechanics and its applications*, vol. 440, pp. 200–209, 2015.

[22] Saleh Basalamah, Sultan Daud Khan, and Habib Ullah, "Scale driven convolutional neural network model for people counting and localization in crowd scenes," *IEEE Access*, 2019.

[23] Habib Ullah, Lorenza Tenuti, and Nicola Conci, "Gaussian mixtures for anomaly detection in crowded scenes," in *Video surveillance and transportation imaging applications*, 2013, vol. 8663, p. 866303.

[24] Philipp Berens et al., "Circstat: a matlab toolbox for circular statistics," *J Stat Softw*, vol. 31, no. 10, pp. 1–21, 2009.