Deep Learning Approach for Divergence Behavior Detection at High Density Crowd

M. U. Farooq, M. N. Mohamad Saad, Y. Saleh, and S. Daud Khan

Abstract At high-density crowd gatherings, people naturally escape from the region where any unexpected event happens. Escape in high-density crowds appears as a divergence pattern in the scene and timely detecting divergence patterns can save many human lives. In this paper, we propose to physically capture crowd normal and divergence motion patterns (or motion shapes) in form of images and train a shallow convolution neural network (CNN) on motion shape images for divergence behavior detection. Crowd motion pattern shape is obtained by extracting ridges of Lagrangian Coherent Structure (LCS) from the Finite-Time Lyapunov Exponent (FTLE) field and convert ridges into the grey-scale image. We also propose a divergence localization algorithm to pinpoint anomaly location(s). Experimentation is carried out on synthetic crowd datasets simulating normal and divergence behaviors at the high-density crowd. Comparison with state-of-the-art methods shows our method can obtain better accuracy for both divergence behavior detection and localization problems.

Keywords Divergence · FTLE · LCS · Motion estimation · Image shape

1 Introduction

Divergence detection at the high-density crowd is a tough task due to several challenges involved in high-density crowd videos e.g., few pixels available per head, extreme occlusion, cluttering and noise, and perspective problems, etc. If crowd

M. U. Farooq (✉) · M. N. Mohamad Saad
Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, Tronoh, Malaysia

Y. Saleh
Electrical Engineering Department, Faculty of Engineering and Islamic Architecture, Umm Al-Qura University, Makkah, Saudi Arabia

S. Daud Khan
Department of Computer Science, National University of Technology, Islamabad, Pakistan
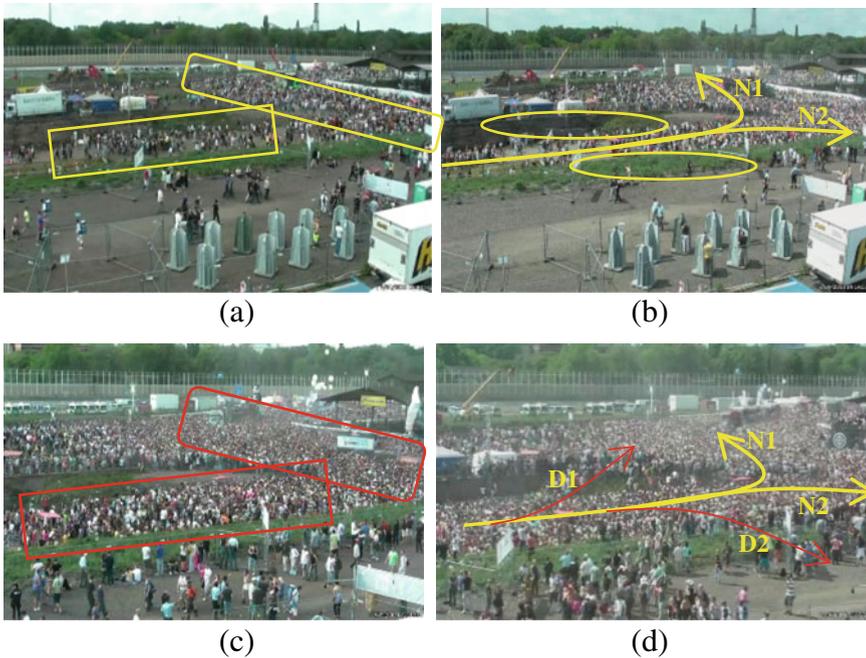
Fig. 1 Demonstration of crowd divergence at Love parade 2010: **a** crowd walking under normal conditions with low density **b** crowd walking paths N1, N2 under normal conditions **c** high-density crowd within the same region **d** crowd diverging through paths D1 and D2

divergence is not detected earlier at its development stage, it may lead to larger disasters like a stampede. Figure 1 shows an example of high-density crowd divergence behavior (Love parade 2010 musical festival [1]) where divergence eventually leads to disastrous stampede. Figure 1a, b demonstrate high-density crowd normal behavior following paths N1 and N2, whereas in Fig. 1c, d, a critical situation is shown where the incoming crowd is blocked by a stationary crowd and diverge through D1, D2 paths. Such divergence situations are common in mass gatherings where the whole crowd is marching towards a common destination and with an increase of density, ends up with a half-stationary half-moving crowd segments that result in divergence behavior.

Previous divergence detection methods [2, 3] learn manual motion features for every individual in the crowd from optical flow (OPF) including location, direction, magnitude, etc. An inherent problem with such methods is with an increase of crowd density, it is almost impossible to capture individual-level motion information and one must learn global crowd features. Later several methods have been developed to capture global crowd motion information e.g., optical flow with pathline trajectories [4–7], pathlines with Lagrangian particle analysis [8], streakflow [9–13], etc. These methods performed well in capturing crowd global motion information under normal

behavior scenes only. Unfortunately, no results are reported in the literature for abnormal behavior detection at very high-density crowd levels.

In this work, we solve divergence detection in the high-density crowd by directly capturing crowd global motion in form of images and learn crowd normal and divergent motion shapes through a neural network that predicts crowd behavior for the unknown scene. We also propose a novel divergence localization algorithm to pinpoint divergence location with the help of a bounding box. Finding a source of divergence can help to efficiently deploy crowd management staff right at the critical locations.

2 Related Work

Motion is one of the key ingredients in the crowd scene analysis and the success of the behavior prediction scheme greatly relies on the efficiency of the motion estimation (ME) method. Therefore, we provide a comprehensive review of ME techniques and the corresponding abnormal behavior detection methods with emphasis on their capabilities for 'high' density crowded scenarios. OPF is considered to be one of the most fundamental motion flow model [14–17] that has been widely employed for motion estimation [18, 19], crowd flow segmentation [20], behavior understanding [21–23] and tracking in the crowd [24]. However, OPF methods suffer from various problems like motion discontinuities, lack of spatial and temporal motion representation, variations in illumination conditions, severe clutter and occlusion, etc.

To overcome problems of OPF ME, researchers employ particle advection concepts from fluid dynamics into the computer vision domain [8] and obtain long-term "motion trajectories" under the influence of the OPF field. We et al. [7] employ chaotic invariants on Lagrangian trajectories to determined either the behavior of the crowd is normal or not. They also perform localization of anomaly by determining the source and size of an anomaly. Unfortunately, no results were reported for the high-density crowd. Similarly, Ali et al. [8] obtain Lagrangian Coherent Structures (LCS) from particle trajectories by integrating trajectories over a finite interval of time termed as Finite-Time Lyapunov exponent (FTLE). LCS appears as ridges and valleys in the FTLE field at the locations where different segments of the crowd behave differently. Authors perform crowd segmentation and instability detection in the high-density crowd using LCS in FTLE, however actual anomalies of the high-density crowd like crowd divergence, escape behavior detection, etc. are not performed. Similarly, authors in [10, 11] obtain particle trajectories using high accuracy variational model for crowd flow and perform crowd segmentation tasks only. Mehran et al. [9] obtain streakflow by spatial integration of streaklines that are extracted from particle trajectories. For anomaly detection, they decompose streakflow field into curl-free and divergence-free components using the Helmholtz decomposition theorem and observe variations in potential and streak functions used with SVM to detect anomalies like crowd divergence/convergence, escape behavior,
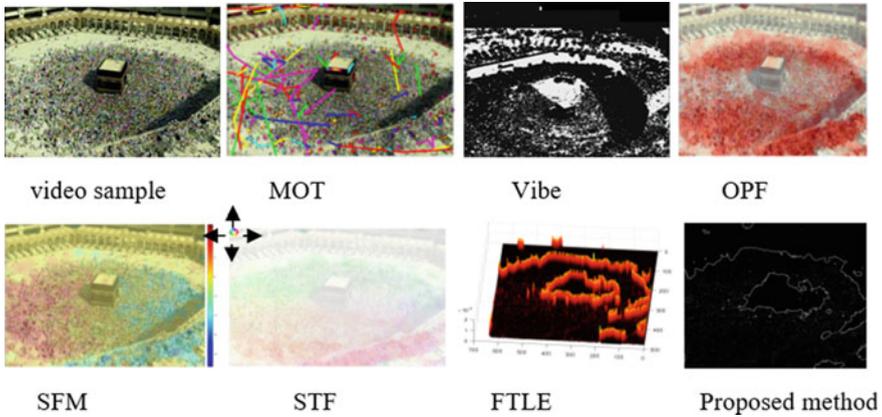
Fig. 2 High-density crowd motion estimation by state-of-the-art methods

etc. However, results are reported for anomaly detection and segmentation at low-density crowd and efficacy is still questionable for anomalies at the high-density crowd. Eduardo et al. [25] obtain long-range motion trajectories by using the farthest point seeding method called streamline diffusion on streamlines instead of spatial integration.

Behavior analysis is performed by linking short streamlines using Markov Random Field (MRF). However, only normal behavior detection and crowd segmentation results are reported. Although particle flow methods discussed above are better candidates for ME of the high-density crowd, but they are rarely employed for abnormal behavior detection at high density crowded scenes. Figure 2 provides a comparison of ME methods for high-density crowd performing Tawaf around Kabbah. Conventional object tracking based ME methods [26, 27] (Fig. 2b, c) works best at low crowd density but completely fails at high crowd density. The OPF method from Brox et al. [15] can estimate motion at high density but motion information is short-term. SFM [28] method can provide better motion estimation in low-density crowd areas but at high density, the performance of SFM also degrades. Streakflow [9] method also performs similarly to the SFM method at a high-density crowd. Unfortunately, all these methods are unable to provide a clean motion-shape for the crowd. FTLE method [9] (Fig. 2g) produce clear ridges at cowd boundaries and can be best to describe high-density crowd motion. Therefore, in this work, we utilize the FTLE method to obtain crowd motion-shape and translate it to a single channel greyscale image (Fig. 2h) for both normal and abnormal behavior analysis. Our framework for divergence detection is shown in Fig. 3 (top portion). It consists of two main phases: Phase 1: low-level FTLE feature extraction and conversion into a grey-scale motion shape image; Phase 2: behavior classification using a CNN. Motion shape images are also used for divergence localization process.
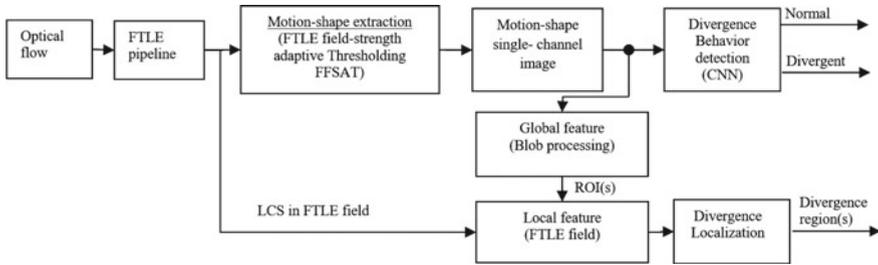
Fig. 3 Framework for divergence behavior detection in high-density crowd

3 Divergence Detection with Motion Shape and Deep Convolution Neural Network

3.1 Data Preparation

Due to the unavailability of a very high-density crowd dataset with divergence behavior, we generate synthetic data by simulating crowd in Massmotion software [29]. We model two crowd scenarios: Stampede at Loveparade 2010 and Tawaf around Kabbah. Example snapshots of normal and divergence crowd behaviors are shown in Figs. 4 and 5.

### 3.2 Global Motion Estimation and Shape Extraction

In this work, high-density crowd motion is computed by the Finite-Time Lyapunov Estimation (FTLE) method [8, 30]. Lagrangian Coherent Structure (LCS) appears as ridges in the FTLE field where two crowd segments behave differently. We extract LCS from FTLE field FTLE using the field-strength adaptive thresholding (FFSAT) scheme and convert it into a grey-scale image. At every integration step in the FTLE pipeline, maximum Eulerian distance (dmax) is calculated between LCS absolute peak value and average FTLE field strength, and a threshold (ffsat_thr) is set for dmax (65% in our work). LCS values crossing ffsat_thr are extracted and converted into a single-channel grey-scale image. FFSAT algorithm ensures only strong magnitude LCS values from the FTLE field are extracted and noise is filtered out.

3.3 Deep Network for Divergence Detection

A deep CNN network developed for normal and divergence classes in the high-density crowd is shown in Fig. 6.

a. Normal behavior

b. Diverging crowd

c. K13 view

d. K5 view

e. K12 view
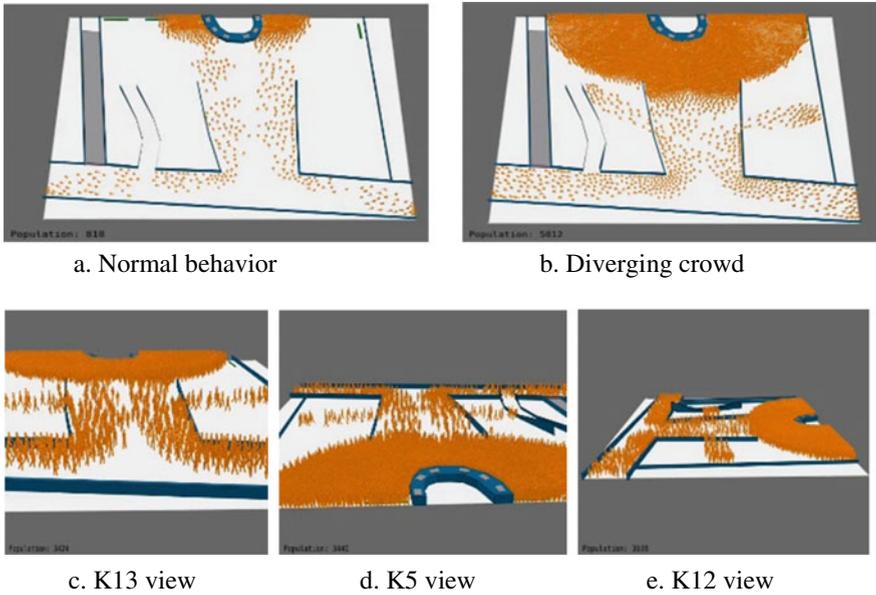
Fig. 4 Synthetic crowd data for Love parade 2010 disaster: **a**, **b** are camera-top views; **c**, **d**, **e**- same crowd with perspective views

a. Normal behavior

b. Diverging behavior
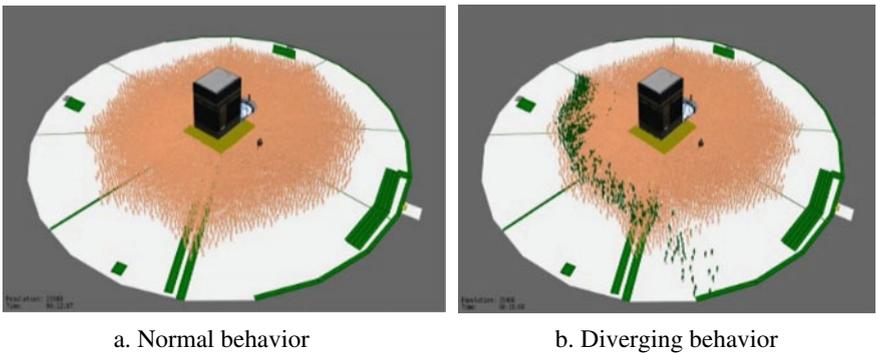
Fig. 5 Synthetic crowd data for Kabbah Tawaf—Normal behavior and divergent crowd

The greyscale image is first rescaled to 50×50 pixels at the input layer. A convolution layer is used (24 filters) with ReLU activation. The purpose of using a large number of convolution filters is to ensure all important receptive fields of CNN are excited about a given motion-shape. ReLU is adopted as the activation function because of its good performance for CNNs [31] and Max pooling is used for each 2×2 region. Finally, two fully connected layers are used and the softmax layer is used for the classification of normal or divergent behavior.
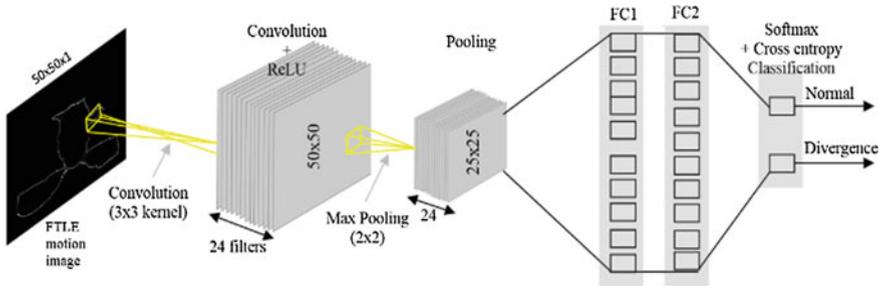
Fig. 6 Deep CNN for divergence behavior detection

3.4 Divergence Localization Algorithm

We propose a novel divergence localization algorithm that analyzes changes in motion shape blob to search for the region of divergence. It was noticed that motion-shape also exhibit undesired local variations (Fig. 7 top row) that could lead to false divergence region detection. These changes are occurred due to the to-and-fro motion experienced by the crowd at high densities crowd [32]. As these oscillatory motions propagate and reach the crowd boundary, the shape does not remain consistent in every frame. Whereas the initial occurrence of divergence also appears as a small shape change and progressively increases in size (as shown in Fig. 7 bottom row). To cater to undesired local shape changes, a blob processing pipeline is implemented shown in Fig. 8.
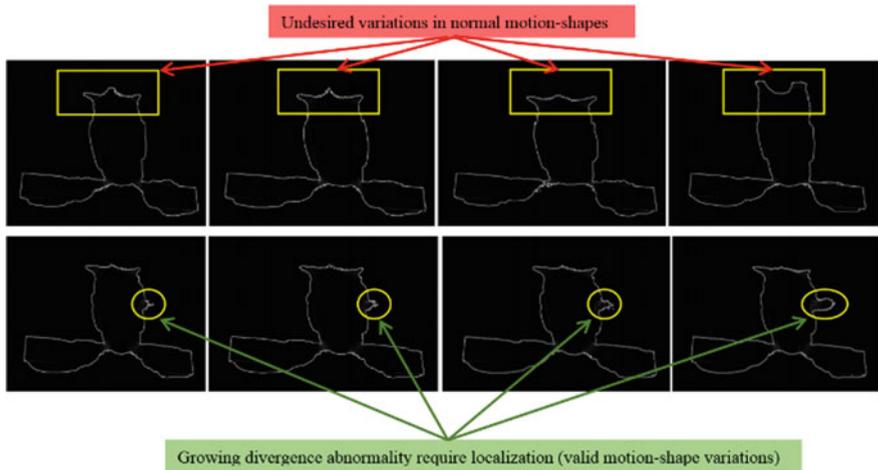

Fig. 7 Top row: Undesired motion-shape variations due to crowd oscillatory motion; Bottom row: Real shape change due to divergence

Fig. 8 Baseline blob extraction pipeline for normal and divergence behaviors

Baseline blob extraction pipeline extracts a baseline blob from the normal and divergence motion shapes and input to divergence localization algorithm is shown in Table 1. The divergence localization algorithm indicates divergence location(s) with the bounding box.

4 Experimentation Results

We evaluate proposed methods of crowd divergence behavior detection and divergence localization using crowd datasets of Love parade and Kabbah (data preparation details in Sect. 83.1). A detailed qualitative and quantitative analysis is provided for both methods on two selected scenarios. We also compare our methods with OPF from Brox et al. [15] by converting the OPF field in binary images.

4.1 Divergence Behavior Detection

For divergence behavior detection at two scenarios, the crowd is simulated to diverge from 25 different locations in each scenario and 1000 motion-shape images are captured (total images for 25 divergence locations $= 25 \times 1000 = 25,000$ divergence images for each scenario). One thousand images for each divergence location are generated to train CNN with minor local motion changes contributed by crowd oscillatory motion. Similarly, 2500 images are generated for normal crowd behavior. The dataset for each scenario is split into two parts: randomly 20 divergence locations data ($20 \times 1000 = 20,000$ images) are used for training/validation purposes, whereas the remaining random 5 divergence locations data (completely unseen to CNN) is used for prediction. Figure 9 provides a confusion matrix of divergence behavior detection for both scenarios and performance is compared with the OPF method.

For both the Love parade and Kabbah scenario, our method can achieve 100% accuracy. However, in both scenarios, OPF was able to detect approx. 50% of divergence behaviors only. Motion-shapes obtained through the OPF method are not as smooth and consistent as produced by our method; hence OPF performance degradation is evident.
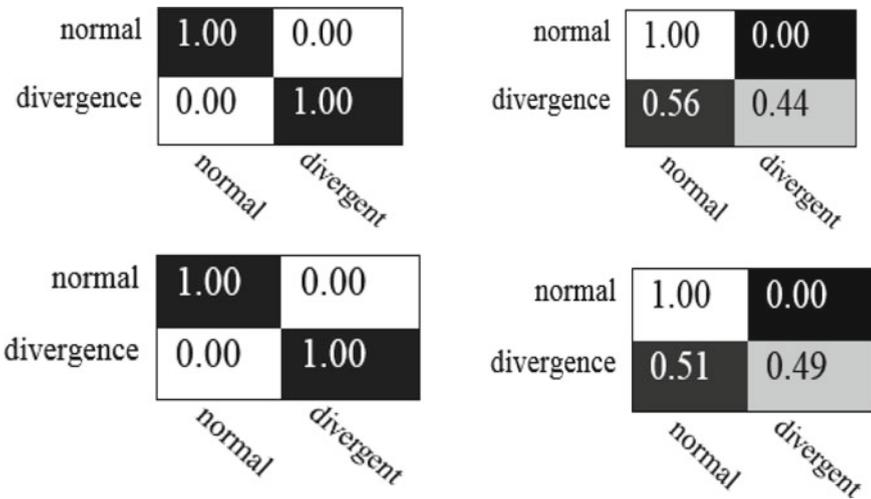
Table 1 Algorithm for divergence localization with bounding box

Algorithm: Divergence localization using the bounding box	
Init	**N**: Total video frames, **M**: Total divergence frames to analyze=50, **s_n**:No. of frames for spatial averaging =5, **t_n**: No. of frames for temporal averaging = 5, **st_thr**: FTLE field spatio-temporal threshold = 0.5, **blb_cnt**: number of small blobs to process = 5, **i_t**: transition frame index
Step 1	**Search normal to divergence-starting Transition frame (global features analysis)**
	FOR frames i=1:*N*
	-Obtain FTLE motion-shape images for two frames i(previous) and i+1(current),
	-Run blob processing pipeline (fig. 8) on previous and current frames FTLE images and obtain large single blobs for both frames
	-Subtract the current large blob from the previous large blob to obtain several smaller blobs
	-Sort blb_cnt small blobs according to the area in descending order
	-Blob shape analysis for ROI
	FOR blob_num=1: blb_cnt
	Compare blob shape with shape-list [square, rectangle, triangle, pentagon, hexagon, semi-circle or nearly circular]
	IF blob shape matches with any of shape-list
	Transition frame 'i_t' with the first ROI, Exit this loop and go to step 2
	ELSE
	Select next blob in area sorted list
	END
	-Increment frame number i
	END
Step 2	**Bounding box search for divergence location (global features analysis)**
	FOR frames j=i_t:i_t+M
	- Obtain FTLE motion-shape images for two frames j(previous) and j+1(current)
	- Run blob processing pipeline (fig. 8) on previous and current frames FTLE images and obtain large single blobs for both frames
	- Subtract current large blob from previous large blob to obtain several smaller blobs
	- Sort blb_cnt small blobs according to the area in descending order
	- **Blob shape analysis for ROI**
	FOR blob_num=1: blb_cnt
	IF blob shape matches with any of shape-list
	Mark this blob as possible ROI and update ROI_list
	ELSE
	Select next blob in area sorted list
	END
	- Jump to step 3 to obtain divergence location bounding box coordinates for all ROI of the current frame
	- Increment j
	END

(continued)

Table 1 (continued)

| Step 3	**FTLE field search for refined divergence location bounding box (local features analysis)**

	FOR ROI=1:length(ROI_list)
	   - **Spatial Averaging**: For each ROI boundary pixel, perform 8-connected spatial averaging of FTLE field (for all frames from j-s_n to j+s_n frames)
	   - **Temporal averaging:** For each ROI boundary pixel in j-t_n to j+t_n frames, perform temporal averaging.
	   - FOR all ROI boundary spatio-temporal averaged values
	          IF spatio-temporal averaging value >= st_thr,
	            keep pixel as valid
	          ELSE
	            discard pixel
	       END
	   - Obtain bounding box coordinates of current ROI by calculating maximum and minimum (x,y) coordinates from selected valid pixels list in above step
	END
	Return bounding box coordinates for all ROIs in ROI_list

Fig. 9 Confusion matrices for divergence behavior detection: Love parade scenario **a** Proposed method **b** OPF method; Kabbah scenario **c** Proposed method **d** OPF method

4.2 Divergence Localization

The performance of the divergence localization algorithm is evaluated by calculating the Intersection over Union (IoU) area of the predicted bounding box and ground truth bounding box for each divergence region. Ground truth bounding boxes are obtained by hand labeling divergence regions of each abnormal frame. IoU is calculated using Eq. (1).

$$\text{IoU} = \frac{\text{Area of overlap}}{\text{Area of union}} \tag{1}$$

Generally, an IoU score greater than 0.5 (50% overlap) is considered a good prediction by any bounding box (b. box) detection algorithm [33]. In this work, the IoU score is calculated for N post i_t frames. IoU score of six selective frames (out of N = 50 post i_t frames) for the Love parade scenario is shown in Fig. 10. The green color b.box represents ground truth and the red color b.box represents prediction by our algorithm. The Final IoU score is obtained by averaging N frames IoU scores. The average IoU score of our algorithm for the Love parade scenario is 0.501 (50% overlap). We also perform divergence region detection using OPF motion images. The average IoU score with the OPF method is found to be 0.15 (15% overlap) which proves our method performs well than OPF for divergence
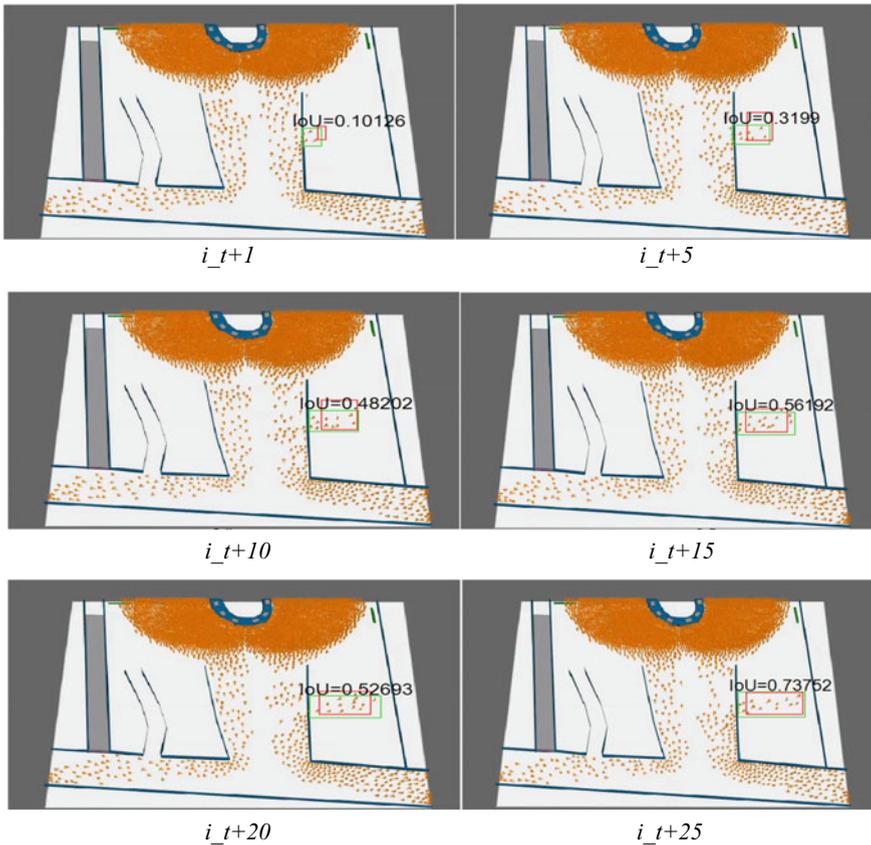
Fig. 10 Divergence region localization with our proposed method. IoU scores shown for six post *i_t* frames

region localization. Similarly, the average IoU score for the Kabbah scene with our algorithm is 0.63 (63% overlap) and 0.18 (18% overlap) for the OPF method.

5 Conclusion

In this work, we propose a deep CNN-based divergence behavior detection framework that extracts high-density crowd motion shapes in form of images to train deep CNN. Experimentation results show that the proposed method can achieve close to 100% accuracy for divergence detection in challenging Loveparade and Kabbah crowding scenarios. Similarly, a novel divergence region detection algorithm efficiently detects divergence regions with IoU of more than 50%. However, we notice there are few limitations of our proposed methodology of converting crowd motion into images using the FTLE method. Motion shape analysis is inefficient in the situations when a crowded segment in high density gets stationary due to any reason. Since there is no more movement at the stationary crowd segment, FTLE is unable to predict crowd motion shape at static crowd portions and results in incomplete or broken motion-shapes. Therefore, for our framework to work efficiently, the crowd needs to keep moving (for consistent motion-shape) that is always not true. Secondly, in the FTLE method, LCS ridges appear only at crowd boundaries, if any anomaly takes place at interior portions of the crowd (far from crowd boundaries towards the center), FTLE is unable to provide any information there. Therefore, in future work, we shall improve our method by incorporating spatial and temporal crowd density variations to capture static crowd behavior. And predict crowd behavior in all segments of the crowd, either crowd is stationary or in motion.

References

1. Helbing D, Mukerji P (2012) Crowd disasters as systemic failures: analysis of the Love Parade disaster. EPJ Data Sci 1(1):1–40. https://doi.org/10.1140/epjds7
2. Chen CY, Shao Y (2015) Crowd escape behavior detection and localization based on divergent centers. IEEE Sens J 15(4):2431–2439. https://doi.org/10.1109/JSEN.2014.2381260
3. Wu S, Wong HS, Yu Z (2014) A bayesian model for crowd escape behavior detection. IEEE Trans Circuits Syst Video Technol 24(1):85–98. https://doi.org/10.1109/TCSVT.2013.2276151
4. Andrade EL, Blunsden S, Fisher RB (2006) Modelling crowd scenes for event detection. Proc Int Conf Pattern Recognit 1:175–178. https://doi.org/10.1109/ICPR.2006.806
5. Courty T, Corpetti N (2007) Crowd Motion Capture. Comput Animat Virtual Worlds 18(September 2007):361–370
6. Nam Y, Hong S (2014) Real-time abnormal situation detection based on particle advection in crowded scenes. J Real-Time Image Process 10(4):771–784. https://doi.org/10.1007/s11554-014-0424-z
7. Wu S, Moore BE, Shah M (2010) Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 2054–2060

8. Ali S, Shah M (2007) A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: IEEE conference on computer vision and pattern recognition, pp 1–6
9. Mehran R, Moore BE, Shah M (2010) A streakline representation of flow in crowded scenes. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) LNCS 6313(PART 3):439–452. https://doi.org/10.1007/978-3-642-15558-1_32
10. Wang X, Gao M, He X, Wu X, Li Y (2014) An abnormal crowd behavior detection algorithm based on fluid mechanics. J Comput 9(5):1144–1149. https://doi.org/10.4304/jcp.9.5.1144-1149
11. Wang X, Yang X, He X, Teng Q, Gao M (2014) A high accuracy flow segmentation method in crowded scenes based on streakline. Opt Int J Light Electron Opt 125(3):924–929. https://doi.org/10.1016/j.ijleo.2013.07.166
12. Wang X, He X, Wu X, Xie C, Li Y (2016) A classification method based on streak flow for abnormal crowd behaviors. Opt Int J Light Electron Opt 127(4):2386–2392. https://doi.org/10.1016/j.ijleo.2015.08.081
13. Huang S, Huang D, Khuhro MA (2015) Crowd motion analysis based on social force graph with streak flow attribute. J Electr Comput Eng 2015. https://doi.org/10.1155/2015/492051
14. Horn BK, Schunck BG (1981) Determining optical flow. Artif Intell 17(1981):185–203
15. Brox T, Papenberg N, Weickert J (2004) High accuracy optical flow estimation based on a theory for warping. In: Computer Vision - ECCV 2004, vol 4, no May, pp 25–36. https://doi.org/10.1007/978-3-540-24673-2_3
16. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. Proc Imaging Underst Work 130:121–130
17. Fortun D, Bouthemy P, Kervrann C, Fortun D, Bouthemy P, Kervrann C (2015) Optical flow modeling and computation: a survey. Comput Vis Image Underst 134:1–21
18. Lawal IA, Poiesi F, Anguita D, Cavallaro A (2016) Support vector motion clustering. IEEE Trans Circuits Syst Video Technol X(X):1–1. https://doi.org/10.1109/TCSVT.2016.2580401
19. Cheriyadat AM, Radke RJ (2008) Detecting dominant motions in dense crowds. IEEE J Sel Top Signal Process 2(4):568–581. https://doi.org/10.1109/JSTSP.2008.2001306
20. Ali S, Shah M (2007) A lagrangian particle dynamics approach for crowd flow simulation and stability analysis
21. Hu MHM, Ali S, Shah M (2008) Learning motion patterns in crowded scenes using motion flow field. In: 2008 19th international conference pattern recognit, pp 2–6. https://doi.org/10.1109/ICPR.2008.4761183
22. Solmaz B, Moore BE, Shah M (2012) Identifying behaviors in crowd scenes using stability analysis for dynamical systems. IEEE Trans Pattern Anal Mach Intell 34:2064–2070. https://doi.org/10.1109/TPAMI.2012.123
23. Chen DY, Huang PC (2011) Motion-based unusual event detection in human crowds. J Vis Commun Image Represent 22(2):178–186. https://doi.org/10.1016/j.jvcir.2010.12.004
24. Hu W, Xiao X, Fu Z, Xie D, Tan T, Maybank S (2006) A system for learning statistical motion patterns. IEEE Trans Pattern Anal Mach Intell 28(9):1450–1464. https://doi.org/10.1109/TPAMI.2006.176
25. Pereira EM, Cardoso JS, Morla R (2016) Long-range trajectories from global and local motion representations. J Vis Commun Image Represent 40:265–287. https://doi.org/10.1016/j.jvcir.2016.06.020
26. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: EEE computer society conference on computer vision and pattern recognition (CVPR'05), pp 886–893
27. Barnich O, Van Droogenbroeck M (2011) ViBe: a universal background subtraction algorithm for video sequences. IEEE Trans Image Process 20(6):1709–1724. https://doi.org/10.1109/TIP.2010.2101613
28. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In: IEEE computer society conference on computer vision and pattern recognition work. CVPR Work 2:935–942. https://doi.org/10.1109/CVPRW.2009.5206641.
29. https://www.oasys-software.com

30. Shadden SC, Lekien F, Marsden JE (2005) Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. Phys D Nonlinear Phenom 212(3–4):271–304. https://doi.org/10.1016/j.physd.2005.10.007
31. Zeiler MD et al (2013) On rectified linear units for speech processing New York University, USA Google Inc ., USA University of Toronto , Canada. In: IEEE international conference on acoustic speech and signal processing (ICASSP 2013), pp 3–7
32. Krausz B, Bauckhage C (2012) Loveparade 2010: automatic video analysis of a crowd disaster. Comput Vis Image Underst 116(3):307–319. https://doi.org/10.1016/j.cviu.2011.08.006
33. Ahmed F, Tarlow D, Batra D (2015) Optimizing expected intersection-over-union with candidate-constrained CRFs. In: Proceedings of the IEEE international conference on computer vision, vol 2015, pp 1850–1858. https://doi.org/10.1109/ICCV.2015.215