



Violent Video Event Detection: A Local Optimal Oriented Pattern Based Approach

B. H Lohithashva[✉] and V. N Manjunath Aradhya[✉]

Department of Computer Applications, JSS Science and Technology University,
Mysuru, Karnataka, India
{lohithashva.bh, aradhya}@sjce.ac.in

Abstract. Video/Camera-based monitoring is a prominent and difficult research problem in the field of machine learning and pattern recognition and posed much interest in our safety in the private and public sectors. Therefore, surveillance cameras have been deploying to control suspicious activity. Consequently, many researchers have worked on developing an automatic surveillance system to detect violent events and assists security guards to take the right decision at the right time. Still, violent event detection is difficult to detect because of illumination, complex background, scale variation, blurriness, occlusion, and low resolution in a surveillance camera. In this paper, the Local Optimal Oriented Pattern (LOOP) texture-based feature descriptor is proposed. Eventually, eminent features are used with a support vector machine (SVM) classifier for violent event detection. Experiments are conducted on the Hockey Fight dataset and Violent-Flows dataset. The five-Fold Cross-Validation approach is used to analyze the performance of the proposed method. The data and results are promising and encouraging.

Keywords: Violent event · Texture features · Spatio-temporal · LOOP descriptor

1 Introduction

Video surveillance is often seen as the process of analyzing violent event scenes in the video. The actions conducted by humans can be analyzed with the help of a surveillance camera that can be manual or automated. An intelligent video surveillance system intends at detecting, tracking, and recognizing objects of interest and further analyzing and interpreting the video activities of the scenes, despite the substantial amount of videos collected by surveillance cameras. Nowadays, we see plenty of surveillance cameras being installed throughout the private and public sectors. The reason behind this is for the safety of human beings and also for the hardware equipment available in the markets at reasonable prices. As indicated by the way that visual information is generally accessible in surveillance systems, we focus on strategies used vision information [7]. The automatic

surveillance system reduces the risk of security persons monitoring prolonged videos. Violent event detection is a difficult action recognition task and it is a branch of computer vision. The pattern, facial expression, and actions to be detected unusual events in the video scenes. A terrorist attack, bomb detection, fraud detection, loitering, slip and fall event, and many more are action recognition problems [22]. Detection of the violent events is highly uncertain to resolve the difficult task. Once the system has experienced violent and non-violent events, we apparently gave the test label to detect the classification of events. This is a more challenging scenario when the normal event drastically alters and it is difficult to learn due to blurriness, variations in scale, complex background, occlusion, and illumination. In this work, we have used the LOOP descriptor to detect violent events in the video sequences.

The Contributions of the Paper are as Follows

- LOOP descriptor used to extract salient features to detect violent events
- Spatial-temporal post-processing approach is used to improve the accuracy of violent event detection.
- To evaluate the efficiency of the proposed method, Five-Fold Cross-Validation approach is used and results are compared with the state of the art techniques.

The rest of the paper is structured as follows. Section 2 is connected with the previous research work. The proposed texture-based descriptors are discussed in Sect. 3. Experimentation results are described in Sect. 4. Finally conclude the research paper in Sect. 5.

2 Previous Works

In recent years, the research community has established a survey on a variety of algorithms based on the handcrafted features [2,5], deep learning features [27,28,30,39] and classifiers [3,18] are used to resolve the major issues of violent event detection [29,32]. Quasim et al. [33] introduced the Histogram of Swarms (HOS) descriptor. The method used the variance of Optical Flow (OF) to extract spatio-temporal information in the sequence of video frames. Ant Colony Optimization (ACO) is used to cluster moving object and it separate salient and non-salient features, finally OF technique is used to extract prominent features to detect normal and violent events. Febin et al. [12] presented a combination of Motion Boundary Scale Invariant Features Transform (MoB-SIFT) and movement filter algorithm. The movement filter algorithm extracts temporal information features of the non-violent event and avoids the normal event. Furthermore, the combination of motion boundary, optical flow, and SIFT feature extract eminent features to detect violent events. Esen et al. [36] used Motion Co-Occurrence Feature (MCF) to detect abnormal events in the video. The method used a block matching algorithm to extract the direction and magnitude of motion features and fed it to the KNN classifier to categorize normal and

abnormal events. Recently, Lohithashva et al. [23] introduced the integration of texture features to detect violent activity. The method extracts prominent texture features to detect suspicious activity. Song et al. [37] introduced the fusion of multi-temporal analysis and multi-temporal perceptron layers to detect unusual events. Zhang et al. [41] presented an entropy model to measure the distribution of enthalpy for abnormal event detection. The authors have used an enthalpy model in the micro point of view to describe crowd energy information. Ryan et al. [35] proposed optical flow and Gray Level Co-occurrence Matrix (GLCM) feature descriptor to detect abnormal events in the video sequence. Lloyd et al. [21] proposed a GLCM texture feature descriptor detect non-violent and violent activity detection. Pujol et al. [10] described events based on features fusion extraction technique of local eccentricity which includes the combination of Fast Fourier Transform, radon transform, projection, and ellipse eccentricity. Deepak et al. [9] introduced the extraction of both spatio-temporal information from texture based feature descriptor. The method extracts local geometric characteristics such as gradients and curvatures which are basic space-time movement properties used to detect normal and abnormal events. Li et al. [20] introduced OF based feature descriptor to detect violent events in the video scene. Initially, they have used background subtraction to remove low variation and noise in the frame and extract the Histogram of Maximal Optical Flow Projection (HMOFP) features. Reconstruction cost (RC) is used to detect violent events in video scenes.

Imran et al. [16] introduced a deep learning method to detect a violent event in surveillance video. MobileNet is used to extract spatio-temporal information from the moving objects after that dominant features are given to a gated recurrent unit (GRU) to detect suspicious events in the video scene. Hason et al. [14] introduced spatiotemporal information using a Spatiotemporal Encoder, Bidirectional Convolution Long Short Term Memory (BCLSTM) deep-learning feature extraction technique to detect unusual events in the video sequence. Asad et al. [4] presented violent event detection based on the spatio-temporal features from a video's uniformly spaced sequential frames. Multi-level processes for two consecutive frames, obtained from the top and bottom layers of the convolutional layers neural network, are integrated using optimized feature fusion strategy, finally, features are fed to Long short term memory (LSTM) to distinguish between violent and non-violent event. Sabokrou et al. [11] introduced Fully Convolution Neural Networks (FCNs) to detect and localize violent events in a sequence of video. Accatolli et al. [1] introduced a 3D-CNN to detect suspicious activity in video. CNN architecture extracts salient features without any prior knowledge and fed them to the SVM classifier to segregate violent and non-violent events. Zhou et al. [36] applied hybrid auto-encoder architecture to extract spatio-temporal features from the crowd and discriminate normal and abnormal events in video frames. Song et al. [38] introduced a modified 3D-CNN to detect an aggressive incidence throughout the video. The method is used a uniform sampling method to reduce the redundancy and conquest the motion coherence and they have illustrated the efficacy of the sampling method.

3 Proposed Methodology

We demonstrate an overview of the proposed approach in this section. The LOOP descriptor extracts prominent texture features from the input video and fed them to the SVM classifier to detect violent events. Figure 1 shows the workflow of violent event detection using the proposed LOOP descriptor. The approach suggested in the sections that follow illustrates the detection of violent events.

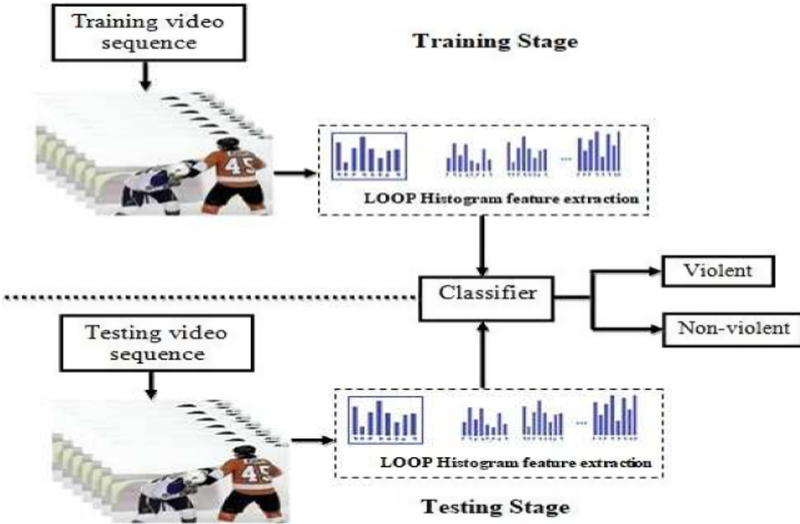


Fig. 1. The description of the proposed method

3.1 LOOP Feature Descriptor

LOOP [6] is a scale and rotational invariance texture-based feature descriptor. To overcome the drawback of previous binary descriptors the LOOP descriptor has used and it is an upgrade of the Local Binary Pattern (LBP) and Local Directional Pattern (LDP) descriptors. Consider p_c be the intensity of the frame F at pixel (a_c, b_c) and $p_n (n = 0, 1, \dots, n - 1)$ and pixel intensity of 3×3 neighborhood of (a_c, b_c) except for the middle pixel p_c . The eight Kirsch masks used previously for the LDP [17] are located in the direction of these eight adjacent pixels $p_n (n = 0, 1, \dots, n - 1)$. Therefore, it provides a measure of the severity of the degree of variability in the direction, separately. The Kirsch eight directions mask as shown in Fig. 2.

The eight respondents of the Kirsch masks are k_n response to the pixels of the intensity $p_n (n = 0, 1, \dots, n - 1)$. Each pixel is assigned an exponential e_n by the size of k_n output of eight Kirsch masks.

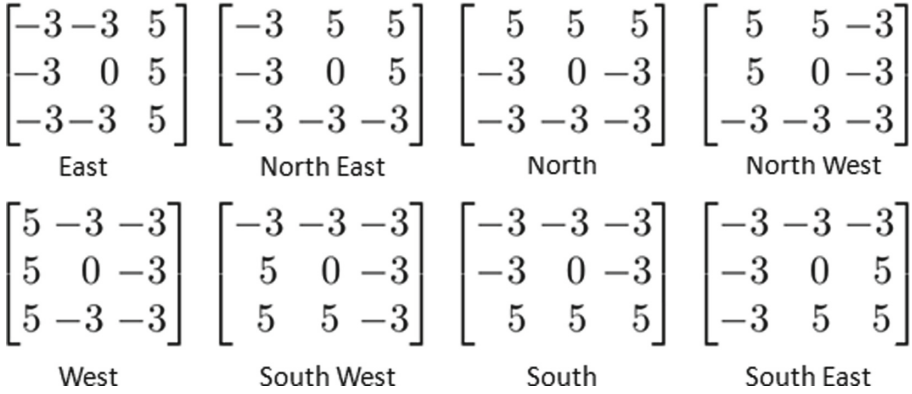


Fig. 2. Kirsch eight directions mask

$$LOOP(a_c, b_c) = \sum_{n=0}^{n-1} s(p_n - p_c) * 2^{k_n} \tag{1}$$

$$s(a) = \begin{cases} 1 & \text{if } a \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The LOOP outcome about the pixel (a_c, b_c) is stated as in (1 & 2) and $s(a)$ represents neighborhood pixels intensity values. Therefore, the LOOP descriptor computes the rotational invariance in the major method. Eventually, pixel intensities are evaluated over the cell at each number that has prominently featured. This descriptor is measured as a $2^8 = 256$ dimensional features for each frame.

3.2 Classification Based on Support Vector Machine (SVM)

SVM [8] is a binary classification approach which is widely used in regression and classification applications. Initially, SVM is introduced for classification and regression and subsequent kernel methodologies are used to implement non-linear classification by processing input information via a high-dimensional feature space. SVM attempts to optimize the distance of the distinguishing borderline among violent and non-violent events by trying to maximize the distance of the separating plane from each of the features. In the binary classification problem, data from a two-class are considered. In our research work, the Gaussian kernel function in SVM is used to violent video scene.

3.3 Post-processing

The post-processing technique [34] significantly increases the accuracy and reduces the false-positive rate. In this work, for the post-processing technique,

we have taken 30 frames for the detection of frames which significantly improves the performance.

4 Experiment Results and Discussion

In this section, we summarize the detailed experimentation study to evaluate the use of violent event detection approaches in two standard benchmark datasets. Thereafter, the experimentation parameter setting is explained. Finally, the results obtained are compared with the existing feature descriptors.

4.1 Violent Datasets

The Hockey Fight (HF) dataset and Violent-Flows (VF) dataset experimentation are conducted to demonstrate the effectiveness of the proposed method and both datasets have complex backgrounds, illumination, blurriness, scale changes, and occlusion. This dataset comprises 1000 action videos of the National Hockey League (NHL) (500 fights and 500 no-fights), initially used to distinguish violent event detection processes [31]. For each clip, there have been battles to fight between two or hardly any hockey players. Each video clip is approximately equal to 1.75 s.



Fig. 3. Violent datasets sample frames. First row: Normal scenes, Second row: Fight scenes.

The Violent-Flows dataset contains 246 action videos (123 fights and 123 no-fights). Maximum possible people to seeing aggressive events that occurred inside the football ground during the match. This dataset is used to assess the detection of violent events [15]. All violent videos in the angered circumstances, each video is roughly equivalent to 3.5 s. Figure 3 illustrates the following frame sequences comprising Hockey Fight and Violent-Flows dataset sample frames of fights and no fight scenes.

4.2 Experimental Setting

In this section, we have used a Five-fold cross-validation test. We have compared our experimental results with existing methods using Hockey Fight and Violent-Flows dataset. Therefore, five different divisions were partitioned into each dataset: four for training and one for evaluation testing. The average accuracy result is estimated each time and the Precision (P), Recall (R), F-measure (F), Accuracy (Acc), and Area Under Curve (AUC) have used as an evaluation method. we employed an SVM classifier with a Gaussian kernel function to differentiate violent and non-violent events in the video sequences.

4.3 Result

In the experiment, we have used the LOOP descriptor to demonstrate for detection of unusual events in the video sequence. Our proposed method shows impressive results compared to existing methods. HF dataset ROC curves with SVM classifier using LOOP descriptor is compared with the existing methods as shown in Fig. 4. The Precision of 94.48%, Recall of 94.09%, F-measure of 94.28%, the accuracy of 92.25%, and AUC of 95.11% as illustrated in Table 1. VF dataset ROC curves SVM classifier using LOOP descriptor compared with the previous methods as shown in Fig. 5. The obtained Precision, Recall, F-measure, Accuracy, and AUC result are successively, 95.64%, 93.38%, 95.17%, 91.54%, and 93.81% on the Violent-Flows dataset as shown in Table 1. Comparative analysis of the proposed method for HF and VF Datasets as shown in Fig. 6. It is noticed that our proposed feature descriptor is capable to detect violent events even if there is a cluttered background, varied illumination, little motion, and scale changes.

Table 1. Performance evaluation metrics is illustrated in percentage

Dataset	P	R	F	Acc	AUC
Hockey Fight	94.48	94.09	94.28	92.25±1.77	95.11
Violent-Flows	95.64	93.83	95.17	91.54±3.42	93.84

4.4 Discussion

Our proposed LOOP descriptor gives good result than Histograms of Oriented Gradients (HOG), Histogram of Optical Flow (HOF), Local Ternary Pattern (LTP), Violent Flow (ViF), Oriented Violent Flow (OViF), ViF+OViF, Distribution of Magnitude Orientation Local Interest Frame (DiMOLIF), GHOG+GIST, LBP+GLCM and Histogram of Optical flow Magnitude and Orientation (HOMO) for both HF and VF datasets. HOG, HOF, LTP, and ViF

descriptors can not work if orientation changes. Therefore, these feature extraction methods are failed to detect violent event detection. The OViF feature extraction method extracts orientation features and obtains good performance for the HF dataset but does not perform well for the VF dataset. To resolve this problem the ViF+OViF feature extraction technique is used to extract both magnitude and orientation features to detect suspicious behavior and is superior to ViF and OViF descriptors. DiMOLIF descriptor extracts magnitude and orientation from the optical flow feature descriptor to detect violent events. This descriptor gives substantial results as compared to ViF and OViF. The GHOG+GIST descriptor uses the fusion of global gradient and texture features. GHOG descriptor is poorly performed if there is a cluttered background and the GIST descriptor does not work for violent crowd activity in the video sequence. LBP+GLCM descriptor uses the fusion of texture features to detect aggressive behavior. The main drawback of LBP is the arbitrarily defined set of binary weights that depend on direction. GLCM feature extraction limitations are the high dimensional of the matrix and the high correlation of the features. HOMO is based on multiple scaling factors being applied to the magnitude and orientation variations of the optical flow. LOOP descriptor is effective for illumination changes, scale, and rotational invariance.

Table 2. Performance comparison result on Hockey Fight and Violent-Flow dataset

Method	HF dataset		VF dataset	
	Acc (\pm SD)	AUC	Acc (\pm SD)	AUC
HOG [19]	87.8	-	57.43 \pm 0.37	61.82
HOF [19]	83.5	-	58.53 \pm 0.32	57.60
LTP [40]	71.90 \pm 0.49	-	71.53 \pm 0.17	79.86
ViF [15]	81.60 \pm 0.22	88.01	81.20 \pm 1.79	88.04
OViF [13]	84.20 \pm 3.33	90.32	76.80 \pm 3.90	80.47
ViF+OViF [13]	86.30 \pm 1.57	91.93	86.00 \pm 1.41	91.82
DiMOLIF [25]	88.6 \pm 1.2	93.23	85.83 \pm 4.2	89.25
GHOG+GIST [24]	91.18 \pm 2.95	93.45	88.86 \pm 5.12	92.00
LBP+GLCM [23]	91.51 \pm 1.51	93.60	89.06 \pm 3.32	93.00
HOMO [26]	89.3 \pm 0.91	95.18	76.83 \pm 1.76	82.84
LOOP	92.25\pm3.13	95.19	91.54\pm1.32	93.84

We have demonstrated the efficiency of our proposed model and this is an immensely important task. We compare our experimental results with existing methods using HF and VF datasets. In the experiment, we have used the LOOP descriptor to demonstrate for violent event detection. Our proposed method shows impressive results compared to existing methods as illustrate in Table 2.

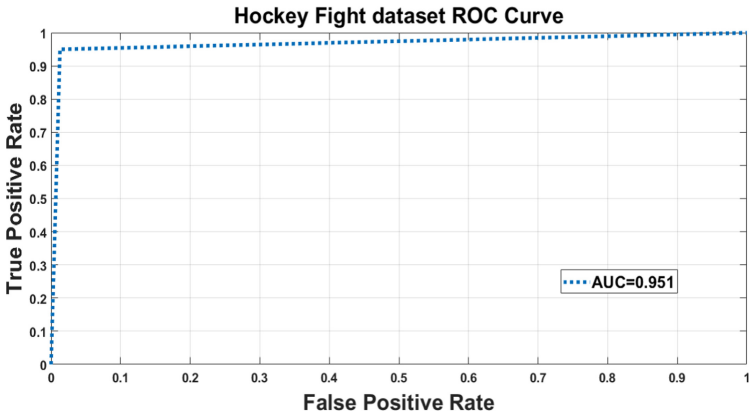


Fig. 4. Hockey Fight dataset ROC Curves of proposed feature descriptor with SVM classifier

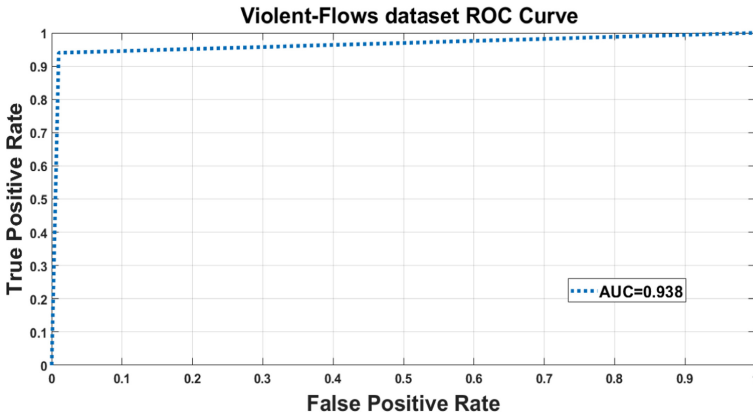


Fig. 5. Violent-Flows dataset ROC Curves of proposed feature descriptor with SVM classifier

It is noticed that our proposed feature descriptor is capable to detect violent event even if there is a cluttered background, varied illumination, little motion, and scale changes. Actually, there are six attributes that need to be intimate for suspicious event detection. Some of the intimates are, magnitude, orientation, the spatial arrangement of the moving objects, number of the objects moving in a video scene, mass, and acceleration. Certainly, our proposed method based on the scale and orientation of the object apparent motion using the extraction of LOOP features to improve the performance of the proposed method. Eventually, we deduce that our proposed method significantly performs well for both Hockey Fight and Violent-Flows dataset.

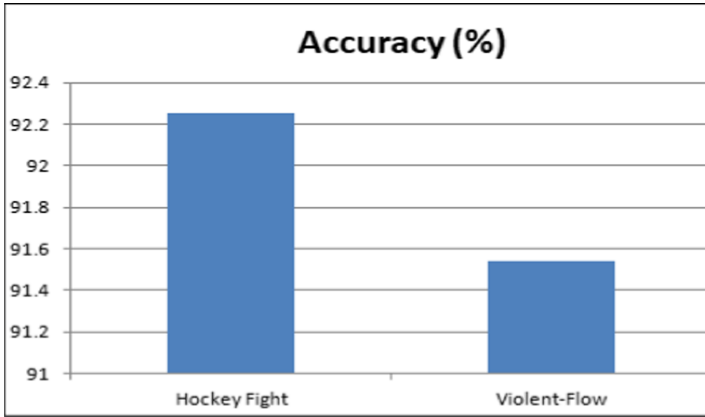


Fig. 6. Comparative analysis of the proposed method for Hockey Fight and Violent Flows Dataset

5 Conclusion

Video monitoring is used as a mechanism of scrutinizing videos to recognize suspicious behavior. Human behavior can be examined with the help of a surveillance video that could be manual or automatic. The research community has failed to develop an effective algorithm because of complex background, illumination, scale changes, etc. Experiments are conducted on the HF dataset and VF dataset and the experimental result shows that our proposed method performs an effective and preferable result to the previous feature descriptors. In the future, we intent to conduct experimentation on complex videos, endeavor to optimize the proposed method to improve the accuracy and reduce the time computation.

Acknowledgment. The first author would like to thank UGC under RGNF for supporting financially, Letter no.F1-17.1/2014-15/RGNF-2014-15-SC-KAR-73791 /(SA-III/Website), JSS Science and Technology University, Mysuru, Karnataka, India.

References

1. Accattoli, S., Sernani, P., Falcionelli, N., Mekuria, D.N., Dragoni, A.F.: Violence detection in videos by combining 3d convolutional neural networks and support vector machines. *Appl. Artif. Intell.* **34**(4), 329–344 (2020)
2. Aradhya, V.M., Basavaraju, H., Guru, D.S.: Decade research on text detection in images/videos: a review. *Evolut. Intell.* **14**, 1–27 (2019)
3. Aradhya, V.M., Mahmud, M., Guru, D., Agarwal, B., Kaiser, M.S.: One-shot cluster-based approach for the detection of covid-19 from chest x-ray images. *Cognit. Comput.* **22**, 1–9 (2021)
4. Asad, M., Yang, J., He, J., Shamsolmoali, P., He, X.: Multi-frame feature-fusion-based model for violence detection. *Vis. Comput.* **37**(6), 1415–1431 (2020)

5. Basavaraju, H., Aradhya, V.M., Pavithra, M., Guru, D., Bhateja, V.: Arbitrary oriented multilingual text detection and segmentation using level set and gaussian mixture model. *Evolut. Intell.* **14**(2), 881–894 (2020)
6. Chakraborti, T., McCane, B., Mills, S., Pal, U.: Loop descriptor: local optimal-oriented pattern. *IEEE Signal Process. Lett.* **25**(5), 635–639 (2018)
7. Cong, Y., Yuan, J., Liu, J.: Sparse reconstruction cost for abnormal event detection. In: *CVPR 2011*, pp. 3449–3456. IEEE (2011)
8. Vapnik, V., Cortes, C.: Support vector machine. *Mach. Learn.* **20**(3), 273–297 (1995)
9. Deepak, K., Vignesh, L., Chandrakala, S.: Autocorrelation of gradients based violence detection in surveillance videos. *ICT Express* **6**(3), 155–159 (2020)
10. Denoeux, T.: A k-nearest neighbor classification rule based on dempster-shafer theory, vol. 25, pp. 804–813. IEEE (1995)
11. Esen, E., Arabaci, M.A., Soysal, M.: Fight detection in surveillance videos. In: 2013 11th International Workshop on Content-Based Multimedia Indexing (CBMI), pp. 131–135. IEEE (2013)
12. Febin, I., Jayasree, K., Joy, P.T.: Violence detection in videos for an intelligent surveillance system using MoBSIFT and movement filtering algorithm. *Pattern Anal. Appl.* **23**, 611–623 (2020)
13. Gao, Y., Liu, H., Sun, X., Wang, C., Liu, Y.: Violence detection using oriented violent flows. *Image Vis. Comput.* **48**, 37–41 (2016)
14. Hanson, A., Pnvr, K., Krishnagopal, S., Davis, L.: Bidirectional convolutional lstm for the detection of violence in videos. In: *Proceedings of the European Conference on Computer Vision (ECCV)* (2018)
15. Hassner, T., Itcher, Y., Kliper-Gross, O.: Violent flows: real-time detection of violent crowd behavior. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–6. IEEE (2012)
16. Imran, J., Raman, B., Rajput, A.S.: Robust, efficient and privacy-preserving violent activity recognition in videos. In: *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pp. 2081–2088 (2020)
17. Jabid, T., Kabir, M.H., Chae, O.: Local directional pattern (LDP)-a robust image descriptor for object recognition. In: 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 482–487. IEEE (2010)
18. Kaiser, M.S., et al.: iworksafe: towards healthy workplaces during Covid-19 with an intelligent phealth app for industrial settings. *IEEE Access* **9**, 13814–13828 (2021)
19. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2008)
20. Li, A., Miao, Z., Cen, Y., Zhang, X.P., Zhang, L., Chen, S.: Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning. *Pattern Recognit.* **108**, 107355 (2020)
21. Lloyd, K., Rosin, P.L., Marshall, D., Moore, S.C.: Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. *Mach. Vis. Appl.* **28**(3-4), 361–371 (2017)
22. Lohithashva, B.H., Manjunath Aradhya, V.N., Basavaraju, H.T., Harish, B.S.: Unusual crowd event detection: an approach using probabilistic neural network. In: Satapathy, S.C., Bhateja, V., Somanah, R., Yang, X.-S., Senkerik, R. (eds.) *Information Systems Design and Intelligent Applications. AISC*, vol. 862, pp. 533–542. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-3329-3_50

23. Lohithashva, B., Aradhya, V.M., Guru, D.: Violent video event detection based on integrated LBP and GLCM texture features. *Rev. d'Intell. Artif.* **34**(2), 179–187 (2020)
24. Lohithashva, B.H., Manjunath Aradhya, V.N., Guru, D.S.: Violent event detection: an approach using fusion GHOG-GIST descriptor. In: Komanapalli, V.L.N., Sivakumaran, N., Hampannavar, S. (eds.) *Advances in Automation, Signal Processing, Instrumentation, and Control*. LNEE, vol. 700, pp. 881–890. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-8221-9_82
25. Mabrouk, A.B., Zagrouba, E.: Spatio-temporal feature using optical flow based distribution for violence detection. *Pattern Recognit. Lett.* **92**, 62–67 (2017)
26. Mahmoodi, J., Salajeghe, A.: A classification method based on optical flow for violence detection. *Expert Syst. Appl.* **127**, 121–127 (2019)
27. Mahmud, M., Kaiser, M.S., McGinnity, T.M., Hussain, A.: Deep learning in mining biological data. *Cognit. Comput.* **13**, 1–33 (2021)
28. Mahmud, M., Kaiser, M.S., Hussain, A., Vassanelli, S.: Applications of deep learning and reinforcement learning to biological data. *IEEE Trans. Neural Netw. Learn. Syst.* **29**, 2063–2079 (2018)
29. Majumder, S., Kehtarnavaz, N.: A review of real-time human action recognition involving vision sensing. In: *Real-Time Image Processing and Deep Learning 2021*. vol. 11736, p. 117360A. International Society for Optics and Photonics (2021)
30. Naveena, C., Poornachandra, S., Manjunath Aradhya, V.N.: Segmentation of brain tumor tissues in multi-channel MRI using convolutional neural networks. In: Mahmud, M., Vassanelli, S., Kaiser, M.S., Zhong, N. (eds.) *BI 2020*. LNCS (LNAI), vol. 12241, pp. 128–137. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59277-6_12
31. Bermejo Nievas, E., Deniz Suarez, O., Bueno García, G., Sukthankar, R.: Violence detection in video using computer vision techniques. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) *CAIP 2011*. LNCS, vol. 6855, pp. 332–339. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23678-5_39
32. Pareek, P., Thakkar, A.: A survey on video-based human action recognition: recent updates, datasets, challenges, and applications. *Artif. Intell. Rev.* **54**, 2259–2322 (2021)
33. Qasim, T., Bhatti, N.: A hybrid swarm intelligence based approach for abnormal event detection in crowded environments. *Pattern Recognit. Lett.* **128**, 220–225 (2019)
34. Reddy, V., Sanderson, C., Lovell, B.C.: Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture, pp. 55–61. *IEEE* (2011)
35. Ryan, D., Denman, S., Fookes, C., Sridharan, S.: Textures of optical flow for real-time anomaly detection in crowds. In: *2011 8th IEEE international conference on advanced video and signal based surveillance (AVSS)*, pp. 230–235. *IEEE* (2011)
36. Sabokrou, M., Fayyaz, M., Fathy, M., Moayed, Z., Klette, R.: Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes. *Comput. Vis. Image Underst.* **172**, 88–97 (2018)
37. Song, D., Kim, C., Park, S.K.: A multi-temporal framework for high-level activity analysis: violent event detection in visual surveillance. *Inf. Sci.* **447**, 83–103 (2018)
38. Song, W., Zhang, D., Zhao, X., Yu, J., Zheng, R., Wang, A.: A novel violent video detection scheme based on modified 3d convolutional neural networks. *IEEE Access* **7**, 39172–39179 (2019)

39. Ye, L., Liu, T., Han, T., Ferdinando, H., Seppänen, T., Alasaarela, E.: Campus violence detection based on artificial intelligent interpretation of surveillance video sequences. *Remote Sens.* **13**(4), 628 (2021)
40. Yeffet, L., Wolf, L.: Local trinary patterns for human action recognition. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 492–497. IEEE (2009)
41. Zhang, X., Shu, X., He, Z.: Crowd panic state detection using entropy of the distribution of enthalpy. *Phys. A Stat. Mech. Appl.* **525**, 935–945 (2019)