IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 34, No. 2, April, 2020, pp. 179-187
Journal homepage: http://iieta.org/journals/ria

Violent Video Event Detection Based on Integrated LBP and GLCM Texture Features

B.H. Lohithashva[1*], V.N. Manjunath Aradhya[1], D.S. Guru[2]

[1] Department of Computer Applications, JSS Science and Technology University, Mysuru 570017, Karnataka, India
[2] Department of Studies in Computer Science, University of Mysore, Mysuru 570005, Karnataka, India

Corresponding Author Email: lohithashva.bh@sjce.ac.in

https://doi.org/10.18280/ria.340208

Received: 15 November 2019
Accepted: 10 January 2020

Keywords:
features fusion, GLCM, LBP, optical flow, spatio-temporal interest points, texture features, videos, violent event

ABSTRACT

Violent event detection is an interesting research problem and it is a branch of action recognition and computer vision. The detection of violent events is significant for both the public and private sectors. The automatic surveillance system is more attractive and interesting because of its wide range of applications in abnormal event detection. Since many years researchers were worked on violent activity detection and they have proposed different feature descriptors on both vision and acoustic technology. Challenges still exist due to illumination, complex background, scale changes, sudden variation, and slow-motion in videos. Consequently, violent event detection is based on the texture features of the frames in both crowded and uncrowned scenarios. Our proposed method used Local Binary Pattern (LBP) and GLCM (Gray Level Co-occurrence Matrix) as feature descriptors for the detection of a violent event. Finally, prominent features are used with five different supervised classifiers. The proposed feature extraction technique used Hockey Fight (HF) and Violent Flows (VF) two standard benchmark datasets for the experimentation.

1. INTRODUCTION

In the modern era, technology has been improving day by day; we are very much fascinated about private and public sector's safety. Nowadays, we have seen numerous video surveillance cameras being deployed in both private and public sectors such as schools, colleges, museums, traffic control, hospitals, airports, railway stations, etc. Mainly because hardware equipment's are available at reasonable prices in the markets. Violent event detection is a prominent research field in action recognition's and it is a branch of computer vision. In the early days manually monitoring abnormal activities in the video. However, it was difficult to detect abnormal activity because of poor quality of information and prolonged videos. The automatic surveillance system is more attractive and interesting in unusual events detection they are, intrusion, loitering, slip and fall event, unattended event, fraud event and action recognition, etc. [1]. Because of the outstanding importance of safety in our daily life and provide reliable security to mankind.

Lohithashva et al. [1] used the Histogram of optical flow orientation (HOFO) optical flow feature descriptor. The authors used Horn-Schunck optical flow to detect the location of the moving crowd peoples. The method extracts magnitude and orientation features and fed into the Probabilistic Neural Network (PNN) to the classification of the normal and abnormal event. Mahadevan et al. [2] used the mixture of dynamic texture (MDT) to detect abnormal events in the crowded scene. The authors identified abnormal events based on the joint modeling of appearance, spatial, and temporal information. Spatial information extracted using discriminate saliency and temporal information extracted based on the low-probability. Lloyd et al. [3] proposed a GLCM texture feature descriptor, the method is used temporal changes of gray level features and inter-frame uniformity, it shows the variation between normal and abnormal event based on the local variations of the spatial relationship among the pixels. Riberio et al. [4] introduced the detection of violent and non-violent using the Rotation Invariant Motion Coherence (RIMOC) feature descriptor, which is based on a Histogram of Oriented Flows (HOF). The method extracts the discriminate and unstructured motion, based on the Eigen values present in the optical flow vectors from the spatio-temporal information and finally combined into a spheric Riemannian manifold. Zhang et al. [5] introduced violent event detection based on detection and localization. Gaussian model of optical flow (GMOF) detected the location of the moving objects; oriented histogram of optical flow (OHOF) feature extraction technique used Lucas-Kanade optical flow to extract magnitude orientation of the moving objects, based on the magnitude orientation authors detected the violent events in the video scenes. Deniz et al. [6] proposed violent event detection established on the acceleration patterns as the discriminating features. The accelerations calculated by employing the Radon transform. Gao et al. [7] presented violent event detection in the public datasets by using the Oriented Violent Flow (OViF) which provides information on statistical motion orientation with respect to motion magnitude change. Ullah et al. [8] used a 3-dimensional convolution neural network (3D-CNN) for feature extraction. After that, the features are passed to a softmax classifier to detect the violent event. Febin et al. [9] introduced a fusion of SIFT, optical flow, and gradient features. The fusion of three features called the MoBSIFT descriptor used to detect violent and non-violent events. Recently, Lohithashva et al. [10] introduced gradient and texture-based feature descriptors. The fusion features descriptor extracted prominent features and fed to support vector machine (SVM) classifier to detect crowded and uncrowded violent event

scenes in the video. Recently, some of the surveys [11-13] have published different feature extraction techniques used to detect violent events in videos. The most existing methods based on the spatio-temporal interest points [8], features fusion [9, 10], optical flow [14], textures [15, 16], trajectories [17], descriptors and deep learning techniques [18]. Some of the researchers have also been focusing on effective segmentation [19], subspace techniques [20, 21], and classifiers [1, 10, 22] used to detect violent events. Although, they are facing difficulty due to complex background, illuminations, scale variation, slow motion in video surveillance. The LBP descriptor is employed to analyze the local patterns in the given video sequence. The local patterns roles important to analyze violent event detection in grayscale changes and low resolution and the GLCM descriptor measures local variations of the spatial relationship among the pixels, and calculated at any angle or direction changes of the frame. The occurrence of local variation is more in a violent event. Hence, in this paper, we propose a new texture features fusion descriptor to increase the performance of the model for the violent event detection in video surveillance.

The contributions of the paper are as follows:

1. An integration of LBP and GLCM descriptor extracts the prominent features effectively to detect the violent and non-violent event.
2. Explored the different classifiers show to the effectiveness of categorization of violent and non-violent event.
3. Post-processing technique which has improved the true positive rate.

The rest of the paper is presented as follows. Section 2 delineates the proposed method, experimental result and discussion presented in section 3. Finally, conclusion is presented in section 4.

2. PROPOSED METHODOLOGY

In this section, we work on the task of detection of violent and non-violent for which we use Hockey Fight [17] and Violent-Flow dataset [23]. We propose the method based on the texture features descriptors. In experimentation, we have LBP, GLCM and features fusion descriptor for the detection of violent events. Work flow of the proposed method as mentioned in Figure 1. Five supervised Classifiers, Decision Tree (DT), Discriminant Analysis (DA), Logistic Regression (LR), SVM and K-Nearest Neighbour (KNN) are used for categorization of violent and non-violent events. The proposed method illustrates detection of violent and non-violent event in the below-mentioned sections.

2.1 LBP features descriptor

LBP is a global texture-based features descriptor introduced by Ojala in 2002 [24] and LBP features labels the pixel value by threshold the neighborhood of each pixel and analyze the outcome as binary numbers. There are different forms of LBP which are briefly explained by Bouwmans et al [25]; the feature descriptor algorithm allows better handling the scale changes and occlusion.

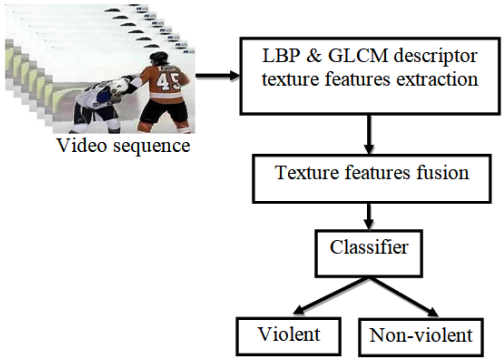

Figure 1. Work flow of the proposed method

Let g_c be the intensity of the frame I (m_c, n_c) and g_a is employed for neighborhood pixels which intends m_c as a distribution of sample points on a circle of radius of n_c and computes LBP. The value LBP of a pixel (m_c, n_c) is given in Eq. (1) and $p(x)$ is given in Eq. (2).

$$LBP_{(m_c,n_c)} = \sum_{a=0}^{1} p(g_a - g_c) * 2^a \quad (1)$$

$$p(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0, & otherwise \end{cases} \quad (2)$$

Initially, the windows are divided into cells and examined each pixel is compared to the 8 neighbors of each pixel in a cell, the pixels follow along in a clockwise direction. Labels the frame pixels by thresholding each pixel for the neighborhood for the center pixel value. If there is a neighborhood pixel value less than the threshold of the center pixel value then we assign "0", otherwise assign "1". From this 8-digit binary number is obtained. Finally, the histograms are calculated over the cell for each number which has frequently occurred. This histogram can be evaluated as 2^8 = 256-dimensional the feature vector for each frame.

2.2 GLCM features descriptor

GLCM is a texture-based feature descriptor that can be used to extract spatial variation of the matrix. According to the Haralick GLCM texture features [26] gives 14 texture features measured from the probability matrix to extract the characteristics of texture statistics of frames. In this work, we have used only four statistical properties; they are Contrast, Correlation, Energy and Homogeneity. Based on the 'x' grayscale pixel occurred in horizontally to the neighborhood pixels with the value of 'y'. Each element (x, y) in GLCM specifies the number of times that the pixel with value 'x' occurred horizontally neighboring to a pixel with value. GLCM metrics used to allow rotational invariance using set of rotational parameter. Generally, 8 orientations separated π /4 radian aside, where 'N' indicates the intensity value present in the frame. The properties (Energy, Correlation, contrast, and Homogeneity) are calculated using the normalized GLCM. The Contrast property is used to measure the local variations in the GLCM. It is also referred to as variance inertia, the whole frames intensity between a pixel and its neighbor is measured by it and its range is measured as Range = [0, size (GLCM, 1)-1]^2 and 0 is for constant frame as shown in Eq. (3).

$$Contrast = \sum_{x,y=0}^{N-1} p_{xy}(x-y)^2 \quad (3)$$

The correlation measures the occurrence of the specified pairs of the pixels of the joint probability (correlate to the neighbor of its pixel over the entire frames). It is unrepresentative for the constant frame and 1 and -1 for the positive and negatively correlated frame (i.e. Range = [-1, 1] as shown in Eq. (4). Where 'μ' and 'σ' represents mean and standard deviation respectively.

$$Correlation = \sum_{x,y=0}^{N-1} p_{xy} \frac{(x-\mu)(y-\mu)}{\sigma^2} \quad (4)$$

Energy is a property used to measure the sum of squared elements and also referred to as angular second moment or uniformity. Its value is 1 for the constant frame, otherwise ranges from 0 to 1 (i.e. Range = [0 1]) as shown in Eq. (5).

$$Energy = \sum_{x,y=0}^{N-1} (p_{xy})^2 \quad (5)$$

The Homogeneity evaluates the nearness of the distribution of the elements diagonally in GLCM. For diagonal elements its value is 1, otherwise it's in between 0 and 1(i.e. Range = [0, 1]) as shown in Eq. (6).

$$Homogenity = \sum_{x,y=0}^{N-1} \frac{p_{xy}}{1+(x-y)^2} \quad (6)$$

These features are referred to as Haralick features. The rotationally invariant frames are detected by texture analysis. The frames are evaluated from various angles of the matrices with respect to relation.

2.3 Features fusion descriptors

LBP feature descriptor has simply analyzed the texture of the frame, it works well of low-resolution, invariance to grayscale changes and gives local information of the frame. Computationally LBP feature descriptor is simple, so the properties of LBP is adopted in the detection of a violent event. But it is not invariant to rotation, only pixel values are used and ignore the magnitude information. Merely, the GLCM feature descriptor measures local variations of the spatial relationship of the pixels which is considered by examining the texture of the statistical function and the frequency of gray occurrences pattern with a specified angle and distance and it is rotational invariance. Therefore, to increase the accuracy and performance we depicted the detection of the violent and non-violent event based on the LBP and GLCM texture features descriptors in video scenes. LBP features constructed from the window regions in feature vectors representing the 256 features for each frame. Afterward, we used the GLCM feature extraction technique which measures local variations of the spatial relationship of the pixels which is considered by examining the texture of the statistical function, 4 features for each frame are obtained. The fusion of LBP and GLCM features get 260 feature vectors for each frame. Therefore, the integration of LBP and GLCM properties extracts the prominent features to increase the performance of the proposed method.

2.4 Classifiers

The violent event detection can be classified by using supervised learning techniques. Classification is a systematic categorization of features according to predefined training knowledge. In this section, supervised classifiers are used in our research work.

2.4.1 DT classifier

A DT classifier [27] is a supervised classifier and also called as a binary classifier, which constructs the hierarchical tree structure by means of training data. It uses the divide and conquer method for training data with labels. The DT contains the internal node and leaf node, these rules are query structure on test frames. The test frames are passing through the tree structure in a top-down manner, will finally end up in leaf node which represents a violent and non-violent event. DT classifier is used to separate the composite decision process into a simple process.

2.4.2 DA classifier

DA classifier [28] is also known as discriminant function analysis which can be used to classification sets of variables into different classes of the same data types. DA learns to discriminate features from the higher dimensional features from the higher dimensional feature space in which classes are well separated. DA evaluates the probability of a new instance that belongs to each class. The prediction is based on the class acquire the highest probability. At this moment, Bayes theorem is used to evaluate the probabilities of the class when the input data is given.

2.4.3 LR classifier

LR classifier is a kind of a statistical model that can be used for logistic functions to predict input training feature descriptor data value based on the prior knowledge or observation. This classifier is based on the relationship between one dependent and one or more independent variables [29]. LR tree which gives the numeric answer, every step in a prediction requires assuring the evaluation of one predictor. LR transforms its output using sigmoid function to return probability value which can be mapped to discrete class.

2.4.4 SVM classifier

SVM classifier [30] is a widely used supervised learning classification algorithm, which is based on the principle of structural risk minimization (SRM). SVM classifier is applicable for the feature vectors that are linearly separable. We use non-linear kernel function, when the two classes could not segregate in linearly. Hyperplane should have the largest margin in high-dimensional space to separate given input features into two or more classes. SVM helps in avoiding over-fitting about the problem. Coarse Gaussian kernel function has used in the experimentation.

2.4.5 KNN classifier

The KNN classifier [31] is a supervised learning classification algorithm which can be used to identify input features that are separated into two classes of a new point. The categorization of test data mainly depends on the labeled data points closest to the new point and that has those majorities of the neighbor's vote. In this work, we have used Medium kernel function and the number of neighbor k is set to 10.

Space-time post-processing

The post-processing technique which has used to improve the true positive rate and very first time Wang et al. [32] used the temporal post-processing technique and space-time post-processing technique introduced by Reddy et al. [33].

Algorithm 1: Detection of violent event: post-processing procedure.

Input: Different supervised classifiers output
Output: Detection of frame-level

```
for i=1, 2, 3... Sequence of frames do
  if All sequences of frames are violent or non violent
    then
     Frame ←Violent/Non-violent
     else
        Space post-processing:
        for j = 1, 2, 3...., all frames do
          Vote (all neighboring frames) ← Violent/Non-
          violent
          Frame ← Violent/Non-violent
         end
  end
end
Time post-processing:
for j = 1, 2, 3...., sequence of frames do
  Vote (30 neighboring frames) ← Violent/Non-violent
  Frame ← Violent/Non-violent
end
```

To improve the accuracy performance, we have used space-time post-processing technique by taking 30 frames for detection of frame level. The space-time post-processing is explained in Algorithm 1.

3. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we manifest the experimentation results obtained to the analysis of the proposed method implementation of violent events on two standard benchmark datasets are used. Subsequently, experimentation setting is described. Eventually, experimentation results are examined and compared with the state of the art methods.

3.1 Violent dataset

In this work, the two standard benchmark datasets Hockey Fight (HF) Dataset and Violent-Flow Dataset are used. Hockey Fight dataset consists 1000 video clips in which 500 clips present the fights and the other 500 clips present the no fight scenarios. Each video clip consists two or more persons with 50 frames. The sample frame of it is as shown in the Figure 2. The clips are dividing into 5 test and train dataset each consists of 100 clips of the fight and no fight scenario.

Violent-Flow (VF) Dataset consists of 246 video clips in which 123 are violent and other 123 are non-violent. The sample frame of it is as shown in Figure 2. The clips in this dataset are of different scenarios such as street and stadium. Both HF and VF datasets videos are downloaded from the web with average 3.60 seconds underneath unconfined, in the undomesticated circumstances. In this dataset divided into 5 tests and train dataset each consist of 25 clips of violence and non-violence. The frames have illumination variation, complex background, and occlusion which make the task more challenging.

Figure 2. Selected violent dataset frames

3.2 Experimentation setting

In this section, we have presented our proposed texture-based features extraction technique performance. Five-fold cross-validation technique [7, 10, 34] have been used for experimentation. Consequently, for each one of the two datasets is separated into five halves. For each time, four halves are employed for the training and left one halves employed for testing. We have computed this procedure five times and the final result is the average of the accuracy determined at each time. We have used different supervised classifiers for training and testing. Besides the accuracy, we have adopted as evaluation measures the area under the receiver operating characteristic (AUC) and classification accuracy of an algorithm. In the receiver operating characteristic (ROC) curve is used to compare with the state of the art methods and it shows correctly classified action events.

3.3 Results and analysis

Experimentation is persuaded on HF and VF separately. HF dataset video clips organized to evaluate two or more people fight scenes in video and VF dataset delineation to evaluate crowded scenes. Both HF dataset and VF dataset have complexity with scale changes, complex background and illumination. HF dataset ROC curves all classifiers using LBP descriptor demonstrated in Figure 3, illustrate that SVM classifier outperforms the other classifiers. Moreover detailed experimental result reported in Table 1. SVM classifier is superior in categorization of violent and non-violent to other classifiers, DA also performs well for the classification. Accuracy of 89.81% and AUC of 91.00% is obtained for HF dataset. Figure 4 explained VF dataset ROC curves; it shows that LR classifier surpasses the other classifiers, DA and SVM classifiers also perform indistinguishable to LR classifier. The attained experimental result of accuracy and AUC are respectively, 82.09% and 85.84% as shown in Table 1.

HF dataset ROC curves all classifiers using GLCM descriptor illustrated in Figure 5 reveal that SVM classifier better the other classifiers, KNN and LR classifier also performs satisfactory result. The acquired accuracy and AUC result of SVM classifier are respectively, 76.79% and 78.60% as shown in Table 2. VF dataset ROC curves all classifiers using GLCM result as shown in Figure 6, LR classifier achieve well classification than other classifiers. Accuracy of 82.16% and AUC of 83.25% are obtained, the result as shown in Table 2.

The fusion features descriptor of ROC curves all classifiers are presented in Figure 7 shows that SVM classifier is superior the other classifiers. Additionally detailed experimental result explained in Table 3. The obtained accuracy and AUC result of SVM classifier are respectively, 91.51% and 93.60% on the HF dataset. VF dataset ROC curves all classifiers using fusion feature descriptor as shown in Figure 8, SVM classifier is better classification than other classifiers. Accuracy of 89.06% and AUC of 93.00 % are obtained, the result as shown in Table 3.

Our proposed method used five classifiers to evaluate the performance of the algorithm, which takes a known training input dataset and its known responses to the output data to learn the categorization of violent and non-violent events.

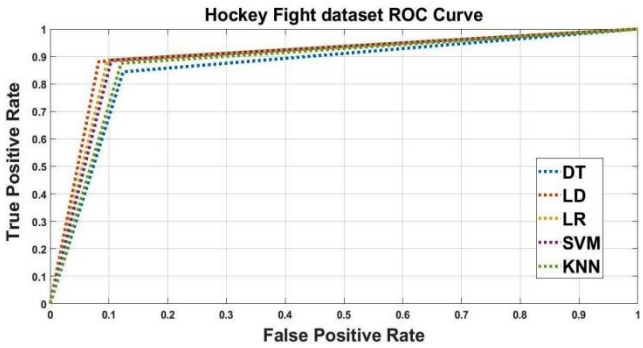

Figure 3. HF dataset ROC Curves all classifiers using LBP feature descriptor

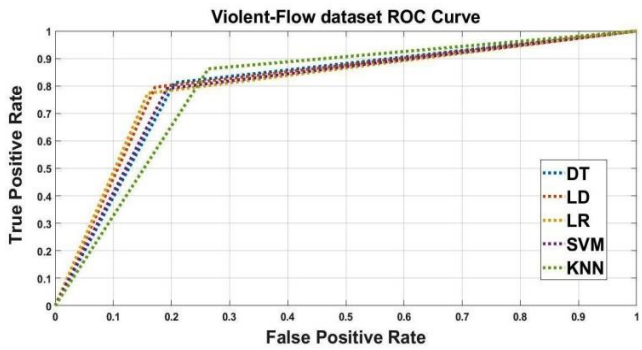

Figure 4. VF dataset ROC Curves all classifiers using LBP feature descriptor

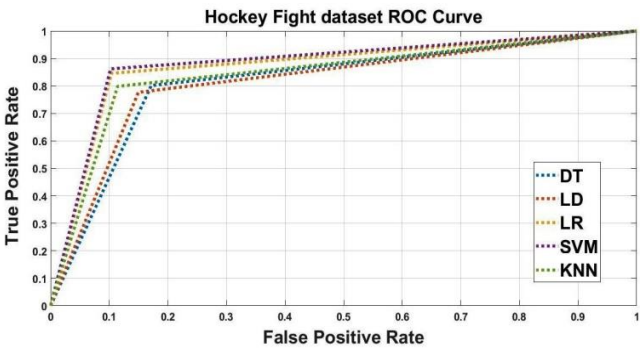

Figure 5. HF dataset ROC Curves all classifiers using GLCM feature descriptor

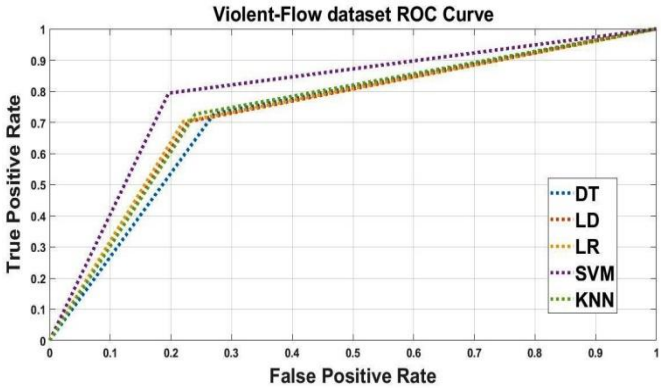

Figure 6. VF dataset ROC Curves all classifiers using GLCM feature descriptor

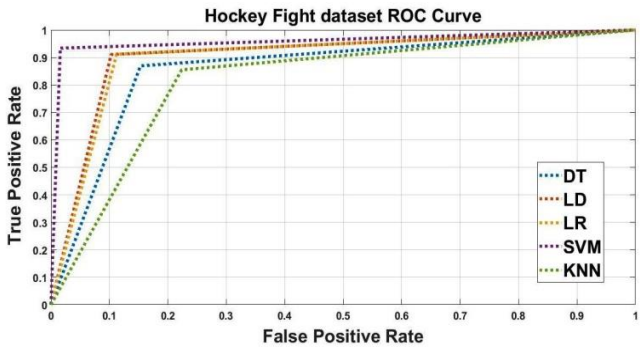

Figure 7. HF dataset ROC Curves all classifiers using texture features fusion descriptor

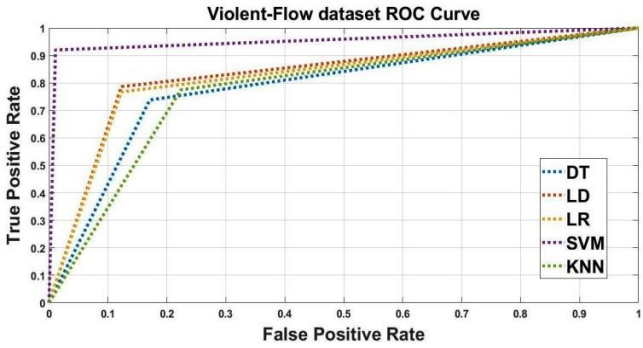

Figure 8. VF dataset ROC Curves all classifiers using texture features fusion descriptor

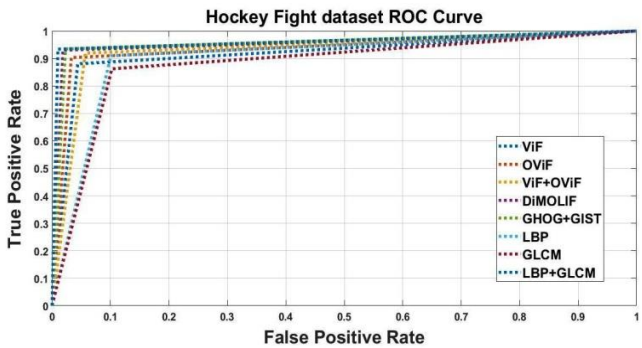

Figure 9. HF dataset ROC Curves of violent detection

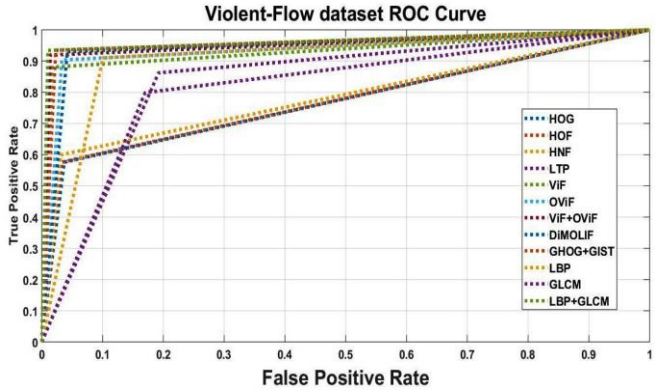

Figure 10. VF dataset ROC Curves of violent detection

Figure 11. Result of HF dataset for detection of nonviolent event and detection of violent event (top row), result of VF dataset for detection of non-violent event detection of violent event (bottom row)

LBP and GLCM descriptor works well used SVM classifier for HF dataset, but LR classifier effectively distinguish violent and non-violent using LBP and GLCM descriptor for VF dataset. Nevertheless, SVM classifier perform well for texture fusion features, it is used linear or non-linear Kernel function to categorize the feature, it handles non-linear solution and outlier better. But, LR classifier only handle linear solution and unreliable with well separated classes. DT failed to extract significance of the features and chance for outfitting. KNN has taken large time computation time cost, DA compute the addition of multivariate distribution and suffer multi collinearity.

HF dataset and VF dataset ROC curves illustrated experimental result of accuracy in Figure 9 and Figure 10 respectively. Detection results of HF and VF dataset are shown in Figure 11. Our proposed texture fusion features descriptor compared with HOG, HOF, HNF, LTP, ViF, OViF, ViF+OViF, DiMOLIF and GHOG+GIST state of the art methods for both HF dataset and VF dataset in Table 4 and Table 5 respectively. The GHOG+GIST feature descriptor also performs equivalent to our proposed method, but feature vectors dimension for each frame is large and computational time is high. Therefore, we deduce our proposed method is outperformance the other state of the art methods.

In this paper, our proposed method extracts texture features in each frame. Our feature descriptor intelligent to representation at all action without any constraint concerning the distribution. For the meantime, HOG, HOF, HNF, LTP and ViF feature descriptors uses only magnitude information, those descriptors cannot work if orientation changes. Therefore, the descriptors are failed to detect different actions changes because of same magnitude of histograms of optical flow. Moreover, OViF is an extension of ViF descriptor, which can use both magnitude and orientation, even though this descriptor is failed to detect crowded behavior compared to ViF. DiMOLIF feature descriptor is used interest frame spatial and temporal magnitude and orientation information to detect the violent event. GHOG+GIST feature descriptor is a fusion of gradient and texture feature descriptor, the method used scale, magnitude and orientation and gives adequate result but compare to our proposed method GHOG+GIST features vector dimension is substantial and required large time computation cost.

Table 1. The comparison of LBP descriptor accuracy with Standard Deviation (SD) and AUC of five classifiers are demonstrated in percentage using HF dataset and VF dataset

Classifiers	HF dataset		VF dataset	
	Acc (±SD)	AUC	Acc (±SD)	AUC
DT	85.47±2.78	87.72	79.15±2.80	81.68
DA	89.41 ± 3.95	90.36	81.33±1.05	81.45
LR	88.75 ± 5.43	89.06	**82.09±4.39**	**85.84**
SVM	**89.81±3.10**	**91.00**	79.75±3.51	79.90
KNN	87.58±2.10	88.90	77.71±2.73	84.24

Table 2. The comparison of GLCM descriptor accuracy with Standard Deviation (SD) and AUC of five classifiers are demonstrated in percentage using HF dataset and VF dataset

Classifiers	HF dataset		VF dataset	
	Acc(±SD)	AUC	Acc (±SD)	AUC
DT	74.99±3.84	74.97	77.12±2.15	79.33
DA	74.07 ±4.68	77.33	78.13±1.93	81.62
LR	75.91 ±4.27	78.18	**82.16±3.72**	**83.25**
SVM	**76.79±2.14**	**78.60**	77.97±2.62	82.69
KNN	76.36±3.24	76.64	75.84±2.99	76.80

Table 3. The comparison of texture features fusion descriptor accuracy with Standard Deviation (SD) and AUC of five classifiers are demonstrated in percentage using HF dataset and VF dataset

Classifiers	HF dataset		VF dataset	
	Acc (±SD)	AUC	Acc (±SD)	AUC
DT	85.22±3.59	86.25	78.99±1.47	79.84
DA	87.66± 5.20	89.72	79.74± 3.64	79.45
LR	87.20 ± 5.26	89.20	81.79±4.29	81.75
SVM	**91.51±1.51**	**93.60**	**89.06±3.32**	**93.00**
KNN	87.58±2.10	88.90	83.10± 2.84	82.40

Table 4. Performance comparisons of all other descriptors accuracy with Standard Deviation (SD) and AUC are presented in percentage using HF dataset

Method	HF dataset	
	Acc (±SD)	AUC
HOG [35]	87.8	-
HOF [35]	83.5	-
HNF [35]	87.5	-
LTP [35]	71.90±0.49	-
ViF [23]	81.60±0.22	88.01
OViF [7]	84.20±3.33	90.32
ViF+OViF [7]	86.30±1.57	91.93
DiMOLIF [34]	88.6±1.2	93.23
GHOG+GIST [10]	91.18±2.95	93.45
LBP	89.81± 3.10	91.00
GLCM	76.79±2.14	78.60
LBP+GLCM	**91.51±1.51**	**93.60**

Table 5. Performance comparisons of all other descriptors accuracy with Standard Deviation (SD) and AUC are presented in percentage using VF dataset

Method	VF dataset	
	Acc (±SD)	AUC
HOG [35]	57.43±0.37	61.82
HOF [35]	58.53±0.32	57.60
HNF [35]	56.52±0.31	59.94
LTP [35]	71.53±0.17	79.86
ViF [23]	81.20±1.79	88.04
OViF [7]	76.80±3.90	80.47
ViF+OViF [7]	86.00±1.41	91.82
DiMOLIF [34]	85.83±4.2	89.25
GHOG+GIST [10]	88.86±5.12	92.00
LBP	82.09 ± 4.39	81.84
GLCM	82.16± 3.72	83.25
LBP+GLCM	**89.06±3.32**	**93.00**

In experimentation, our proposed features descriptor performed on an 8GB RAM, Intel core i7 computer running Windows 10. MoSIFT [36] took 0.661 s/frame to compute each frame, which is comparatively very high. To reduce time complexity MoBSIFT [9] feature descriptor has been used, which has taken 0.032 s/frame for each frame. Deniz et al. [6] have taken 0.0419 s/frame but Gracia et al. [37] have taken 0.0225 s/frame less time to compute each frame. Our proposed LBP and GLCM feature descriptor have taken 0.1322 s/frame and 0.0282 s/frame to process the system for both violent and non-violent event with less time complexity, which shows that our proposed feature descriptors has very nearly equal to state of the art methods. The comparison of time computation as shown in Table 6.

Table 6. Duration taken to process frames

Method	Duration (sec/frame)	
	Violent	Non-violent
BoF (STIP) [36]	0.293	0.293
BoF (MoSIFT) [36]	0.661	0.661
Deniz et al. [6]	0.0419	0.0419
Serrano et al. [37]	0.0225	0.0225
BoF (MoBSIFT) [9]	0.257	0.257
BoF (MoBSIFT)+MF [9]	0.257	0.032
LBP	**0.1322**	**0.1322**
GLCM	**0.0282**	**0.0282**

4. CONCLUSIONS

In this work, our proposed texture features descriptors efficiently detect both crowded and uncrowned violent events in the video. We have used LBP and GLCM texture based feature extraction technique to detect the violent and non-violent event and the first time we have introduced texture features fusion descriptor. The experiment shows the performance very well for the two benchmark datasets namely Hockey fight dataset and Violent Flow dataset. In an experimentation, our proposed texture features fusion descriptors gives good result than other state of the art methods. In future, we aim to continue different texture, gradient and optical flow feature descriptor to more complex video events.

ACKNOWLEDGMENT

Lohithashva B.H. is very grateful to University Grant Commission (UGC) under RGNF (Rajiv Gandhi National Fellowship) for the financial support, Letterno.F1-17.1/2014-15/RGNF-2014-15-SC-KAR-73791/(SAIII/Website), Department of Computer Applications, JSS Science and Technology University, (Formerly Sri Jayachamarajendra College of Engineering), Mysuru, Karnataka, India.

REFERENCES

[1] Lohithashva, B.H., Aradhya, V.N.M., Basavaraju, H.T., Harish, B.S. (2019). Unusual crowd event detection: An approach using probabilistic neural network. In Information Systems Design and Intelligent Applications, pp. 533-542. https://doi.org/10.1007/978-981-13-3329-3_50

[2] Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N. (2010). Anomaly detection in crowded scenes. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 1975-1981. https://doi.org/10.1109/CVPR.2010.5539872

[3] Lloyd, K., Rosin, P.L., Marshall, D., Moore, S.C. (2017). Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. Machine Vision and Applications, 28(3-4): 361-371. https://doi.org/10.1007/s00138-017-0830-x

[4] Ribeiro, P.C., Audigier, R., Pham, Q.C. (2016). RIMOC, a feature to discriminate unstructured motions: Application to violence detection for video surveillance. Computer Vision and Image Understanding, 144: 121-143. https://doi.org/10.1016/j.cviu.2015.11.001

[5] Zhang, T., Yang, Z., Jia, W., Yang, B., Yang, J., He, X. (2016). A new method for violence detection in surveillance scenes. Multimedia Tools and Applications, 75(12): 7327-7349. https://doi.org/10.1007/s11042-015-2648-8

[6] Deniz, O., Serrano, I., Bueno, G., Kim, T.K. (2014, January). Fast violence detection in video. Proceedings of the 9th International Conference on Computer Vision Theory and Applications, Lisbon, Portugal, pp. 478-485. https://doi.org/10.5220/0004695104780485

[7] Gao, Y., Liu, H., Sun, X., Wang, C., Liu, Y. (2016). Violence detection using oriented violent flows. Image and Vision Computing, 48-49: 37-41. https://doi.org/10.1016/j.imavis.2016.01.006

[8] Ullah, F.U.M., Ullah, A., Muhammad, K., Haq, I.U., Baik, S.W. (2019). Violence detection using spatiotemporal features with 3D convolutional neural network. Sensors, 19(11): 2472. https://doi.org/10.3390/s19112472

[9] Febin, I.P., Jayasree, K., Joy, P.T. (2019). Violence detection in videos for an intelligent surveillance system using MoBSIFT and movement filtering algorithm. Pattern Analysis and Applications, 23: 611-623. https://doi.org/10.1007/s10044-019-00821-3

[10] Lohithashva, B.H., Aradhya, V.N.M., Guru, D.S. (2020). Violent event detection: An approach using GHOG-GIST descriptor. International Conference on Automation, Signal Processing, Instrumentation & Control (iCASIC), pp. 1-8.

[11] Dhiman, C., Vishwakarma, D.K. (2019). A review of state-of-the-art techniques for abnormal human activity recognition. Engineering Applications of Artificial

Intelligence, 77: 21-45. https://doi.org/10.1016/j.engappai.2018.08.014
[12] Vashistha, P., Singh, J.P., Khan, M.A. (2020). A comparative analysis of different violence detection algorithms from videos. In: Kolhe, M., Tiwari, S., Trivedi, M., Mishra, K. (eds) Advances in Data and Information Sciences. Lecture Notes in Networks and Systems, Springer, Singapore, pp. 577-589. https://doi.org/10.1007/978-981-15-0694-9_54
[13] Manjunath Aradhya, V.N., Basavaraju, H.T., Guru, D.S. (2019). Decade research on text detection in images/videos: A review. Evolutionary Intelligence, 1-27. http://dx.doi.org/10.1007/s12065-019-00248-z
[14] Roshan, S., Srivathsan, G., Deepak, K., & Chandrakala, S. (2020). Violence detection in automated video surveillance: Recent trends and comparative studies. The Cognitive Approach in Cloud Computing and Internet of Things Technologies for Surveillance Tracking Systems, pp. 157-171. https://doi.org/10.1016/B978-0-12-816385-6.00011-8
[15] Constantin, M.G., Stefan, L.D., Ionescu, B., Demarty, C.H., Sjoberg, M., Schedl, M., Gravier, G. (2020). Affect in multimedia: Benchmarking violent scenes detection. IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2020.2986969
[16] Manjunath Aradhya, V.N., Pavithra, M.S. (2016). A Comprehensive of transforms, Gabor filter and k-means for text detection in images and video. Applied Computing and Informatics, 12(2): 109-116. http://dx.doi.org/10.1016/j.aci.2014.08.001
[17] Nievas, E.B., Suarez, O.D., Garcia, G.B., Sukthankar, R. (2011). Violence detection in video using computer vision techniques. In: Real P., Diaz-Pernil D., Molina-Abril H., Berciano A., Kropatsch W. (eds) Computer Analysis of Images and Patterns. CAIP 2011. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 332-339. https://doi.org/10.1007/978-3-642-23678-5_39
[18] Yan, M.J., Meng, J.J., Zhou, C.L., Tu, Z.G., Tan, Y.P., Yuan, J.S. (2020). Detecting spatiotemporal irregularities in videos via 3D convolution autoencoder. Journal of Visual Communication and Image Representation, 67: 1-18. https://doi.org/10.1016/j.jvcir.2019.102747
[19] Mahantesh, K., Manjunath Aradhya, V.N., Niranjan, S.K. (2014). An impact of complex hybrid color space in image segmentation. In: Thampi S., Abraham A., Pal S., Rodriguez J. (eds) Recent Advances in Intelligent Informatics. Advances in Intelligent Systems and Computing, Springer, Cham, pp. 73-82. https://doi.org/10.1007/978-3-319-01778-5_8
[20] Gopala Krishna, M.T., Aradhya V.N.M., Ravishankar, M., Ramesh Babu, D.R. (2012). LoPP: Locality preserving projections for moving object detection. Procedia Technology, 4: 624-628. https://doi.org/10.1016/j.protcy.2012.05.100
[21] Amith, R., Aradhya V.N.M. (2017). Linear projective approach for moving object detection in video. In Proceedings of the 1st International Conference on Internet of Things and Machine Learning, pp. 1-4. https://doi.org/10.1145/3109761.3109767
[22] Praskash, B.A., Ashoka, D.V., Aradhya, V.M., Naveena, C. (2016). Exploration of neural network models for defect detection and classification. International Journal of Convergence Computing, 2(3-4): 220-234. https://doi.org/10.1504/IJCONVC.2016.090081
[23] Hassner, T., Itcher, Y., Kliper-Gross, O. (2012). Violent flows: Real-time detection of violent crowd behavior. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, pp. 1-6. https://doi.org/10.1109/CVPRW.2012.6239348
[24] Ojala, T., Mäenpää, T., Viertola, J., Kyllönen, J., Pietikäinen, M. (2002). Empirical evaluation of MPEG-7 texture descriptors with a large-scale experiment. In Proc. 2nd International Workshop on Texture Analysis and Synthesis, Copenhagen, Denmark, pp. 99-102.
[25] Baumann, F., Ehlers, A., Rosenhahn, B., Liao, J. (2016). Recognizing human actions using novel space-time volume binary patterns. Neurocomputing, 173: 54-63. https://doi.org/10.1016/j.neucom.2015.03.097
[26] Partio, M., Cramariuc, B., Gabbouj, M., Visa, A. (2002). Rock texture retrieval using gray level co-occurrence matrix. Proc. of 5th Nordic Signal Processing Symposium.
[27] Quinlan, J.R. (1986). Induction of decision trees. Machine Learning, 1(1): 81-106. https://doi.org/10.1007/BF00116251
[28] Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N. (2003). Regularized discriminant analysis for the small sample size problem in face recognition. Pattern Recognition Letters, 24(16): 3079-3087. https://doi.org/10.1016/S0167-8655(03)00167-3
[29] Peng, C.Y.J., Lee, K.L., Ingersoll, G.M. (2002). An introduction to logistic regression analysis and reporting. The Journal of Educational Research, 96(1): 3-14. https://doi.org/10.1080/00220670209598786
[30] Cortes, C., Vapnik, V. (1995). Support vector machine. Machine Learning, 20(3): 273-297. https://doi.org/10.1007/BF00994018
[31] Kuncheva, L.I. (1995). Editing for the k-nearest neighbors rule by a genetic algorithm. Pattern Recognition Letters, 16(8): 809-814. https://doi.org/10.1016/0167-8655(95)00047-K
[32] Wang, T., Snoussi, H. (2014). Detection of abnormal visual events via global optical flow orientation histogram. IEEE Transactions on Information Forensics and Security, 9(6): 988-998. https://doi.org/10.1109/TIFS.2014.2315971
[33] Reddy, V., Sanderson, C., Lovell, B.C. (2011). Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture. CVPR 2011 WORKSHOPS, Colorado Springs, CO, pp. 55-61. https://doi.org/10.1109/CVPRW.2011.5981799
[34] Mabrouk, A.B., Zagrouba, E. (2017). Spatio-temporal feature using optical flow based distribution for violence detection. Pattern Recognition Letters, 92: 62-67. https://doi.org/10.1016/j.patrec.2017.04.015
[35] Yeffet, L., Wolf, L. (2009). Local trinary patterns for human action recognition. 2009 IEEE 12th International Conference on Computer Vision, Kyoto, pp. 492-497. https://doi.org/10.1109/ICCV.2009.5459201
[36] Chen, M.Y., Hauptmann, A. (2009). Mosift: Recognizing human actions in surveillance videos. Technical report, Carnegie Mellon University, Pittsburgh, USA, pp. 1-17.
[37] Gracia, I.S., Suarez, O.D., Garcia, G.B., Kim, T.K. (2015). Fast fight detection. PloS One, 10(4).

https://doi.org/10.1371/journal.pone.0120448

NOMENCLATURE

c	center of the pixel value
p	pixels
N	intensity value present in the frame
x	threshold of the center pixel value

Greek symbols

μ	mean
σ	standard deviation
π	pi

Subscripts

m_c, n_c	neighbor pixels
xy	the occurrence of specified pairs of the pixels of the joint probability