Abnormality Detection Using LBP Features and K-Means Labelling based Feed-Forward Neural Network in Video Sequence

Ruchika, Ravindra Kumar Purwar

Abstract— **Video surveillance is widely used in various domains like military, commercial and consumer areas. One of the objectives in video surveillance is the detection of normal and abnormal behavior.It has always been a challenge to accurately identify such events in any real time video sequence. In this paper, abnormality detection method using Local Binary Pattern and k-means labeling basedfeed-forward neural network has been proposed. The performance of the proposed method has also been compared with four other techniques in literature to show its worthiness. It can be seen in the experimental results that an accuracy of up to 98% has been achieved for the proposed technique.**

Index Terms—**NN, k-mean labeling, abnormality detection, video surveillance**

I. INTRODUCTION

Digital video processing is widely used in the application areas of surveillance and security, traffic monitoring, night vision, event detection, duplication detection, military fields, commercial and consumer application, space missions, advertisements, scientific research, medical applications and many more.

Over the years, automated video surveillance has gained an increased focus amongst researchers due to its widespread application in government, residential and commercial sectors. In general terms, video surveillance can be characterized as the action of monitoring activities of any moving objects particularly the actions of people in both secluded and crowded environments. It is accomplished via placing a camera, usually at the top of the rigid structure in order to record the activities that can be monitored via a closed loop system. The camera can either be static or dynamic depending on the geographical area to be covered. From these videos the objects are tracked and these detected objects are categorized into semantic categorizes labeled as human, crowd, car, truck etc.on the basis of color and shape of the objects. These labels help to enhance the tracking of objects using consistency constraints applied on the basis of time. The typical overview of video surveillance system is illustrated in fig. 1. The real-time video stream is fed in to the system from camera. Once the pixel data is received,region of interest detection module executes robust object detection procedures to discover the objects in the video sequence. The next module is video tracking that involves objecttracking. The essential parameters are extracted and updated in this module like inter-frame displacement of pixels of the object under consideration. The camera feedback and video tracking module both use these parameters. Depending upon the updated value of these parameters in present frame,the camera feedback module produce essential commands for deliberate camera movement [1] in vital directions in order to survey the tracked object. The video tracking module output is also used to produce the alarming signals, so that the security persons can detect the unauthorized movement in sensitive areas. Display screens are deployed at results module to identify the unauthentic eventsor objects.

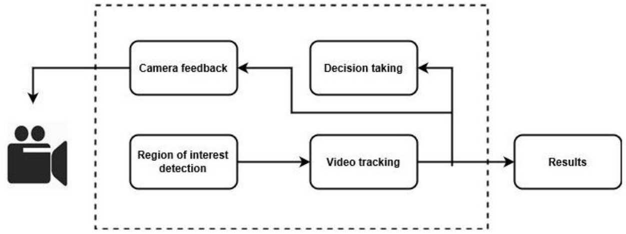


Fig 1: Block diagram of Video Surveillance [1]

II. RELATED WORK

Authorsin [2] present a CNN based approach for detecting an abnormal event. This method is capable for providing spatial reliability. Temporal CNN Pattern (TCP) is combined with traditional optical-flow to provide corresponding information of appearance and motion designs.The advantages of this method are spatial reliability and semantic embedding with low dimensions. A deep learning method to detect the video abnormality is proposed in [3]. convolutional long short term memory (ConvLSTM) is a temporal sequencer and spatialfeature extractor that perform the abnormality detection. False alarms are the drawback produced, based on the complex activities present in the scene. Hierarchical Dirichlet Process (HDP) connects different video events like low-level visual features, simple atomic activities, and multi-agent interactions [4].Low -level visual features are extracted on the basis of optic flow.

Revised Manuscript Received on June 15, 2019.

Ruchika, Research Scholar, USIC&T, Guru Gobind Singh Indraprastha University, Dwarka, Delhi, India

Ravindra Kumar Purwar, Associate Professor, USIC&T, Guru Gobind Singh Indraprastha University, Dwarka, Delhi, India

HDP model decides the types of atomic activities and interactions.For classification accuracy the sequential dependences among the video events recognized using HDP-HMM are combined with Gaussian Process (GP) classifier. Energy-based Anomaly Detection (EAD) [5] consists of two phases: training phase and detection phase. Energy Based Model (EBM) training is done in first phase, whereasin second phase error map thresholding for high probable abnormal region selection, filtering of small anomaly objects over graphical connected components for surviving regions and incremental update of EBM with video stream data. Restricted Boltzman Machines (RBM) and DeepBoltzman Machines (DBM), the two EBM techniques are experimented over three benchmark datasets: Avenue, UCSD Ped1, Ped2 for performance evaluation.Cost effective CNN based fire detection system [6] enables the video surveillance system to avoid dangerous events. CNN architecture is shown to be efficient in detecting flames correctly. High false alarm rate is the limitation of this architecture.

Theabnormal behavior of the psychiatric patient is monitored in [7]. Spatialand temporal features are used to characterize the normal behavior of humans. An unsupervised learning based on N-cut algorithm accompanied by SVM,labels the segments of video sequence to be normal and Condition Random Field (CRF) with an adaptive threshold performs the abnormality detection. S. Wang et al [8] proposed video localization and anomaly detection using local motion based joint video representation and One-Class Extreme Learning Machine (OCELM).In spatio-temporal video cuboidsSpatially Localized Histogram of Optical Flow (SL-HOF) portrays the motion of forefront object using 3D local region motion demonstration. With this foreground localization scheme, the motion of local texture is characterized in the video foreground byUniformLocal Gradient Pattern based Optical Flow (ULGP-OF)descriptor. Automated learning of human behavior characteristics from a large number of videos consist of normal and abnormal behavior using CNN is proposed in [9]. An abnormal behavior detection is performed in diverse conditions with differences in background, number of subjects like individual, two humans or crowd, and a variety of uncommon human activities. The effect of frame rate onrecognizing the human activities with four differentmethodologyover three bench mark datasets is explained in [10]. The methods evaluated are Bag of Visual Words (BoVW), Action Snippets, Dense Trajectories, and Motion Interchange Patterns. For action recognition key frame selection method selects the combination of framesfrom action sequences.

Inthis paper, afeed-forward neural networkbased abnormal behavior detection technique using k-mean labeling is proposed. Neural networks are sufficiently capable of detecting humans in the image frames.First step here is pre-processing thatinvolves image frame extraction from videos and thenapplying the median filter for noise removal.The background subtraction is performed using frame differencing of current frame with reference frame. Moving objects are detected using image morphological dilation operation. Local Binary Pattern (LBP) features of this dilated image is extracted, and given as input to feed-forward neural network for classification. The neural network is pretrained for abnormality detection using k-mean labeling and based on this training the test frames are fed to the network for normal or abnormal behavior detection. Rest of the paper is organized as follows. In section III, the proposed classification method is described which is evaluated and compared with other methods experimentally in section IV. Finally, the paper is concluded in Section V.

III. PROPOSED ABNORMALITY DETECTION TECHNIQUE

In this section, we give the detailed explanation of proposed methodology. The main aim of this work is to detect the abnormal behavior based on the interaction of two persons. The benchmark dataset used here is UT-Interaction dataset [11]. The proposedwork is summarized through a flow chart given in fig. 2. Initially,in the preprocessing stage, input video sequence is converted into image frames in which noise removal is done using median filter. In background suppression, the background is subtracted to obtain the foreground objects from the image using frame differencing method. From extracted foreground, moving objects are detected using morphological dilation operation. LBP features are then extracted and used for classification purpose using k-means labeling based feed-forward neural network.

A. Pre-processing

Preprocessing is an important stage where the frame conversion occurs i.e. the image frames are obtained from input video.

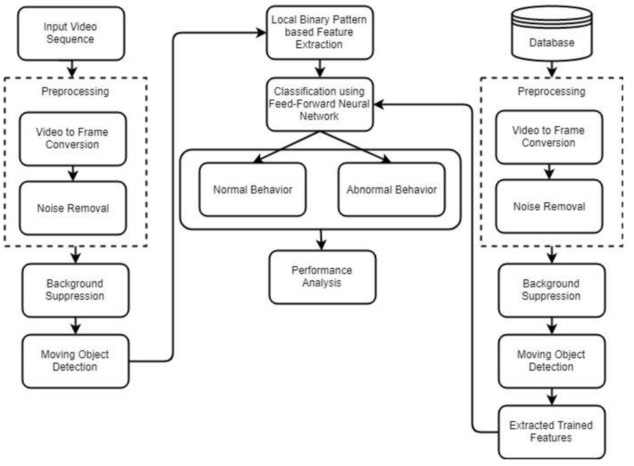


Fig.2: Flowchart of proposed abnormality detection method

Median filter of size 3×3 is used to remove the noise resulting in improved quality of image frames. It is anon-linear filter, that removes the noise from the digital images.

B. Background Suppression

Background suppression is performed over the image frames obtained from the videos in previous step using frame differencing technique. It is the technique in which current frame of video is subtracted from the reference frame using equation (1) and foreground is extracted from the image frame. Since, stationary camera has been used in the referred dataset keeping background information fixed, all the frames of video are subtracted from the first frame.

Let I_1 denote the first frame and I_j ($j \geq 2$) be the j^{th} frame of given video frame sequence then the differential image D_j is obtained as

$$D_k = |I_1 - I_j| \qquad (1)$$

C. Moving Object Detection

The background of the images is suppressed to remove the unwanted distortions for acquiring required foreground part in the images. Morphological dilation [12] is applied on the extracted foreground frames to detect moving objects in the video.

In dilation, depending on the shape and size of the structuring element used for processing the image, number of pixels are added to the object boundary in image.Output pixel value is the largest of the pixel values in neighborhood.

D. Feature Extraction

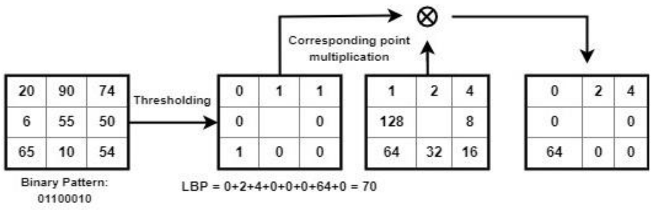
Feature extraction is the process of extracting required features from input data. These features are intended to be non-redundant and having relevant information that is desired by the user or machine. In proposed work, the features are extracted using Local Binary Pattern (LBP).LBP (Ojala et. al. [13])for neighborhood of size N is defined by the following

$$LBP = \sum_{k=0}^{N-1} r(p_k - p_c)2^i \qquad (2)$$

$$r(y) = \begin{cases} 1, & y \geq 0 \\ 0, & y < 0 \end{cases} \qquad (3)$$

$p_c$represent the central pixel value, whereas, p_k ($0 \leq k \leq N-1$) denotes the pixel values of N neighborhoods.

Thesample calculation of LBP is shown in Fig. 3.



20	90	74
6	55	50
65	10	54

0	1	1
0		0
1	0	0

1	2	4
128		8
64	32	16

0	2	4
0		0
64	0	0

Thresholding
Corresponding point multiplication
⊗
Binary Pattern: 01100010
LBP = 0+2+4+0+0+0+64+0 = 70

Fig.3Illustration for LBPcomputation [14]

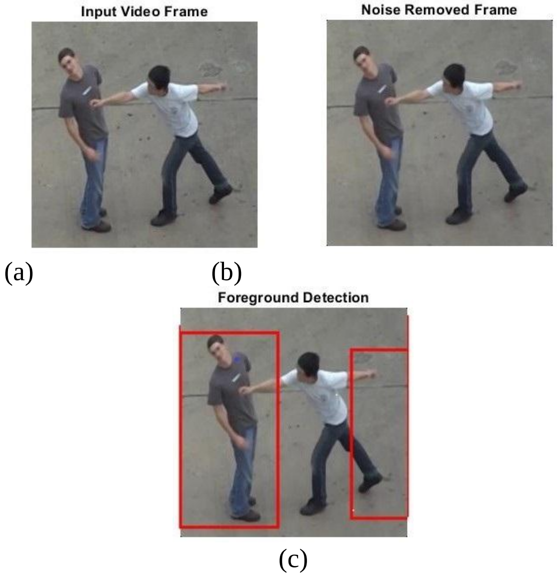
These LBP features are fed into the feed-forward neural network for classification.

E. Classification

In this stage,the extracted feature patternsare classified using k-mean labeling and feed-forward neural network.Initially, the unlabeled data defined through LBP is clustered using k-means labeling. The aim of k-mean labeling is the grouping of data in k number of clusters based on the similarity of features.The algorithm produces centroids that are used to label new data and these labels are used for the training of the data through feed-forward neural network. During training the label is split into binary values as 0’s and 1’s. Normal behavior is labeled as 0 and abnormal as 1. Then this label is used for classification of normal and abnormal behaviorin the videos under test.

IV. PERFORMANCE ANALYSIS

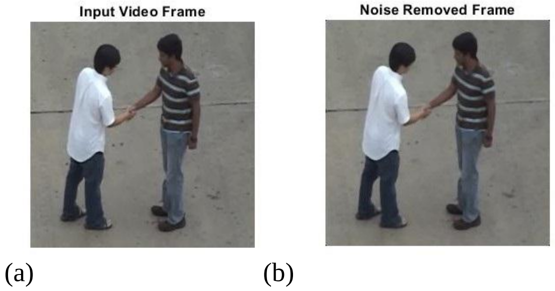
UT-Interaction dataset is used in proposed work, where the six classes of actions are executed continuously in videoswith interaction between two human beings like pointing, hugging, hand-shaking, pushing, kicking and punching. In proposed work these interactions are divided into two categories as abnormal and normal behavior.



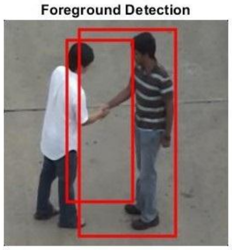
(a) (b)

(c)

Fig.4Abnormal event (a) Input video frame (b) Noise removed frame (c) foreground detection



(a) (b)



(c)

Fig.5Normal event (a) Input frame (b) Noise removed frame (c) foreground detection

The simulation has been carried out in MATLAB 2018b on computing machine with processing speed of 2.21GHz, RAM of 8GB and disk of 1TB in Window 10 environment. Results of the proposed method for detecting both the abnormal and normal activity is illustrated in fig.4 and fig. 5 respectively.

Performance of the proposed method has been measured in term of precision,recall, F1-Score, dice coefficient, kappa coefficient, jaccard coefficientand accuracy.

Definitions of performance parameters

Precision [15](P)is used to find the appropriate results while retrieving an information. Recall can be determined by measuring the number of truthfully appropriate results.

Precision is defined as the ratio of true positive (TP) observations to total positive predictionsi.e. the sum of true positive and false positive (FP)

$$P = \frac{TP}{TP+FP} * 100\% \quad (4)$$

Recall [15](R)is measured as the ratio of correct positive predictions i.e. true positives to all actual observations i.e. sum of true positives and false negatives(FN)

$$R = \frac{TP}{TP+FN} * 100\% \; (5)$$

F1-Score [15](F)is weighted average of recall and precision. This performance measure works upon both false positive and false negative parameters.

$$F = \frac{2*(Precision*Recall)}{Precision+Recall}* 100\%(6)$$

Dice coefficient [16](D) measures the similarity of two samples under consideration. Jaccard (J) [15][16] is defined as the ratio of true positive to the sum of true positive, false positive, and false negative. If the value is 100%, the two objects are identical.

$$D = \frac{2TP}{2TP+FP+FN} * 100\% \qquad (7)$$

$$J = \frac{TP}{TP+FP+FN} * 100\% \qquad (8)$$

Kappa coefficient [15](K) comparison measures total accuracy and random accuracy of the system. Total accuracy (TA) is an observational probability whereas random accuracy (RA) is hypothetical expected probability.

$$\mathrm{K} = \frac{TA-RA}{1-RA} * 100\% \qquad (9)$$

$$TA = \frac{TP+TN}{TP+TN+FP+FN} \qquad (10)$$

$$RA = \frac{(TN+FP)*(TN+FN)+(FN+TP)*(FP+TP)}{(TP+TN+FP+FN)*(TP+TN+FP+FN)} \qquad (11)$$

Accuracy [15](A)is dignified by identifying the correctness of the classified results. It is the proportion of correct predictions to total predictions.

$$A = \frac{TP+TN}{TP+FN+TN+FP} * 100\% \qquad (12)$$

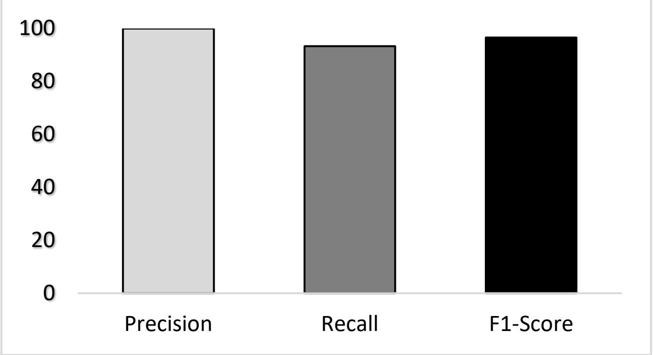


Fig. 6 Precision, Recall and F1-Score for proposed method

Fig. 6depicts precision, recall and F1 Scorefor proposed approach. The precision, recall and F1 Score evaluate to 100%, 93.33% and 96.55% respectively.

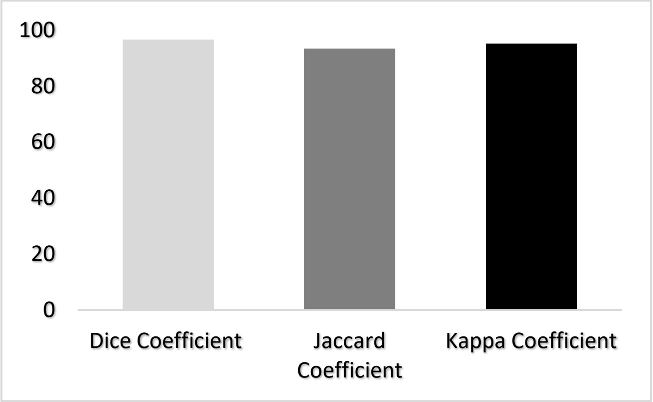


Fig. 7 Dice, Jaccard, and Kappa coefficient for proposed method

Fig. 7 demonstrates the coefficient results of proposed approach. Values of Dice, Jaccard, and Kappa coefficient are 96.55%, 93.33%, 95.14% respectively. Accuracy of abnormality detection for various methods is shown in table 1.

Table 1: Accuracy comparison of proposed method and existing techniques

Technique	Accuracy Percentage
SVM-STIP	85.00%
SVM context	88.33%
Model BM	93.33%
DEEP model	95.00%
Proposed method	98.00%

The performance of the proposed method has been compared with four parallel techniques in literature – SVM - STIP, SVM context, Model BM and DEEP model [17]. It can be seen that the accuracy of the proposed method is 98% whereas, DEEP model based algorithm is the second best algorithm with accuracy up to 95%.

V. CONCLUSION

A Local Binary Pattern based abnormality detection technique for video surveillance has been proposed in this paper. Initially, video frames are preprocessed using median filter for possible noise removal and foreground information is extracted through background subtraction method which is followed by moving object retrieval using dilation operation. LBP features are then extracted and used for classification purpose with the help of k-means clustering based feed-forward neural network. Performance of the proposed method has been measured using various parameters like precision, recall, F1-Score, accuracy etc. and compared with other methods as well. The maximum accuracy has been observed for the proposed technique.

ACKNOWLEDGMENT

This research is supported by Visvesvaraya PhD Scheme ofMinistry of Electronics and Information Technology, Government of India, being implemented by Digital IndiaCorporation.

REFERENCES

1. Singh, Sanjay, Sumeet Saurav, Chandra Shekhar, and Anil Vohra. "Prototyping an automated video surveillance system using FPGAs." International Journal of Image, Graphics and Signal Processing 8, no. 8,2016, pp : 37
2. Ravanbakhsh, Mahdyar, Moin Nabi, Hossein Mousavi, Enver Sangineto, and Nicu Sebe. "Plug-and-play cnn for crowd motion analysis: An application in abnormal event detection." In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1689-1698. IEEE, 2018.
3. Chong, Yong Shean, and Yong Haur Tay. "Abnormal event detection in videos using spatiotemporal autoencoder." In International Symposium on Neural Networks,Springer, Cham, 2017,, pp. 189-196..
4. Yang, Michael Ying, Wentong Liao, Yanpeng Cao, and Bodo Rosenhahn. "Video Event Recognition and Anomaly Detection by Combining Gaussian Process and Hierarchical Dirichlet Process Models." Photogrammetric Engineering & Remote Sensing 84, no. 4,2018, pp: 203-214.
5. Vu, Hung, Tu Dinh Nguyen, and Dinh Phung. "Detection of Unknown Anomalies in Streaming Videos with Generative Energy-based Boltzmann Models." arXiv preprint arXiv:1805.01090 (2018).

6. Muhammad, Khan, Jamil Ahmad, Irfan Mehmood, Seungmin Rho, and Sung Wook Baik. "Convolutional neural networks based fire detection in surveillance videos." IEEE Access 6,2018,pp: 18174-18183.

7. Hsu, Shih-Chung, Cheng-Hung Chuang, Chung-Lin Huang, Ren Teng, and Miao-Jian Lin. "A video-based abnormal human behavior detection for psychiatric patient monitoring." In 2018 International Workshop on Advanced Image Technology (IWAIT), IEEE, 2018,pp. 1-4.

8. Wang, Siqi, En Zhu, Jianping Yin, and Fatih Porikli. "Video anomaly detection and localization by local motion based joint video representation and OCELM." Neurocomputing 277,2018,pp: 161-175.

9. Tay, Nian Chi, Tee Connie, Thian Song Ong, Kah Ong Michael Goh, and Pin Shen Teh. "A robust abnormal behavior detection method using convolutional neural network." In Computational Science and Technology, Springer, Singapore, 2019,pp. 37-47..

10.Harjanto, Fredro, Zhiyong Wang, Shiyang Lu, Ah Chung Tsoi, and David Dagan Feng. "Investigating the impact of frame rate towards robust human action recognition." Signal Processing124 (2016): 220-232.

11.http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html

12.Gonzalez and Woods, Digital Image Processing, 4th edition, pearson, 2017.

13.Ojala T, Pietik¨ainen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 1996, 29(1) pp: 51−59

14.Ke-Chen, Song, et al. "Research and perspective on local binary pattern." Acta Automatica Sinica 39.6 2013,pp: 730-744.

15.Akosa, Josephine. "Predictive accuracy: A misleading performance measure for highly imbalanced data." In Proceedings of the SAS Global Forum. 2017.

16.Taha, Abdel Aziz, and Allan Hanbury. "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool." BMC medical imaging 15, no. 1 2015, pp: 29.

17.Wang, Xiaoyang, and Qiang Ji. "Hierarchical context modeling for video event recognition." IEEE transactions on pattern analysis and machine intelligence 39, no. 9,2016, pp: 1770-1782.

AUTHORS PROFILE

Ruchika has obtained her M. Tech. in Information Technology from University School Of Information, Communication and Technology (USICT) Guru Gobind Singh Indraprastha University, Delhi. She is pursuing her Ph.D. from USICT, GGSIP University, Delhi. Her research interest includes machine learning, image processing&video processing.Sheis currently working on video processing. She is student member of ACM.

Dr. R.K. Purwar has obtained his ME (computer science & engineering) degree from MNREC Allahabad (currently known as MNNIT Allahabad). He has pursued his doctorate from university school of information and communication technology (USICT), GGSIP University, Delhi. He is a life member of Computer Society of India (CSI) and Indian Society of Technical Education (ISTE). He has various publications in peer reviewed quality international journals and conferences. His research interests include image/video processing, pattern recognition, video security and database management.