INTERNATIONAL JOURNAL OF ADVANCED INNOVATIVE TECHNOLOGY IN ENGINEERING

Published by Global Advanced Research Publication House Journal

Home page: www.ijaite.co.in

State of the Art Human Activity Detection

[1]Veena K. Katankar, [2]Piyush A. Fukat, [3]Swapnil S. Urade, [4]Tushar S. Tajne, [5]Puja S. Delikar, [6]Arti M. Waghmare

[1,2,3,4,5,6]Department of Computer Engineering, Suryodaya College of Engineering and Technology, Nagpur, Maharashtra, India

[1]veenakatankar@gmail.com, [2]piyushfukat7@gmail.com, [3]swapnilurade18@gmail.com, [4]ttajne29@gmail.com, [5]pujadelikar@gmail.com, [6]artiw362@gmail.com

Article History

Received on: 06 April 2022

Revised on: 17 April 2022

Accepted on: 06 May 2022

Keywords: Deep Learning using CNN (Convolutional neural Network), Surveillance camera, dataset videos, Snippet prediction, support vector machine.

e-ISSN: 2455-6491

Production and hosted by
www.garph.org

ABSTRACT

The mission of the Human Fight Detection is to control the growing violence in day-to-day life of human by implementing the appropriate violence detecting algorithms. Surveillance scenarios like prisons, psychiatric centres or even embedded in camera phones, but cannot be more capable to solve the problem or cannot be more effective to take action on needed time such types of violence are now on growing tremendously. As a consequence, there is rising interest in developing violence detection algorithms. Recent the work considered by the well-known Bag of Words framework for the specific problem of violence detection. Under the framework of this, spatial-temporal features are extracted from the videos sequences and used for classification. The dataset of violence collected, which consists of fight scenes from surveillance camera videos available in sources of online platforms. The dataset is made publicly available. From the extensive experiments conducted on Hockey Fight, Pellicle, and the newly collected fights datasets, the overall research have been totally helpful for us for finding the new potential for the future references.

1. INTRODUCTION

The thought of developing this project comes to do some good for the fight prevention of human beings. Few years ago, the cameras and other, surveillance equipment were used on different places like streets, banks, hospitals, educational institutes, offices, etc., for monitoring the suspicious activities of humans. Observing and monitoring the behaviours of suspicious activities are normal or not this made it difficult tasks to do and also access the more storages of recorded videos. Overall finding the suspicious activities managing all data is very difficult task. And also, the necessary action was taken which are not implemented on time. These all-different methods were used by todays. To deals with this, different methods have been developed to recognize the violence of real life.

These methods were helps to detect the suspicious activities in the surveillance equipment. In these methods different approaches are proposed that which runs with different input parameters. Flow, time, appearance, acceleration, etc., are the

different attributes of parameters. In the suspicious activity detection process, the first process is to divide a whole video into frames and segments. And the second process is to detect the object from the video frames which is captured. Third process is extracting the attributes of the video according to the implemented methods. And lastly, it successfully detects the suspicious activities from the frames. Different methods of the detection process from the surveillance videos by using computer vision are discussed and explored successfully using systematic review.

2. RELATED WORK

Initial proposals adopted the methodology of violence detection using the degree of motion, recognizing sounds features by exploiting audio visual correlation, skin and blood patterns exposure and discovering scream like cues in audio exploiting audio video correlation for violent scenes detection. Then, audio features are used to detect gunshots, explosions and car breaking activities, using Hidden Markov models (HMM) and Gaussian mixtures model. Time and frequency domain are classified using Support Vector Machine (SVM).

Chen et al. used spatial temporal videos cubes and local binary motion detectors. Lin and Wang exploited weekly supervised audio classifier co-trained with video features of motion, explosion. Giannakopoulous et al. performed audio-visual features analysis using statistics, Nearest neighbors (KNN) performed average motion. And detection of faces and fighting are suggested by Chen et al. for determining potential violent contents in videos.

Later Bermejo et al. performed encouraging results with 90% accuracy using MoSIFT feature descriptor revealing two potential datasets “Movies dataset” and “Hockey dataset”, specifically designed for violence detection job. Following that, Kernel Density Estimation (KDE) was exploited to gather feature selection on MoSIFT descriptor with sparse coding reporting accuracy 94.3% on Hockey dataset determining aggressive human behaviours. Motion blob, another form of motion features is used to recognized fight and non-fight video frames, by extracting basic features of blobs (perimeter, area etc.) yielding 97.8% accuracy. Another approach describes fuzzy region emerges in image frames due to abrupt violence motion patterns, reporting 98.9% accuracy on Movies dataset.

Recently, 3D ConvNets based model with the prior knowledge is studied on Hockey dataset. 3D Convolutional Neural Network architecture C3D is experimented on Movie sequences and Hockey. More recently, using Hough Forest features 2D-CNN model is proposed. This system is revealing finest accuracy results 94.6%, 99% on Movies and Hockey datasets respectively, as compare to all previous techniques of hand-defined features deep representation models and detectors.

In short, a significant number of algorithms perform motion-visual cues analysis by recognizing violent activities or by examining motion patterns using hand-defined features. A few deep learning-based approaches also incorporated 3DCNN and CNN architectures. Moreover, a 2D-CNN model, by taking advantage of network deep representations, in combination with hand-defined Hough Forest features, is constructing finest classifier to discriminate human violence behaviour. However, deep learning models have certain restriction. With enormous amount of domain specific data, they require huge computational power. Developing huge amount of labelled dataset is time-consuming task and laborious. For target domain this shortcoming is leading to a major bottleneck in training deep learning model from scratch. The approach of transfer learning becomes useful. A huge dataset is re-trained on the target domain specific dataset for a source network pre-trained. This scheme eliminates the need of producing training model from scratch as well as huge dataset. In this regards, winner models of ImageNet Large Scale Visual Recognition Challenge (ILSVRC), such as GoogleNet, AlexNet, VGGNet, and ResNet trained on 15 million annotated images for 1000 categories, are publicly available as open-source pre-trained models. These models can be used as pre-trained networks learning to develop specific target networks domain, such as for the task of human fight detection.

3. METHODOLOGY

Violence is suspicious events or day to day life activities were happened in normal life. Detection of such events in surveillance videos through computer vision becomes the active topic in the field of action detection. There are many researchers who proposed different techniques and method for detection of abnormal events which is rapidly increasing of crime rate for more accurate detection. Different techniques of violence detection are used in the recent years. There are techniques of violence detection which are mainly classified into three categories: VDT using machine learning, VDT using SVM and VDT using deep learning.

A. Violence Detection Using Machine Learning Techniques

With a computer vision, the recognition of activities become an active topic. Here are the different traditional algorithms of machine

learning such as Adaboost, KNN, etc. Motion Blob (AMV) acceleration measures vectors methods for detections of fast violence or fast-moving object from video. The motion blobs have a specific shape and position, the difference between consecutive frames is computed for absolute images. After that the resulting image is binarized which is leading towards the number of motion blobs and recognized the largest one on a fight sequence and on a non-fighting scene. With this different parameter are calculated such as area, centroid, perimeter and the distance between the blobs as well. After that the blobs are characterized as fight and non-fight. The dataset of videos has thousands of clips available on YouTube. The results of the proposed method is outperformed by state of the art methods considered that are LMP, Vif, BoW (MoSIFT), BoW (SIFT), variant v-1 and variants v-2 which is used KNN, SVM and Ada boost as a classifier in terms of ROC and accuracy.

B. Violence Detection Techniques Using SVM

By using the Support Vector Machine (SVM) as a classifier techniques of violence detection are discussed. In SVM within two classes we plot the data on dimension space and differentiate it. SVM is an algorithm which is used to classify the problems using supervised learnings. SVM is widely used method in computer vision and it is also for the tasks related for binary classification. SVM is based on the kernel, that converts the input to the high dimensional spaces where the problem can be solved. There are many methods in SVM can be used. Various methods are: real-time detection of violence in crowded scenes, Bag of words framework using acceleration, action detection, BiChannel Convolutional neural network for real, time detection, GMOF framework with tracking and detection module, Multi model features framework on the base of the subclass, Solve detecting problem by dividing the Objective, in depth and clear format using ConvNet, To determine the occurrence of violent purpose extended form of IFV (Improved Fisher vector) and sliding windows, method for detection anomalies in the video, Violence detection using Oriented Violent Flow.

C. Violence Detection Techniques Using Deep Learning

The violence detection techniques use the algorithms of deep learning in the proposed frameworks. Various method which is used to used recognized the methods are co9nvolutional neural network (CNN, ConvNet) base classification. Deep learning base on neural networks. To classify the violence recognition based on the extracted features and data set by using more convolutional layers. There are many methods of violence detection which uses the of algorithms of deep learning, such types of methods are: Deep architectures for place recognition, violence scene detection using CNN and deep audio features, detect violence videos using convolutional long short-term memory, detecting human violence behaviour by integrating trajectory and deep CNN, violence detection using 3D CNN, fight recognition method.

4. RESULT ANALYSIS

In hand-defined domain; With the introduction of Hockey and Movies datasets Bermejo et al. proposed this method and achieved 90% as benchmark accuracy. Following that, Deniz et al. suggested technique using Adaboost and SVM reported 98.9% accuracy on Movies dataset. Later on, The Violent Flows (ViF), LMP methods using Random Forests classifiers, SVM and Adaboost are reported.

Moreover, in deep learning domain; Ding et al. with train/test split scheme 3D-CNN model implemented. In recent times, Serrano et al. evaluated 2D-CNN and C3D models. After that Author proposed finest approach incorporating 2D-CNN with Hough Forest features. This approach elevated accuracies to 99±0.5%, 94.6±0.6% for Movies datasets and Hockey correspondingly, setting the accuracy bar to the next level.

The number of frames were added of violence and non-violence to detect the state of situation are violent or not.

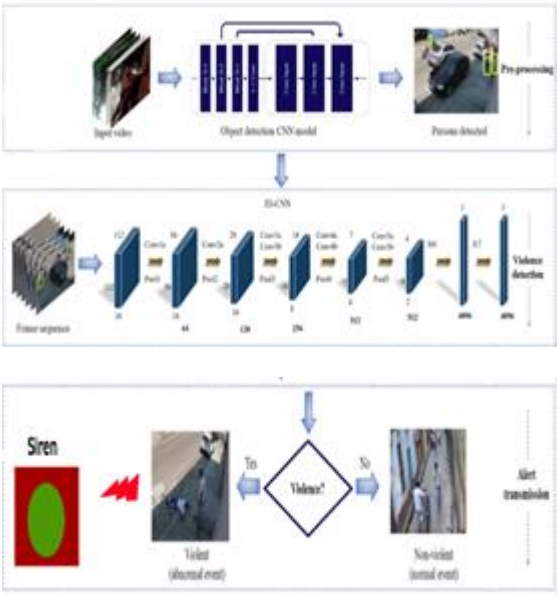

Figure 1: Violence Detection Technique

There are 25 frames of violence benchmark when its recognized 25 frames or more the siren will sound and then it generates the graph and the

graph is knowns as violence graph as shown in figure 1.

Figure 2: Violence Graph

If it counts the number of frames below 25 or then the siren will not be sound and graph will generate and the graph is knowns as non-violence graph as shown in figure 2.

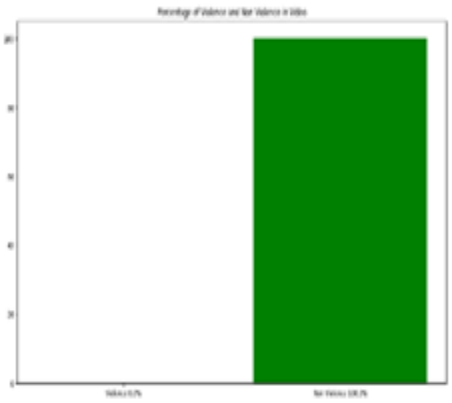

Figure 3: Non-violence Graph

CONCLUSION

With the increase of surveillance cameras in different fields of life to monitor the human activity, also grow the demand of such system which recognize the violent events automatically. In computer vision, violent action detection becomes hot topic to attract new researchers. Indeed, many researchers proposed different techniques for detection of such activities from the video. The goal of this systematic review is to explore the state-of-theart research in the violence detection system. The systematic review delivers details of methods using SVM, CNN and traditional machine learning classification-based violence detection. These techniques are deliberated. Moreover, datasets and video features that used in all techniques, Accuracy is depending upon the techniques of features extraction, object recognition and classification along with dataset being used. Our study potentially contributes in highlighting the techniques and methods of violence activity detection from surveillance videos.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

FUNDING SUPPORT

The author declares that they have no funding support for this study.

REFERENCES

[1] [1] W. Sultani, C. Chen, and M. Shah, ‘‘Real-world anomaly detection in surveillance videos,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 6479–6488.

[2] [2] I. S. Gracia, O. D. Suarez, G. B. Garcia, and T.-K. Kim, ‘‘Fast fight detection,’’ PLoS ONE, vol. 10, no. 4, Apr. 2015, Art. no. e0120448.

[3] [3] O. Deniz, I. Serrano, G. Bueno, and T.-K. Kim, ‘‘Fast violence detection in video,’’ in Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP), vol. 2, Jan. 2014, pp. 478–485.

[4] [4] L. Tian, H. Wang, Y. Zhou, and C. Peng, ‘‘Video big data in smart city: Background construction and optimization for surveillance video processing,’’ Future Gener. Comput. Syst., vol. 86, pp. 1371–1382, Sep. 2018

[5] [5] C. Dhiman and D. K. Vishwakarma, ‘‘A review of state-of-the-art techniques for abnormal human activity recognition,’’ Eng. Appl. Artif. Intell., vol. 77, pp. 21– 45, Jan. 2018.

[6] [6] P. Zhou, Q. Ding, H. Luo, and X. Hou, ‘‘Violent interaction detection in video based on deep learning,’’ J. Phys., Conf. Ser., vol. 844, no. 1, 2017, Art. no. 12044.

[7] [7] S. Chaudhary, M. A. Khan, and C. Bhatnagar, ‘‘Multiple anomalous activity detection in videos,’’ Procedia Comput. Sci., vol. 125, pp. 336–345, Jan. 2018.

[8] [8] A. B. Mabrouk and E. Zagrouba, ‘‘Abnormal behaviour recognition for intelligent video surveillance systems: A review,’’ Expert Syst. Appl., vol. 91, pp. 480–491, Jan. 2018.

[9] [9] Z. Mushtaq, G. Rasool, and B. Shehzad, ‘‘Multilingual source code analysis: A systematic literature review,’’ IEEE Access, vol. 5, pp. 11307– 11336, 2017.

[10] [10] T. Zhang, Z. Yang, W. Jia, B. Yang, J. Yang, and X. He, ‘‘A new method for violence detection in surveillance scenes,’’ Multimedia Tools Appl., vol. 75, no. 12, pp. 7327–7349, 2016.