Available onlinewww.ejaet.com

European Journal of Advances in Engineering and Technology, 2022, 9(4s):316-321
International Conference on Tech Trends in Science & Engineering (ICTTSE) 2022
Suryodaya College of Engineering & Technology, Suryodaya Polytechnic, Nagpur, Maharashtra, India

Research Article **ISSN: 2394 - 658X**

Human Fight Detection

[1]Prof. Veena K. Katankar, [2]Mr. Swapnil S. Urade, [3]Mr.Tushar S. Tajne

Asst. Prof., Computer engineering, Suryodaya College of Engineering and Technology, Nagpur, India
Student Computer engineering, Suryodaya College of Engineering and Technology, Nagpur, India
[1]veenakatankar@gmail.com, [2]swapnilurade18@gmail.com, [3]Ttajne29@gmail.com

---

ABSTRACT

The detection of specific events with direct practical use such as fights or in general aggressive behavior has been comparatively less studied. Such capability may be extremely useful in some video surveillance scenarios like prisons, psychiatric centers or even embedded in camera phones. As a consequence, there is growing interest in developing violence detection algorithms. Recent work considered the well-known Bag-of- Words framework for the specific problem of fight detection. 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.

Key words: Fight detection, Violence detection algorithms etc.

INTRODUCTION

In the last few years, the problem of human action recognition from video has become tractable by using computer vision techniques, see surveys. Within this topic, there is a vast literature in which experimental results are given for recognition of human actions like walking, jumping or hand waving. However, action detection has been devoted less effort. Action detection is a related task in which only a specific action must be detected. Action detection may be of direct use in real-life applications, fight detection being a clear example. Whereas there is a number of well-studied datasets for action recognition, significant datasets with violent actions (fights) have not been made available until the work. A violence detector has, however, immediate applicability in the surveillance domain. The primary function of large-scale surveillance systems deployed in institutions such as schools, prisons and psychiatric care facilities is for alerting authorities to potentially dangerous situations. However, human operators are overwhelmed with the number of camera feeds and manual response times are slow, resulting in a strong demand for automated alert systems. Similarly, there is increasing demand for automated rating and tagging systems that can process the great quantities of video uploaded to websites.

One of the first proposals for violence recognition in video is proposed recognizing violent scenes in videos using flame and blood detection and capturing the degree of motion, as well as the characteristic sounds of violent events. Change recognizes gunshots, explosions and car-braking in audio using a hierarchical approach based on Gaussian mixture models and Hidden Markov models (HMM). Giannakopoulos et at. Also propose a violence detector based on audio features. Clarine *et al* present a system that uses a Korhonen self-organizing map to detect skin and blood pixels in each frame and motion intensity analysis to detect violent actions involving blood. Zande et al., introduced the CASSANDRA system, which employs motion features related to articulation in video and scream-like cues in audio to detect aggression in surveillance videos.

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision. The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms.

LITERATURE SURVEY

An automated visual surveillance system requires a combination of image processing and artificial intelligence techniques. Image processing techniques are providing low level image features and artificial intelligence techniques provide expert decisions [Teddy K]. Based on modeling and classification of human activities with certain rules the doubtful human behavior is detected. Due to uncertainty and complex nature of used a subspace technique to achieve fast and accurate video event detection. The statistical approach, syntactic approach, and description-based approach for hierarchical recognition and recognition of human- object interactions and group activities are discussed the motion detection, segmentation and object classification are usually grouped as low- level processing. Object tracking, Action recognition & behavior understanding and semantic description are known as high-level processing. 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. 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-the-art research in the violence detection system. The systematic review delivers details of methods using SVM, CNN and traditional machine learning classification-based violence detection.

PROPOSED WORK

Social signal processing under affective computing aims at recognizing and extracting useful human social interaction patterns. Fight is a common social interaction in real life. A fight detection system finds wide applications. This paper aims to detect fights in a natural and low-cost manner. CNNs eliminate the need for manual feature extraction, so you do not need to identify features used to classify images. The CNN works by extracting features directly from images. The relevant features are not pretrained; they are learned while the network trains on a collection of images. Machine learning is a technique or technology which is used to prepare this model from the videos in this project these videos are used to prepare this model bye contextual neural network CNN. Contextual neural network is an algorithm or technique which is create a a neural network like human brain which characterize the videos from non-violence and violence human videos to prepare this model Differentiation Algorithm we have to find out the correspondence between the images, for this we use Horn-Schnuck algorithm which is used to recognize the motion is happening or not. There is an option for browsing the images and field for showing the motion happened or not. If two images are same then displaying no motion detected. This process aims to observe the pixel movement between two sequences of image or frame and so on. 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.

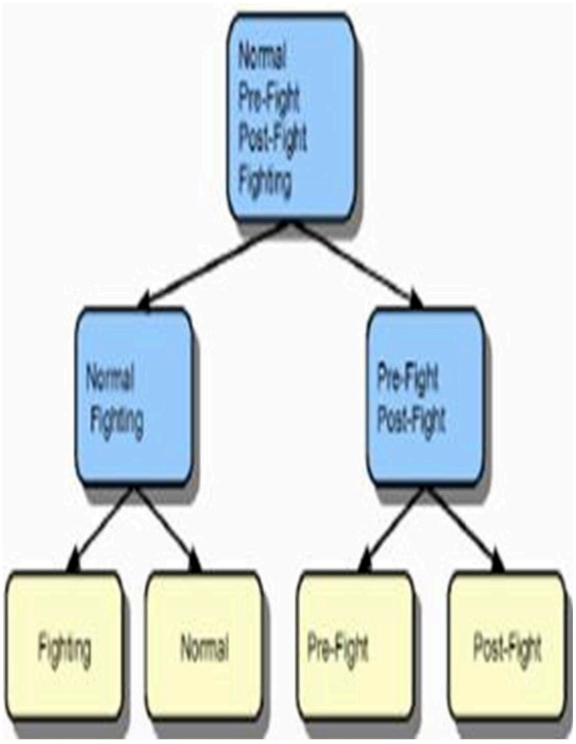

Fig. 1 Block Diagram

METHOD USED FOR IMPLEMENTATION

As can be seen the method of human aggressive movement is composed of the three steps. Preprocessing converts the input video into individual frames that is suitable for further processing. The format used in processing the raw video file is "stable". During Feature Extraction, the video file being extracted into frames and the Horn Schnuck optical flow is implemented to these frames in order to find the velocity of moving objects. The idea of motion detection is based on finding amount of difference in two consequent frames of a video sequence. According, "two frames difference" method is suitable for simple motion detection only it cannot highlight a specific region of moving objects. The experimental result in section Results and Discussion presents a set of images to help in understanding the process achieved in the present method.

PREPARE THE DATA

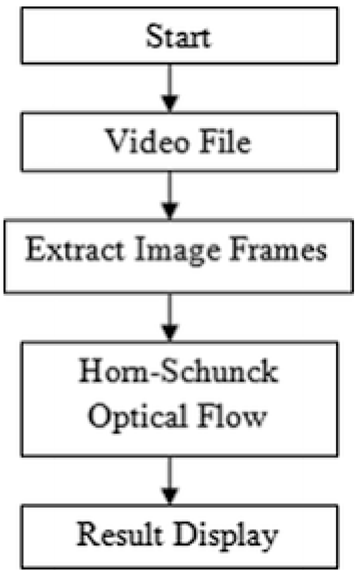

Fig. 2 Flowcharts of Human Aggressive

MOVEMENT DETECTION

Four groups of five people for 20 individuals participated in the study. The individuals were all dressed in casual attire and were aged in the range of 22 to 45. Eighteen participants were male and two were females. The groups gathered for different games of basketball finals. Three different cameras were used to record their reactions for 5 min at the beginning of each half and 5 min before the end of each half. The games usually get interesting at the beginning and towards the end. Tight games invoke more emotions compared to one-sided matches. The non-intrusive way of data capture allowed the subjects to watch the event and express their reactions without consciousness about being recorded on camera. After the data was captured, the canny edge detection was applied on every third frame from the video sequence. This caused down sampling of a 24 frame per second to an eight frames per second sequence. For each image frame, the edge detection filter was applied. Then the frame was divided into 20 x 20 meshes and the intersection of the grid lines with the edges was considered as feature co-ordinates. For a consistent feature vector length, each line was further divided into 5 divisions and the features were counted as one for an intersection and zero for no intersection. Thus, 20 x 5 features for vertical lines and 20 x 5 features for horizontal lines on the mesh, for a total of 200 feature co-ordinates were obtained to form the feature vector. Additionally, the temporal features were also tracked.

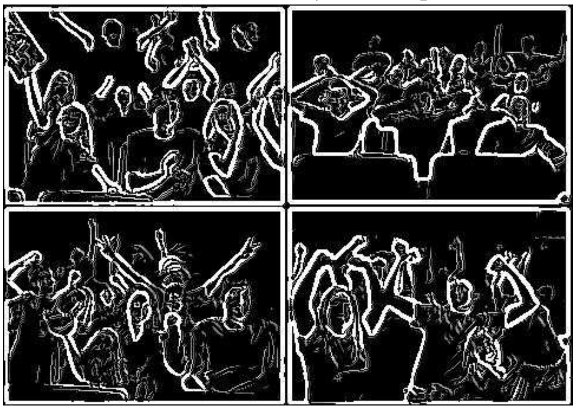

Fig. 3 Step 1 with edge detection applied to the frames

In the pre-processing step shown in Fig edge detection is applied to extract the series of high intensity change edges. These edges assisted in extracting further set of more discriminating feature points. This strategy was useful inextracting potential edges without depending on specifichuman behavioral actions, gestures and simply relied on view-independent image processing technique. Additionally, the occlusion caused by multiple people moving in the scene didnotaffecttheprocess.

As the next step in the feature extraction process, the optimized grid was applied and super-imposed on each frame to find the intersections between the grid lines and the detected edges. Each intersection point was used as thefeature for the frame. Next, the movement of each of thesefeatures was measured across the eight frames to find thetemporal, kinetic and motion pattern of the feature. This allowed the method to rely on a limited yet discriminating set of features instead of tracking individual human activities.

The method also allowed overcoming the limitation of wearing tracking devices to detect each human in the scene. It also allowed view-independent and occlusion resistance mechanism and purely relied on the visual layout of the scene based on the available data from the extracted features across various video frames and series of images. In addition, to limit the number of intersections between the edges and the lines of the grid only the N equally spaced points were considered. This number N was set to the same threshold as the number of blocks within the grid. For instance, if the grid size was 20 then the spacing was set to 20 as well. As a result, any other intersections were discarded. The settings for grid threshold at 5, 10, and 15 up to 50 showed that this strategy did not affect the overall accuracy results and the N simply contributed in limiting the feature vector size and eliminating the redundant tracked points.

The above figure shows the super-imposed grid and theintersection of the grid lines with the edges extracted fromthe first pre-processing step. The next figure shows the processing work flow to extract the edges, then the application of grid, mesh and extraction of potential points for tracking. This step is followed by tracking the motion of the points across the series of video frames.

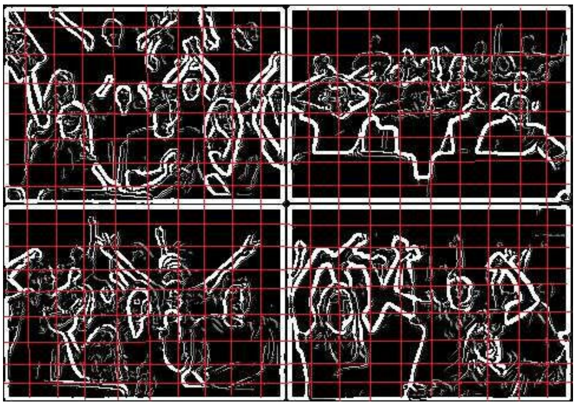

Fig. 4 Step 2 with the mesh applied to the edges

The movement of each co-ordinate was tracked across the eight frames. Thus, the final feature vector consisted of 200 static and 200 velocity values across eight frames for 400 features. The best first search technique was used for feature selection. This resulted in reduced dimensionality of feature vector size and 23 discriminating features were chosen. The sequence of all actions was annotated using three observers to avoid inter-annotator disagreement.

The classes used were happy, angry, surprised, sad, disgust, fear and neutral. After the annotation was done, the feature vector was used to train the classifiers using support vector machine and radial basis function as the kernel function with 0.4 as the slack

CONCLUSION

The action-based recognition techniques cannot be applied in such scenarios. The techniques mentioned in the paper showed promising results to overcome this limitation of view-dependence and lack of sufficient training data. This paper mostly looked into indoor scenes and a limited set of outdoor spontaneous scenes with crowds of people in the scene reacting emotionally expressive manner for sports events. As a future scope, the study could be extended in outdoor bigger crowd settings and a comparative study could be done between various techniques with the processing steps described in this paper.

Differentiation Algorithm we have to find out the correspondence between the images, for this we use Horn-Schnuck algorithm which is used to recognize the motion is happening or not. There is an option for browsing the images and field for showing the motion happened or not. If two images are same then displaying no motion detected. This process aims to observe the pixel movement between two sequences of image or frame and so on. 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.

REFERENCES

[1]. W. N. W. Samsudin, K. H. Ghazali, M. F. M. Jusof.: Analysis of Motion Detection Using Social Force Model. 25–26 (2013).

[2]. C. Anderson, B. Bushman.: Human Aggression. 27–5 1 (2002).

[3]. J. Lynn &S. Baron-cohen.: Psychology Relational Aggression. (2014).

[4]. E. F. Torrey, J. Stanley, J. Monahan & H. J. Steadman.: The MacArthur Violence Risk Assessment Study revisited: two views ten years after its initialpublication.

[5]. G. Wei, Z. Hou, W. Li & W. Yu.: Color Image Optical Flow Estimation Algorithm with Shadow Suppression. 2013 Seventh Int. Conf. Image Graph. 423–427 (2013).

[6]. S. Ramli, M. Talib, T. Khalisah, N. Zainudin.: Implementing Optical Flow Algorithm for Human Aggressive Movement Detection. 264–268

[7]. B. K. P. Horn, & B. G. Schunck.: Determining Optical Flow: a retrospective. Artif. Intell. 59, 81– 87 (1993).

[8]. D. Kesrarat, & V. Patanavijit.: Tutorial of Motion Estimation Based on Horn Schunk Optical Flow Algorithm in MATLAB ® Horn-Schunk Algorithm (HS). 15, 8–16 (2011).

[9]. D. Raju &P. Joseph.: Motion Detection and Optical Flow. 5, 5716–5719 (2014).

[10]. Tianyi Hong, Huabiao Qin, Qianshu Sun, "An improved real time eye state identification system in driver.

[11]. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Dynamic Probabilistic Network Based Human Action Recognition, Computer Science and Emerging Research Journal, vol 4, 2016.