Situation Analysis Based on Object Detection and Recognition in Video Surveillance

[1] Ms. Sandhya Karkhele, [2] Prof. Mrs. Madhura Sanap

[1] *Student*, [2]*Assistant Professor*

Department of Computer Engineering,

Sinhgad Academy of Engineering.Savitribai Phule Pune University,

Pune, Maharashtra.

Abstract : The demand for automatic action recognition systems has increased due to the rapid increase in the number of video surveillance cameras installed in cities and towns. Automatic action recognition system can be effectively used to generate on-line alarm in case of abnormal activities to assist human operators and for offline inspection. Although the action recognition problem has become a hot topic within computer vision, detection of violent scenes receives considerable attention in a surveillance system which is justified by the need of providing people with safer public spaces. This survey discusses the current state of the art methods and techniques that are being applied for the task of automated detection of fight, gun, fire. This survey emphasizes on motivation and challenges of this very recent research area by presenting approaches to fight recognition in the surveillance video. gun recognition in the surveillance video and fire recognition in the surveillance video and show update result to dropbox. This paper aims at being a driving force for researchers who wish to approach the study of different activity recognition and gather insights on the main challenges to solve in this emerging field.

Index Terms - Vision Systems, Fire Detection, Smart Cameras, Computer Vision, Object Detection.

I. INTRODUCTION

There are number of equipment for monitoring in video surveillance .Many fields are using this tolls such as terrorist attacks, unusual behaviours, bomb placement, traffic-related issues, ATM attacks etc. Conventional smoke sensors have complexity detecting anomaly in open spaces. This system chiefly proposed for monitoring woodland fires automatically through video processing. For reducing fire damage in forest video surveillance is needed.. Flames or smoke are very useful to recognize the forest fire is happened. As compare to flames and smoke that flames may not be visible to the monitoring camera when flames happen a long distance or are concealed by obstacles like mountains or buildings. In the forest smoke is useful for detecting forest fire but it is not good for images because it does not have a distinct shape or color patterns.

Typically, there are two methods for detection fire which are smoke detection and flame detection. Smoke detection methods are useful for color and motion information from digital image detect flames from video images and detect flames using IR images. In [1], Toreyin et al. proposed background subtraction, and temporal and spatial wavelet transformation for smoke detection method based. Using wavelet transforms identify smoke based on The area of decreased high frequency energy component . In [2], For the segment of smoke regions from images fractal encoding technique is used and the again those regions classify based on self-similarity of smoke boundary shapes In Smoke is detected using features like speed, column growing, volume, and height. It is pretended that the camera is stunning on a pan/tilt device. This method consist three steps. The first step is to identify whether the camera is moving or not. While the camera is moving, we do not perform the further steps. The second is regions of interest, between backdrop image and current input frame; that is, to extract changed regions against the background. Representing as blobs by connected components to The changed regions. The blobs in close adjacency are combining together with one another. The block based approach which is used in both the first and second step having advantages which are speed and robustness. The final step is to evaluate, using temporal information of color and shape in the detected blobs, whether each blob of the current input frame is smoke.

2. LITERATURE SURVEY

Ismael Serrano, Oscar Deniz, Jose L [1],outlined the Fight acknowledgment in video using Hough Forests and 2D Convolutional Neural Network. One of the first proposals for violence recognition in video. One of the first proposals for violence recognition in video. investigational results have been obtained for recognition of actions such as walking, jogging, pointing or hand waving [4]. However, action detection has been devoted comparatively less effort. Violence detection is a task that can be leveraged in real-life applications. While there is a large number of studied datasets for action recognition, specific datasets with a relevant number of violent sequences (fights) were not available until , where the authors created two specific datasets for the fight/violence problem testing state-of-the-art methods on them.

Daniel Bernhardt[3] outlined Detecting emotions from everyday body movements. The human body is a complex hierarchical structure which has evolved to enable us to perform sophisticated tasks. At the same time, movements and posture of our limbs, head and torso communicate affect and inter-personal attitudes. To a large extent our functioning as socially intelligent individuals relies on our ability to decode the affective and expressive cues we perceive through facial or body gestures. Research suggests that our responses to avatars in Immersive Virtual Environments (IVEs) are governed by our expectations about the presence and correct exhibition of those expressive cues.

Ginevra Castellano [5] outlined the Recognizing Human Emotions from Body Movement and Gesture Dynamics. One critical aspect of human-computer interfaces is the ability to communicate with users in an expressive way. Computers should be able to recognize and interpret users' motional states and to communicate expressive-emotional information to them. Recently, there has been an increased interest in designing automated video analysis algorithms aiming to extract, describe and classify information related to the emotional state of individuals. In this paper we focus on video analysis of movement and gesture as indicators of an underlying emotional process. Our research aims to investigate which are the motion cues indicating differences between emotions and to define a model to recognize emotions from video analysis of body movement and gesture dynamics

Domenico D. Bloisi [6] A paper proposed on "Online real-time crowd behavior detection in video sequences" state an algorithm called FSCB. FSCB is made of three main steps: (1) Feature detection and temporal filtering; (2) image Segmentation and blob extraction; (3) Crowd Behavior detection. In this paper, a real-time and online crowd behavior detection algorithm for video sequences is described. The algorithm, called FSCB, is based on a pipeline made of the following stages: (1) stable features are tracked between frames of the sequence; (2) a temporal mask is extracted; (3) moving blobs are found using segmentation; (4) anomalous events are detected using two measures, i.e., instant entropy and temporal occupancy variation. Quantitative experiments have been conducted on different publicly available data sets: UMN, PETS2009, and AGORASET. For PETS 2009 and AGORASET, ground truth data have been produced and made available at the FSCB website. Furthermore, a novel annotated data set, containing crowded scenes from the start of a marathon, has been created. FSCB has been quantitatively compared with other stateof-the art methods for online crowd event detection. The results of the comparison demonstrate the effectiveness of the proposed approach, that works without the need of a training stage and obtain real-time performance on 320×240 images .

H. Yeh, C. Y. Lin, K. Muchtar, H. E. Lai and M. T. Sun [7] A paper proposed on "Three-Pronged Compensation and Hysteresis Thresholding for Moving Object Detection in Real-Time Video Surveillance" proposed moving object detection method. This method is a three-pronged approach to compensate in order to extract foreground objects as complete as possible. First, use a texture background modeling method, which only detects the texture of the foreground object but can resist illumination changes and shadow interference. Second, apply hysteresis thresholding on both texture and color background models to generate predominant and supplementary images. The combination of predominant images shows the skeleton of moving objects, PCT. Then use several supplementary images to mend the shape of PCT with the goal of completing moving object extraction without shadows. Finally, the proposed motion history applies spatial-temporal information to alleviate the cavity and fragment problems in foreground objects. The combined approach thereby offers a three-pronged compensation by leveraging texture, color, and spatial-temporal information .

S. Coşar, G. Donatiello, V. Bogorny, C. Garate, L. O. Alvares and F. Brémond,[8] A paper proposed on "Towards Abnormal Trajectory and Event Detection in Video Surveillance" focused on trajectory-based and pixel based approaches for unsupervised abnormal behavior detection. A) Object and Group Tracking: As the first step of this approach, it take the input video and extract all trajectories in the scene. In this step, it run the object tracking algorithm and group tracking algorithm to generate all individual trajectories of objects/groups moving in the scene. B) Grid-Based Analysis: This step takes the extracted trajectories and bounding boxes of each object as input and performs grid-based analysis. In the grid-based analysis, three main steps are performed: trajectory snapping, zone discovery, and trajectory-based anomaly detection

3. PROPOSED WORK

The most popular and probably the simplest way to detect faces using Python is by using the OpenCV package.

The algorithm may have 30 to 50 of these stages or cascades, and it will only detect a object if all stages pass. one more library file use alternatives to OpenCV, that is dlib – that come with Deep Learning based Detection and Recognition models. The most popular and probably the simplest way to detect faces using Python is by using the OpenCV package.

The algorithm may have 30 to 50 of these stages or cascades, and it will only detect a object if all stages pass.one more library file use alternatives to OpenCV, that is dlib – that come with Deep Learning based Detection and Recognition models

The proposed work will be in the form of modules,

Module 1 : Image module/ Web Camera module / Video module/ IP Camera

Module 2: Face detection

Module 3 : Activity Detection real time module

Module 3 .1: Weapon /Gun detection

Module 3 .2: Fire detection and smote detection

Module 3 .3: Fight detection

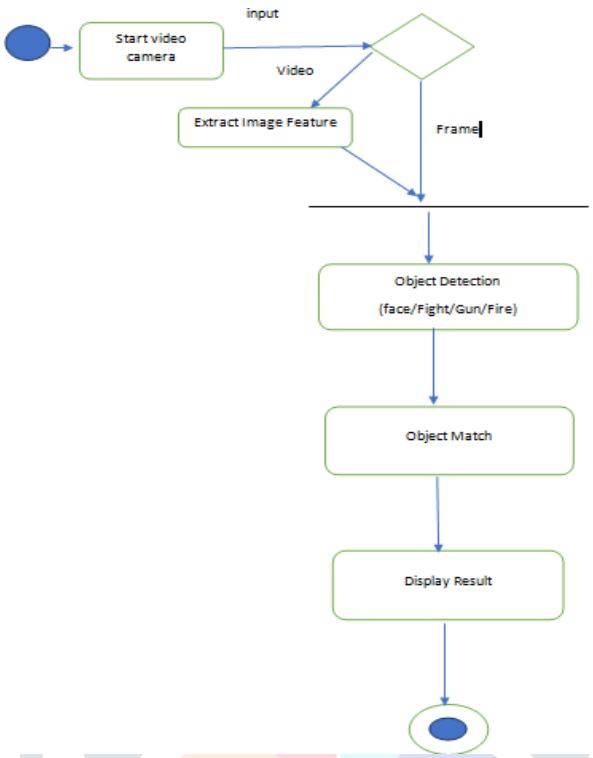


Fig 1. System flow diagram

4 RELEVANT MATHEMATICS ASSOCIATED WITH THE PROJECT

System Description through Mathematical Model

S={S, F, I, O, F, T, DD, NDD, Success, Failure}

• S: Initial Stage.

• F: Final Stage

• I : Input: Video /IMAGE / WEB / IP [5].

• O : Output: Action Response & Completed Task.

•F: Functions: IMAGE,VIDEO,WEB CAMERA,IP CAMERA (DVR/ MOBILE IP) ,Set Alarm, SCREENSHOPOT ,ETC,

• T : Steps

1. Take Image/ Video /IP Camera (live Streaming) from user.

2. Converted that voice command to text.

3. Search answer(Datasets) for video input on server.

4. Server send the answer(Screen shoot) to the client.

5. Take Action perform the activity.

• DD : Deterministic Data

• NDD : Non Deterministic Data

• Success Conditions: Understood Input (video/ Image) Command Properly And Task

Completed Successfully.

• Failure Conditions: Video /Input Command not Understood and given job not completed. ,

Internet on for activity Detection.

4.1 CLASSIFICATION OF FIRE

From each BLOB (t, i), i = 0, ..., N (t) -1, we calculate the characteristics that include the area, the bounding rectangle, the mean and the standard deviation of the Y value, and the average and the UV value of the standard deviation. The statistics of the UV values can be calculated from the current input image I (t, x, y). We maintain the characteristics, F (t), F (t -1), F (t - 2), ..., F (t - k -1) which are calculated from k previous time frames in which k is a dimension for the spot tracking. Where F (t) are the characteristics calculated by BLOBs segmented at time t. For each BLOB (t, i), i = 0, ..., N (t) -1 of instance t, we conclude whether it is burn or not. We classify as smoke if it continually changes form and area and has similar statistics in the Y value in all k frames.

Fig.-2 Fire detections in video

4..2 FIGHT RECOGNITION

One of the first proposals for the recognition of violence in video is Nam et al. [18], which proposes to recognize violent scenes in video using the detection of flames and blood and to capture the degree of movement, as well as the characteristic sounds of violent events. Cheng et al. [5] recognizes car shots, explosions and audio brakes using a hierarchical approach based on Gaussian mixing models and Hidden Markov models (HMM). Giannakopoulos et al. [10] also proposes a violence detector based on audio characteristics. Clarin et al. [6] presents a system that uses a self-organizing Kohonen map to detect skin and blood pixels in each frame and analysis of movement intensity to detect blood-related violent actions. Zajdel et al. [19], presents the CASSANDRA system, which uses motion functions related to video articulation and signals similar to audio cries to detect aggression in surveillance videos. More recently, Gong et al. [11] to propose a violence detector that uses low-level visual and auditory functions and high-level audio effects that identify the potential for violent content in films. Chen et al. [3] uses binary local motion descriptors (space-time video cubes) and a word-bag approach to detect aggressive behavior. Lin and Wang [15] describe a weakly supervised audio violence classifier that is combined with a joint workout with a video motion, an explosion, and a blood classifier to detect violent scenes in the movies. Giannakopoulos et al. [9] presents a method for detecting violence in films based on audiovisual information using audio characteristics statistics and the variance of average orientation and video motion combined in a Nearest Neighbor classifier to decide if the given sequence is violent. In summary, a number of previous jobs

require audio signals to detect violence or rely on color to detect signals such as blood. In this sense, we observe that there are important applications, in particular surveillance, where audio is not available and where the video is grayscale. Finally, while explosions, blood and rush can be useful signals for violence in action films, they are rare in real-life surveillance videos. In this paper, we focus on reliable signals for the early detection of violence in such environments

5. DATASET

Most of the widely recognized and publicly available data sets in recognition of the action, such as KTH [13], focus on individual actors who perform a simple action such as walking, jumping or greeting on a light background; These are clearly inadequate to evaluate the detection of violence. Data sets like INRIA IXMAS, which show that a single calcium or stroke can be used to train (but not evaluate) combat detection systems. Some data sets like CAVIAR, BEHAVE or Care Media contain some cases of people involved in aggressive behavior, but this is not their main goal. Our intention is to present a new set of video data specifically designed to evaluate violence detection systems, in which normal and violent activities are carried out in similar and dynamic environments. To this end, we collect 1000 action clips from the National Hockey League (NHL) hockey games, as shown in Fig. 1. Each clip consists of 50 frames of 720 × 576 pixels and is labeled manually as "Fight". "or" do not fight. "This data set allows us to easily and reliably measure the performance of a variety of violence recognition approaches, as shown in Section 5. Our wrestling data set is available upon request from authors.

6. MODULE DESCRIPTIONS

Conversion of video into image frame. Image Enhancement for object detection Feature Extraction to classify all the objects in the frame like Weapon, Face, Fire etc.

Object detection with highlighting shapes as rectangle.

- **Conversion of video into image frame**

Cascade Object Detector to detect the location of a face in a video frame. Convert the video into frame by using Cascade classifier for detect the object like weapon, fire ,face etc. which can be achieved by Opencv. Opencv library can be used to perform multiple operations on videos. Take a video as input and break the video into frame.

- **Image Enhancement for object detection**

-Image enhancement is a very common field in the area of Computer Vision. It is the extraction of meaningful information from videos or images by using Opencv. Image enhancement techniques are increasingly needed for improving object detection.

- **Feature Extraction**

Feature extraction involves reducing the amount of resources required to describe a large set of data. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

- **Fight detection**

Fight detection has been widely used for face detection, vehicle detection, pedestrian counting, web images, security systems and driverless cars. Here we will work with face detection. The algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier. A simple face tracking system by dividing the tracking problem into three separate problems:

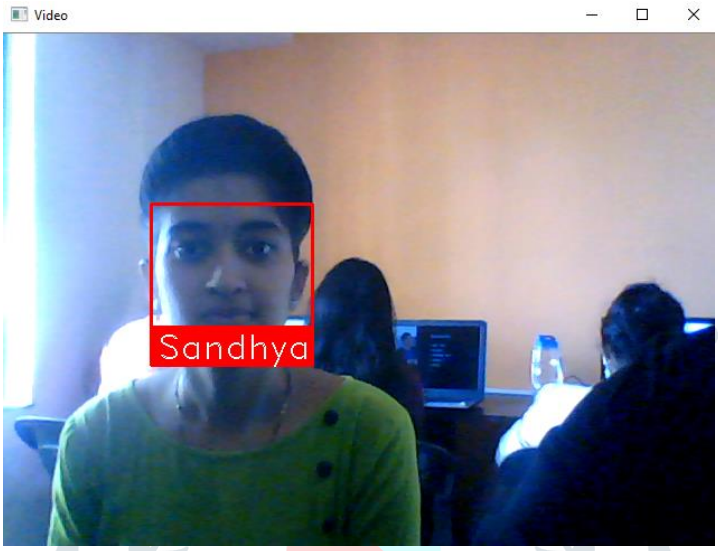
- Detect a fight to track
- Identify facial features to track
- Track the fight

- **Fire Detection**

In the fire detection mostly used edge detection algorithm detects the coarse and superfluous edges in a fire image first and then identifies the edges of the flame and removes the irrelevant artifacts . The auto adaptive feature of the algorithm ensures that the primary symbolic flame/fire edges are identified for different scenarios. Here used opencv for image capture and for accuracy purpose using tensor flow in deep learning for accuracy purpose, And 2D neural network is used for train data set

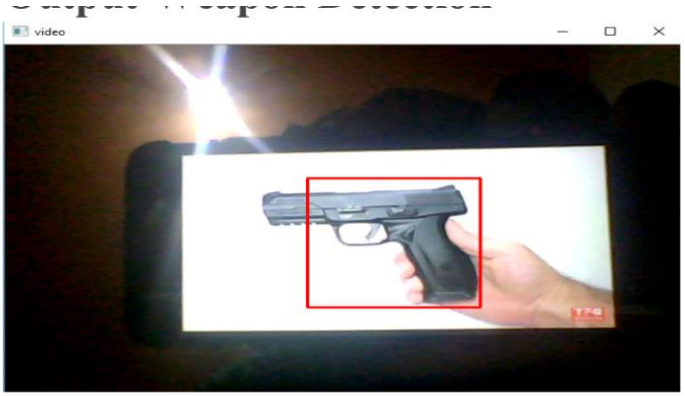
7. RESULTS

7.1 Face recognition in video



7.2 .Fire detections in video

7.3. Gun detection



7.4. Fight detection

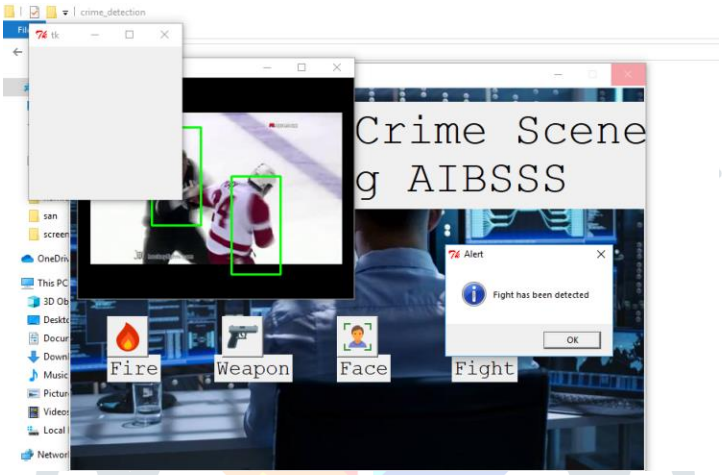


Figure-3 Detections of crimes

Result of this dataset

1) Detection in surveillance
2) Detection in video /Image

And sending output to the drop box on cloud

Result uploaded on drop box as cloud

8. CONCLUSIONS

Activity recognition could be a difficult task because of the liability of the organic structure. The survey provides a short discussion regarding the vital problems associated with the detection of activities as well as current problems, methodology. An outline of various datasets that may be used for Fight, gun and fire detection is additionally listed in the survey. Currently, several models are planned to acknowledge varied activities and specific activities associated with violence. In each model there are many limitations like some models do not work in complex environments like scenes involving crowds and large amounts of occlusion in real time which still needs to be considered by researchers in the future. However, until automatic video surveillance systems provide the same reasoning capabilities as that of monitored person are able to do there is still a lot of research ahead. It is an ultimately endless pursuit and hence continued improvement of such algorithms is important..

REFERENCES

[1] Ismael Serrano, Oscar Deniz, Jose L. Espinosa-Aranda, Gloria Bueno [2018- IEEE TRANSACTIONS

[2] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features through 3d convolutional networks. In 2015 IEEE worldwide Conference on Computer Vision (ICCV), pages 4489–4497. IEEE, 2015.

[3]DanielBernhardt,“Detecting emotions from everyday body movements) University of Cambridge.

[4] Justin Lai, “Developing a Real-Time Gun Detection Classifier”, World academy of science, Stanford University,

[5]Ginevra Castellano1, Santiago D. Villalba2, “Recognising Human Emotions from Body Movement and Gesture Dynamics”, University of Genoa

[6]AndreaPennisi, Domenico D. Bloisi, Luca Iocchi, Online real-time crowd behavior detection in video sequences, In Computer Vision and Image Understanding, Volume 144, 2016.

[7]C. H. Yeh, C. Y. Lin, K. Muchtar, H. E. Lai and M. T. Sun, "Three-Pronged Compensation and Hysteresis Thresholding for Moving Object Detection in Real-Time Video Surveillance," in IEEE Transactions on Industrial Electronics, vol. 64, no. 6, pp. 4945-4955, June 2017.

[8]S.Coşar, G. Donatiello, V. Bogorny, C. Garate, L. O. Alvares and F. Brémond, "Toward Abnormal route and Event Detection in Video Surveillance," in IEEE Transactions on circuit and system used for Video Technology, vol. 27, no. 3, pp. 683-695, March 2017

[9]MoezBaccouche, et al. Sequential deep learning for human action recognition. International Workshop on Human BehaviorUnderstanding.Springer Berlin Heidelberg, 2011.

[10]TobiasSenst, Volker Eiselein, “A Local Feature based on Lagrangian Measures for Violent Video Classification(IEEE), TechnischeUniversitat Berlin, Germany

[11]Samir K. Bandyopadhyay, Biswajita Datta, and Sudipta Roy Identifications of concealed weapon in a Human Body Department of Computer Science and Engineer, University of Calcutta, 2012.