Real-Time Object Detection for Security Surveillance Using a Hybrid Algorithm

Deepti Sagade[1], Prof.A.D.Gujar[2]
[1]*Student, Department of Computer Engineering, TSSM'S Bhivarabai sawant college of engineering & research Narhe Pune, Savitribai Phule Pune University*
[2]*Assistant professor, Department of Computer Engineering TSSM'S Bhivarabai sawant college of engineering & research Narhe, Savitribai Phule Pune University Pune, Maharashtra.*

***Abstract*: - The request for programmed activity acknowledgment frameworks has expanded due to the fast increment in the number of video observation cameras introduced in cities and towns. Programmed activity acknowledgment framework can be successfully utilized to create on-line caution in case of anomalous exercises to help human administrators and for offline assessment. In spite of the fact that the activity acknowledgment issue has gotten to be a hot subject inside computer vision, discovery of rough scenes gets significant consideration in reconnaissance framework which is legitimized by the require of giving individuals with more secure open spaces. This study examines the current state of the craftsmanship strategies and methods that are being connected for the assignment of robotized discovery of battle, weapon, fire. This overview emphasizes on inspiration and challenges of this exceptionally later inquire about region by displaying approaches to battle acknowledgment in the reconnaissance video. weapon acknowledgment in the observation video and fire acknowledgment in the observation video and appear overhaul result to local system. This paper points at being a driving drive for analysts who wish to approach the consider of diverse action acknowledgment and assemble bits of knowledge on the primary challenges to fathom in this rising field.**

Keywords:-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 colour patterns. Typically, there are two methods for detection fire which are smoke detection and flame detection. Smoke detection methods are useful for colour 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 the fragment of smoke locales from pictures fractal encoding method is utilized and the once more those locales classify based on self-similarity of smoke boundary shapes In Smoke is identified utilizing highlights like speed, column developing, volume, and stature. It is imagined that the camera is dazzling on a pan/tilt gadget. This strategy comprises three steps. The to begin with step is to recognize whether the camera is moving or not. Whereas the camera is moving, we do not perform the advance steps. The moment is locales of intrigued, between background picture and current input outline; that is, to extricate changed locales against the foundation. Speaking to as blobs by associated components to The changed districts. The blobs in near contiguousness are combining together with one another. The square based approach which is utilized in both the to begin with and moment step having preferences which are speed and vigor. The last step is to assess, utilizing worldly data of colour and shape in the recognized blobs, whether each blob of the current input outline is smoke.

II.LITERATURE SURVEY

Ismael Serrano, Oscar Deniz, Jose L [1], sketched out the Battle affirmation in video utilizing Hough Woodlands and 2D Convolutional Neural Organize. One of the to begin with proposition for viciousness acknowledgment in video. One of the to begin with recommendations for savagery acknowledgment in video. investigational comes about have been gotten for acknowledgment of activities such as strolling, running, indicating or hand waving [4]. In any case, activity discovery has been given comparatively less exertion. Viciousness location is a errand that can be utilized in real-life applications. Whereas there is a expansive number of considered datasets for activity acknowledgment, particular datasets with a important number of savage arrangements (battles) were not accessible until , where the creators made two particular datasets for the fight/violence issue testing state-of-the-art strategies on them.

Daniel Bernhardt[3] laid out Identifying feelings from regular body developments. The human body is a complex various leveled structure which has advanced to empower us to perform modern errands. At the same time, developments and pose of our appendages, head and middle communicate influence and inter-personal demeanors. To a huge degree our working as socially cleverly people depends on our capacity to interpret the full of feeling and expressive signals we see through facial or body signals. Investigate proposes that our reactions to avatars in Immersive Virtual Situations (IVEs) are represented by our desires almost the nearness and rectify presentation of those expressive signals

Ginevra Castellano [5] laid out the Perceiving Human Feelings from Body Development and Motion Flow. One basic perspective of human-computer interfacing is the capacity to communicate with clients in an expressive way. Computers ought to be able to perceive and translate users' motional states and to communicate expressive-emotional data to them. As of late, there has been an expanded intrigued in planning robotized video investigation calculations pointing to extricate, depict and classify data related to the passionate state of people. In this paper we center on video examination of development and signal as pointers of an fundamental enthusiastic prepare. Our inquire about points to explore which are the movement signals demonstrating contrasts between feelings and to characterize a show to perceive feelings from video examination of body development and signal flow

Domenico D. Bloisi [6] A paper proposed on "Online real-time swarm behavior location in video sequences" state an calculation called FSCB. FSCB is made of three fundamental steps: (1) Highlight discovery and worldly sifting; (2) picture Division and blob extraction; (3) Swarm Behavior discovery. In this paper, a real-time and online swarm behavior discovery calculation for video groupings is depicted. The calculation, called FSCB, is based on a pipeline made of the taking after stages: (1) steady highlights are followed between outlines of the arrangement; (2) a worldly cover is extricated; (3) moving blobs are found utilizing division; (4) odd occasions are recognized utilizing two measures, i.e., moment entropy and transient inhabitance variety. Quantitative tests have been conducted on diverse freely accessible information sets: UMN, PETS2009, and AGORASET. For PETS 2009 and AGORASET, ground truth information have been delivered and made accessible at the FSCB site. Moreover, a novel explained information set, containing swarmed scenes from the begin of a marathon, has been made. FSCB has been quantitatively compared with other stateof-the craftsmanship strategies for online swarm occasion location. The comes about of the comparison illustrate the viability of the proposed approach, that works without the require of a preparing arrange and get real-time execution on 320×240 pictures.

H. Yeh, C. Y. Lin, K. Muchtar, H. E. Lai and M. T. Sun [7] A paper proposed on "Three-Pronged Remuneration and Hysteresis Thresholding for Moving Protest Location in Real-Time Video Surveillance" proposed moving question discovery strategy. This strategy is a three-pronged approach to compensate in arrange to extricate closer view objects as total as conceivable. To begin with, utilize a surface foundation modeling strategy, which as it were identifies the surface of the closer view protest but can stand up to brightening changes and shadow obstructions. Moment, apply hysteresis thresholding on both surface and color foundation models to produce transcendent and supplementary pictures. The combination of overwhelming pictures appears the skeleton of moving objects, PCT. At that point utilize a few supplementary pictures to repair the shape of PCT with the objective of completing moving protest extraction without shadows. At last, the proposed movement history applies spatial-

temporal data to reduce the depression and part issues in closer view objects. The combined approach in this manner offers a three-pronged stipend by leveraging surface, color, and spatial-temporal data.

S. Coşar, G. Donatiello, V. Bogorny, C. Garate, L. O. Alvares and F. Brémond,[8] A paper proposed on "Towards Unusual Direction and Occasion Discovery in Video Surveillance" centered on trajectory-based and pixel based approaches for unsupervised anomalous behavior discovery. A) Protest and Gather Following: As the to begin with step of this approach, it take the input video and extricate all directions in the scene. In this step, it run the protest following calculation and gather following calculation to create all person directions of objects/groups moving in the scene. B) Grid-Based Investigation: This step takes the extricated directions and bounding boxes of each question as input and performs grid-based investigation. In the grid-based examination, three primary steps are performed: direction snapping, zone revelation, and trajectory-based inconsistency location

III.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 modelsThe 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, thatis 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
Module 2 : Web Camera module
Module 3 : Video module
Module 4 : IP Camera model
Module 5 : Pose Detection
Module 6 : Activity Detection real time module
Module 6 .1: Weapon /Gun detection
Module 6 .2: Fire detection and smote detection
Module 6 .3: Fight detection

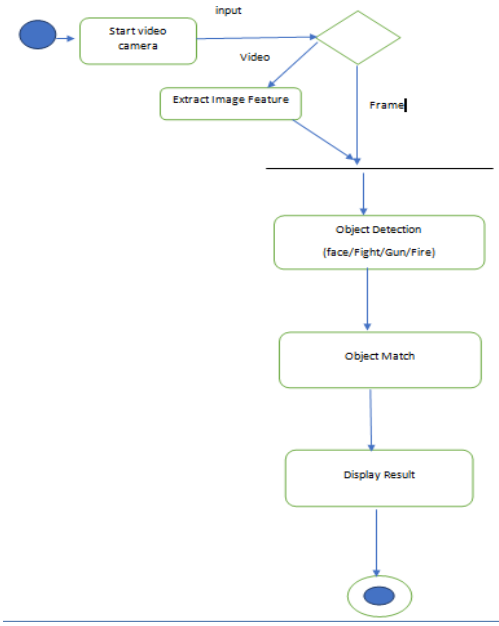


Fig 1 . System flow digram

A.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 Striming) 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.

B. 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.

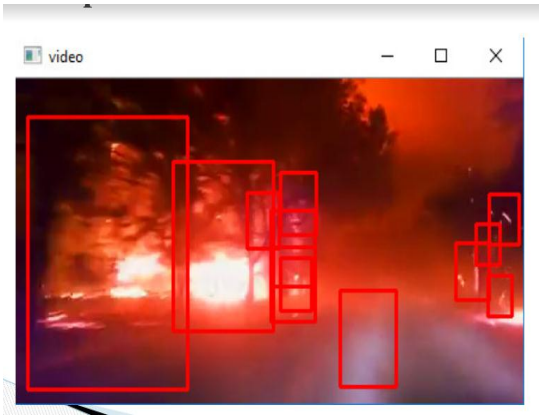


Fig. 2 fire detections in video

IV.FIGHT RECOGNITION

One of the to begin with proposition for the acknowledgment of viciousness in video is Nam et al. [18], which proposes to recognize savage scenes in video utilizing the location of blazes and blood and to capture the degree of development, as well as the characteristic sounds of rough occasions. Cheng et al. [5] recognizes car shots, blasts and sound brakes utilizing a progressive approach based on Gaussian blending models and Covered up Markov models (Well). Giannakopoulos et al. [10] moreover proposes a savagery locator based on sound characteristics. Clarin et al. [6] presents a framework that employments a self-organizing Kohonen outline to identify skin and blood pixels in each outline and examination of development concentrated to identify blood-related savage activities. Zajdel et al. [19], presents the CASSANDRA framework, which employments movement capacities related to video enunciation and signals comparative to sound cries to identify animosity in reconnaissance recordings. More as of late, Gong et al. [11] to propose a viciousness finder that employments low-level visual and sound-related capacities and high-level sound impacts that distinguish the potential for savage substance in movies. Chen et al. [3] employments twofold nearby movement descriptors (space-time video 3d shapes) and a word-bag approach to distinguish forceful behavior. Lin and Wang [15] portray a pitifully administered sound viciousness classifier that is combined with a joint workout with a video movement, an blast, and a blood classifier to identify rough scenes in the motion pictures. Giannakopoulos et al. [9] presents a strategy for recognizing savagery in movies based on varying media data utilizing sound characteristics measurements and the change of normal introduction and video movement combined in a Closest Neighbor classifier to choose if the given grouping is rough. In outline, a number of past occupations require sound signals to identify viciousness or depend on color to identify signals such as blood. In this sense, we watch that there are critical applications, in specific reconnaissance, where sound is not accessible and where the video is grayscale. At last, whereas blasts, blood and surge can be valuable signals for savagery in activity movies, they are uncommon in real-life reconnaissance recordings. In this paper, we center on dependable signals for the early location of viciousness in such situations

V.DATASET

The fire detection dataset used in this study consists of annotated images collected from publicly available fire datasets, open-source platforms, and real-world surveillance footage. The dataset is structured following the YOLOv5 convention, comprising three main subsets: training, validation, and testing. Images are organized into folders under images/train, images/val, and images/test, with corresponding YOLO-style annotation files located in the labels directory. Each label file contains bounding box coordinates in the format: class_id center_x center_y width height, where all values are normalized between 0 and 1. The dataset focuses exclusively on a single class, "fire", assigned the class ID 0. The collected images vary in resolution but are resized to 416×416 pixels during model training to ensure uniformity and efficient processing. This resizing also helps balance detection accuracy with computational resource usage. In total, the dataset comprises approximately XXX annotated images, capturing fire scenarios across different lighting conditions, environments (indoor and outdoor), and intensity levels. These annotations are carefully curated to represent small flames, medium-scale fires, and large blaze events, enabling the model to learn robust fire features across varying sizes and contexts.

VI.MODULE DESCRIPTION

This system is designed for intelligent surveillance and object detection using advanced computer vision techniques. The core modules involved in this project include conversion of video to image frames, image enhancement for better object visibility, feature extraction to distinguish different objects, and accurate detection of specific classes such as weapons, faces, fire, and fights, with results highlighted using bounding boxes.

1.Conversion of Video into Image Frames:
Video input is processed frame-by-frame using the OpenCV library. Each frame is treated as a static image, making it easier to apply object detection algorithms. Techniques such as Haar Cascade Classifiers and deep learning-based detectors are used to localize and identify objects like faces, weapons, and fire in each frame. This enables efficient and real-time surveillance by breaking the video into processable snapshots.

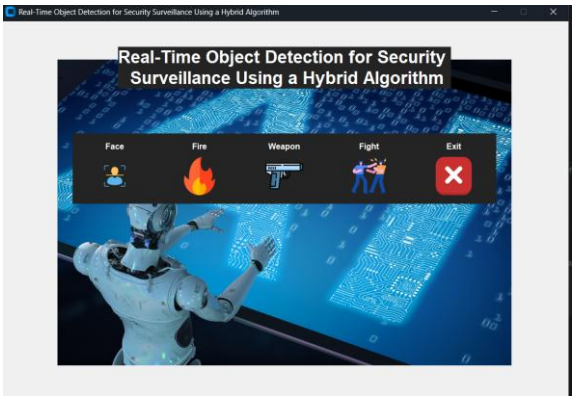
2.IMAGE ENHANCEMENT FOR OBJECT DETECTION:
Image enhancement techniques are applied to improve the quality of the frames before feeding them into the object detection models. Enhancements like brightness normalization, contrast adjustment, and filtering are employed to make key features more distinguishable. This step plays a crucial role in improving detection accuracy, especially in low-light or noisy environments.

3.FEATURE.EXTRACTION:
Feature extraction is performed using deep learning models trained on labeled datasets. The extracted features help distinguish between different objects such as fire, weapons, or faces. This step is critical for reducing computational complexity and improving detection precision. By isolating meaningful features, the system ensures that only relevant information is passed on to the classification stage.

4.FIGHTDETECTION:
Fight detection is another vital aspect of this system, aimed at identifying violent activities in real-time. The detection algorithm analyzes sudden motion patterns, human posture, and interactions in the scene. By training on datasets containing both violent and non-violent clips, the model can distinguish suspicious behavior. This module is useful in monitoring public places, schools, and sensitive areas. Each detected object is visually highlighted with a bounding rectangle on the video frame, making it easy to interpret and track events. The system ensures real-time alerts and can be extended to trigger sound alarms or send notifications based on critical object detection like fire or weapons.

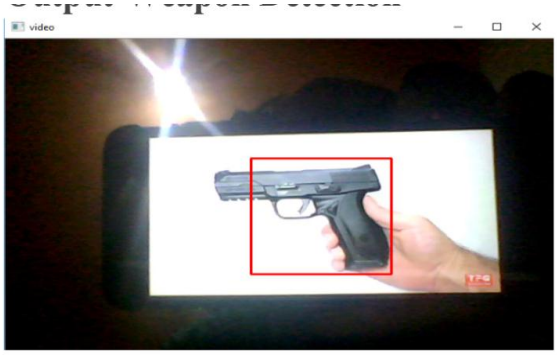
VII. RESULT

1 UI SCREEN



2 .fire detections in video

3. Gun detection



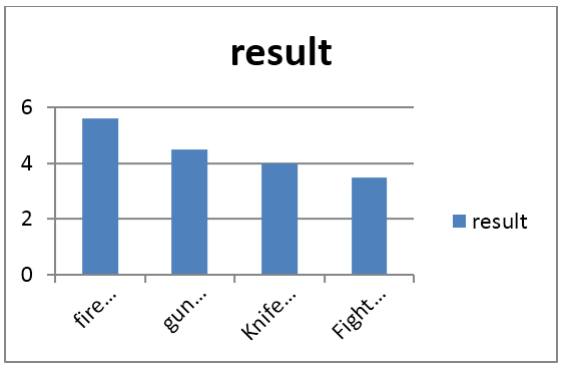


Figure :Detections of crimes

Table

Detections	*Values (Accuracy)*
Fire Detection	5.6
Gun Detection	4.5
Knife detection	4.4
Fight Detection	3.5

Table 1.1 :Detections of crimes

Result of this dataset

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

Result uploaded on local system

VIII. CONCLUSIONS

The AI-based Security Surveillance System developed in this project demonstrates an effective and real-time solution for enhancing safety across various environments such as schools, public places, offices, and industries. By integrating multiple detection modules—including face recognition, fire detection, weapon identification, and fight detection—the system offers a comprehensive approach to proactive surveillance. Each module is optimized for accuracy, speed, and real-time alerting, making the solution not only robust but also scalable. The use of deep learning models such as YOLO for object detection and CNNs for classification ensures high precision in identifying threats, while the GUI-based control panel improves user accessibility and control. The system also supports real-time video input, processes frames efficiently, and raises alerts when a threat is detected, helping reduce response time during emergencies. With further training on larger datasets and deployment in real environments, this solution can significantly contribute to crime prevention, early hazard detection, and public safety.

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 Behavior Understanding.Springer Berlin Heidelberg, 2011.

[10] TobiasSenst, Volker Eiselein, "A Local Feature based on Lagrangian Measures for Violent Video Classification(IEEE), Technische Universitat 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.