



# Intelligent Video Surveillance System Based On Machine Learning

Akshata Suryawanshi, Sukanya Suryawanshi, Ashish Bhosale, Anuradha Yadav, Prof.Nihal Shikalgar

Department of Information Technology, Zeal College Of Engineering, Pune, India

**ABSTRACT:** Now a day's million monitoring cameras have been equipped for surveillance systems in all over world. So we have implemented video surveillance by giving video contents containing early fire events detection, abnormal activities and smart parking system and crowd estimation. We have overcome existing drawbacks of post investigation techniques of video surveillance systems by providing pre alert generation system. Our work is based on machine learning techniques for video analysis with better performance and event detection with advantages of alert generation. Video surveillance system has become a important part in the security and protection of modem cities. Since Video surveillance system has become a critical part in the security and protection of modem cities, since smart monitoring cameras equipped with intelligent video analytics techniques can monitor and pre-alert system by capturing abnormal activity and events. Recent years, more and more video surveillance devices are deployed as the increasing demands on public security and smart city.

**KEYWORDS:** Intelligent Video Surveillance, Fire Detection, Deep neural network.

## I. INTRODUCTION

To intensive task of monitoring surveillance regions as well as explore richly valuable information from the big surveillance data, researchers seek the advanced computer vision algorithms to develop intelligent video surveillance (IVS). Motive behind proposed work is to parsed meaningful structured information from the raw non-structured video. The Video Surveillance and Monitoring has become the richest source of security and investigation. In which motion detection, object recognition, tracking, and some are higher level analysis modules for specific applications, e.g., people counting, activity recognition etc. The current world is completely under CCTV or video surveillance systems. The video recorded is used to find out robbery investigation, crime investigation and abnormal activity detection. After event happened these video sequences is used to catch criminals. But problem is that after event happened we are unable to save loss done by that event or accidents. So there is need of such systems which is able to early event detection and pre alert generation systems.

## II. LITERATURE SURVEY

### 1. "Traffic Monitoring using Video Stream with Machine Learning: Based on Big Data Process with Cloud".

Patel Parin and Gayatri Pandi [1] states Traffic Monitoring is a difficult undertaking on jam-packed streets. Traffic Monitoring methods are manual, costly, time expending and include human administrators. Huge scale stockpiling and investigation of video streams were unrealistic due to constrained accessibility. Notwithstanding, it is currently conceivable to actualize object discovery and following, conduct examination of traffic designs, number plate acknowledgment and reconnaissance on video streams created by traffic checking. In Big Data, video streams (datasets) are enormous to such an extent that average database frameworks are not ready to store and examination the datasets. Putting away and preparing huge volume of information requires adaptability, Fault Tolerance and accessibility. In this manner, Big Data and Cloud figuring are two perfect ideas as cloud empowers Big Data for traffic observing utilizing Hadoop innovation with AI calculation. Examination results are put away in Hive, which is an information stockroom based over Hadoop. This information incorporates by and large traffic speed, traffic volume, and individual vehicle speed. Hive gives information rundown, question, and investigation. To dodge the necessities of path control and sensor exhibits, we propose information driven way to deal with street traffic investigation utilizing advanced video.

### 2. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks."

Shaoqing Ren et.al [2] introducing best in class object recognition systems rely upon area proposition calculations to speculate object areas. Advances like SPPnet and Fast R-CNN have diminished the running time of these discovery

systems, uncovering area proposition calculation as a bottleneck. In this work, we present a Region Proposal Network (RPN) that offers full-picture convolutional highlights with the recognition arrange, along these lines empowering almost without cost district recommendations. A RPN is a completely convolutional arrange that at the same time predicts article limits and objectness scores at each position. RPNs are prepared start to finish to create highquality district recommendations, which are utilized by Fast R-CNN for location. With a basic substituting streamlining, RPN and Fast R-CNN can be prepared to share convolutional highlights.

### 3. “Multi-Object Tracking with Quadruplet Convolutional Neural Networks”

Jeany Son et.al [3] proposed Quadruplet Convolutional Neural Networks (Quad-CNN) for multi-object following, which figure out how to partner object discoveries crosswise over edges utilizing quadruplet misfortunes. The proposed systems consider target appearances together with their transient adjacencies for information affiliation. In contrast to traditional positioning misfortunes, the quadruplet misfortune implements an extra imperative that makes transiently contiguous identifications more firmly situated than the ones with enormous transient holes. We likewise utilize a perform various tasks misfortune to mutually learn object affiliation and bouncing box relapse for better limitation. The entire system is prepared start to finish. For following, the objective affiliation is performed by minimax mark engendering utilizing the measurement gained from the proposed system. We assess execution of our multi-object following calculation on open Witticism Challenge datasets, and accomplish extraordinary outcomes.

### 4. “IBM Smart Surveillance System (S3): Event Based Video Surveillance System with an Open and Extensible Framework”

Ying-li Tian et.al [4] introducing the increasing need for sophisticated surveillance systems and the move to a digital infrastructure has transformed surveillance into a large scale data analysis and management challenge. Smart surveillance systems use automatic image understanding techniques to extract information from the surveillance data. While the majority of the research and commercial systems have focused on the information extraction aspect of the challenge, very few systems have explored the use of extracted information in the search, retrieval, data management and investigation context. The IBM smart surveillance system (S3) is one of the few advanced surveillance systems which provides not only the capability to automatically monitor a scene but also the capability to manage the surveillance data, perform event based retrieval, receive real time event alerts thru standard web infrastructure and extract long term statistical patterns of activity. The IBM S3 is easily customized to fit the requirements of different applications by using an open-standards based architecture for surveillance.

### 5. “Joint Detection and Identification Feature Learning for Person Search”

Tong Xiao et.al [5] states Existing individual re-recognizable proof benchmarks and techniques mostly center around coordinating trimmed passerby pictures among inquiries and competitors. Be that as it may, it is unique from genuine situations where the explanations of passerby bouncing boxes are inaccessible and the objective individual should be looked from a display of entire scene pictures. To close the hole, we propose another profound learning structure for individual pursuit. Rather than separating it into two separate undertakings—walker location and individual re-recognizable proof, we together handle the two perspectives in a solitary convolutional neural system. An Online Instance Matching (OIM) misfortune capacity is proposed to prepare the system viably, which is adaptable to datasets with various characters. To approve our methodology, we gather and comment on an enormous scale benchmark dataset for individual hunt. It contains 18, 184 pictures, 8, 432 personalities, and 96, 143 walker bouncing boxes. Investigations demonstrate that our system beats other separate approaches, and the proposed OIM misfortune capacity merges a lot quicker and better than the ordinary Softmax.

### 6. “Similarity Estimation for Large-Scale Human Action Video Data on Spark”

Weihua Xu et.al. [6] Introducing the measure of human activity video information is expanding quickly because of the development of interactive media information, which builds the issue of how to process the enormous number of human activity recordings efficiently. Along these lines, we devise a novel methodology for human activity closeness estimation in the distributed environment. The efficiency of human action similarity estimation depends on feature descriptors. Existing element descriptors, for example, Local Binary Pattern and Local Ternary Pattern can just separate surface data however can't acquire the item shape data. To determine this, we present another element descriptor, in particular Edge based Local Pattern descriptor (ELP). ELP can concentrate article shape data other than surface data and ELP can likewise manage power fluctuations. In addition, we investigate Apache Spark to perform include extraction in the disseminated condition.

At long last, we present an observational adaptability assessment of the undertaking of extricating highlights from video datasets.

### III. EXISTING SYSTEM APPROACH

In existing on machine learning and surveillance techniques we come on conclusion that there is not any promising solution for pre- event identification and alert generation. All alert systems are based on sensors and hardware devices which is very expensive. Existing work on video surveillance is used recorded video sequences for crime investigation which is post investigation process. In which chances of overcome risk and loss is very less. So there is need of pre-event identification systems in video surveillance and monitoring with the addition of alert generation for better accident prevention techniques. So we are working on smart surveillance system for giving most promising solution over existing post prevention methods.

### IV. ALGORITHM AND MATHEMATICAL MODEL

#### 1] Convolutional Neural Network (CNN):-

In proposed work we are using CNN which takes video frames as a input. After getting frames from video it will processed using image processing techniques for feature evaluation. We extract different features from those images regardless of their events in it consists. By using a series of mathematical functions we are going to identify the abnormal events. Every layer in CNN has capability to find out weights of images by using matrix evaluations which converts input to output with valuable functions. Layers of CNN used to identify fire events from extracted frames and give prediction by preserving high accuracy and less time.

- Step 1- Input event and normal event video's
- Step 2- Frame extraction from video
- Step 3- Image processing by using open-cv
- Step 4- Feature Extraction from images
- Step 5- Model generation
- Step 6- event recognition
- Step 7- Alert generation in the form of voice

Four main layer working approach of CNN explained below:-

#### a) Convolutional Layer

We are going to extract different features of frames like pixel weight matrix calculations by using feature kernels. Perform mathematical convolutions on frames, where every function uses a unique filter. This outcome will be in different feature maps. At the end, we will collect all of these feature maps and draft them as the destination output matrix of the convolution layer.

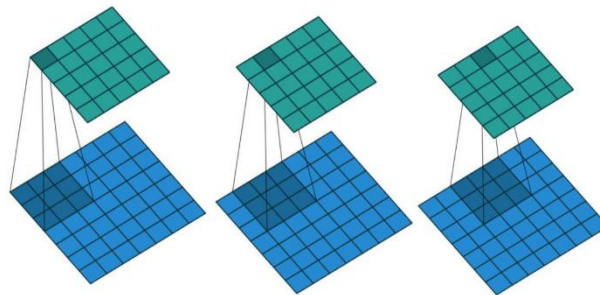


Fig.1 Convolutional Layer

#### b) Pooling

The expression of pooling is to constantly decrease the dimensionality to limits the number of factors and calculation in the network. This limits the time of training and maintains over fitting problem. The max Pooling extracts out the largest pixel value out of a feature. While pooling average is calculated for the average pixel value that has to be evaluated.

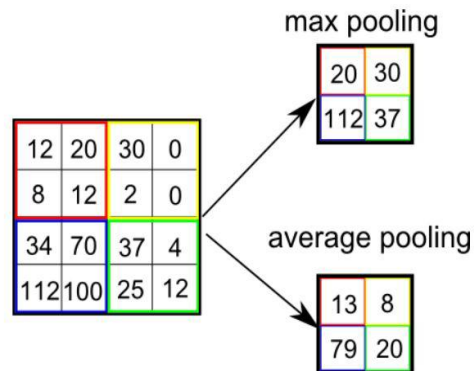


Fig.2 Pooling Layer

c) **Flattening**

Generally here we put the pooled feature into a single column as a sample input for further layer (transform the 3D matrix data to 1D matrix data)



Fig. 3 Flattening Matrix

d) **Fully Connection**

A fully connected layer has full connections of neurons to all the nodes in the previous layer. The fusion of more neurons to evaluates accurately.

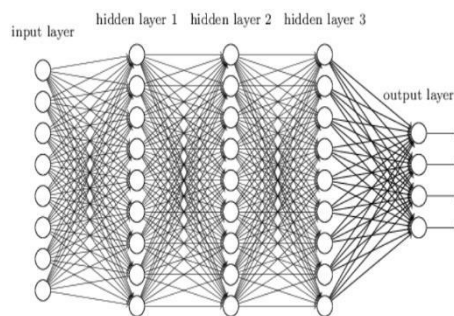


Fig.4 Fully Connected Layer

2] **Mathematical Model:-**

a] **Testing Model 1:**

**Sign System:**

Mathematical Model: Let us consider S as a system for Smart video surveillance system.

$$S = \{s, H, I, O, V, e\}$$

**INPUT:**

1. S=start of program
2. H= {h1, h2, h3 ..... , hn} where H is the set of video frames/images of testing dataset.
3. I= i1, i2, i3 testing image submitted by the User, i.e. fire sequence , parking vehicle sequences
4. O= o1, o2, o3 Set of voice outputs from the function sets, Output of desired event detection, i.e. fire detection
5. V=alert voice as a output
6. e = End of the program.

**b) Training Model 2:**

Set Theory

$$T=\{s, e, D, M\}$$

Where,

s = Start of the program.

T=Train set

D=video Dataset contains {L1,L2,L3,....Ln} where L is label sets of total dataset images

L1=set of label one fire data

L2= set of label two vehicle parking

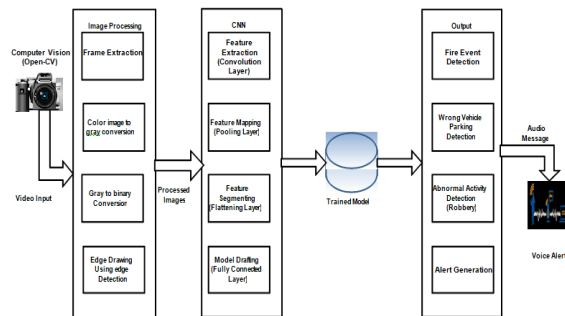
L3=set of label three robbery video's

M=Trained model

**V. PROPOSED SYSTEM APPROACH**

In a proposed system, we are proposing experiment on

Fire detection and pre-alarm generation, Abnormal activity detection and prevention, Smart Parking by using video surveillance.



**Fig.5 Block Diagram of Proposed System**

In a proposed system, we are going to overcome existing drawbacks of post investigation techniques of video surveillance systems by providing pre alert generation system. Our work is based on machine learning techniques for video analysis with better performance and event detection with advantages of alert generation. We are going to develop following modules:

- 1) Fire detection and pre-alarm generation
- 2) Abnormal activity detection and prevention
- 3) Smart Parking by using video surveillance

1) Fire recognition result:



Fig. 6 Fire recognition from video

2) Robbery detection result:

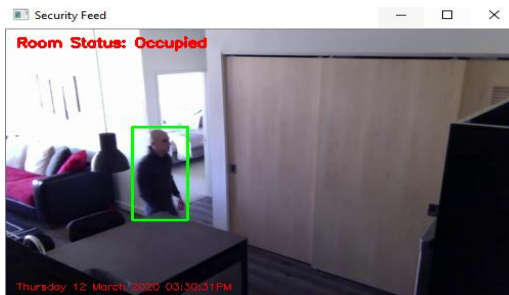


Fig.7 Person recognition from video

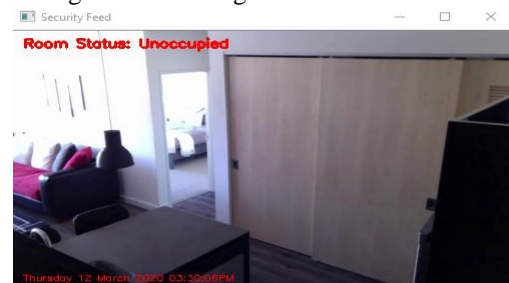


Fig.8 Person recognition from video

3) Smart parking result:



Fig.9 Motor bike recognition from video

Main motive behind this system is to give promising solution by pre-event recognition which alternatively increase prevention rate of accidents. To intensive task of monitoring surveillance regions as well as explore richly valuable information from the big surveillance data, researchers seek the advanced computer vision algorithms to develop intelligent video surveillance (IVS). In proposed system, motion detection, object recognition, tracking, and some are higher level analysis modules for specific applications, e.g., people counting, activity recognition etc.

## VI. COMPARATIVE RESULTS

In our experimental setup, as shown in table , the total numbers of fire and no fire video frames were tested. These frames go through event detection framework by following feature extraction using our image processing module. Then our trained model of fire detection get classifies the video into fire and no fire categories. Same procedure followed by robbery detection which classifies the video into robbery and no robbery categories. Smart parking video sequence get classified into car, bike and other vehicle categories. The common action will take on event detection by us after getting appropriate voice alert.

Sr. No	Category	Number of Frames
1	Positive Frames	750
2	Negative Frames	250

Table1 : Classification of frames

## VII. RESULTS

From above data, as shown in graph, the numbers of frames goes through test module some of found fire detected, some of found no fire.

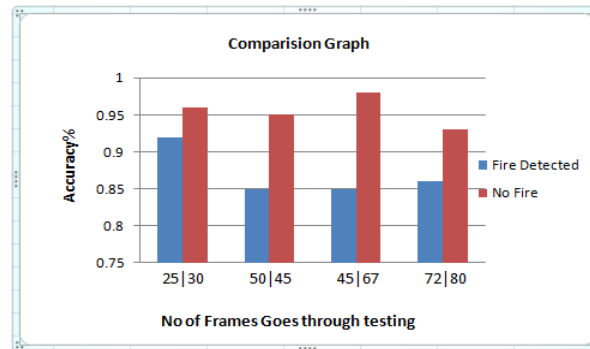


Fig.10 : Classification of fire videos

In our experimental setup, as shown in graph , the total numbers of frames were 414. These frames were then divided into Two subcategories; among which 25,50,45,72 found Fire detected and 30,45,67,80 found No Fire respectively We classified video data into fire and no fire categories based on accuracy factor which is our main motive.

## VIII. CONCLUSION

We have proposed smart surveillance systems based on video surveillance and monitoring techniques by saving cost of hardware. And give best result over post-event recognition by our pre-event recognition and alert generation work. Future work will be based on real time CCTV based video surveillance systems mounted on real time areas.

## ACKNOWLEDGMENT

This work is supported in a smart surveillance system of any state in india. Authors are thankful to Faculty of Engineering and Technology (FET), SavitribaiPhule Pune University,Pune for providing the facility to carry out the research work.

## REFERENCES

- [1] Patel Parin, Gayatri Pandi “ Traffic Monitoring using Video Stream with Machine Learning: Based on Big Data Process with Cloud”. International Journal of Innovations & Advancement in Computer Science 2017
- [2] Shaoqing Ren,Kaiming He Ross Girshick Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 3, March 2015



- [3] Jeany Son, Mooyeol Baek, Minsu Cho, Bohyung Han “Multi-Object Tracking with Quadruplet Convolutional Neural Networks” Dept. of Computer Science and Engineering, POSTECH, Korea 2017
- [4] Ying-li Tian, Lisa Brown, Arun Hampapur, Max Lu, Andrew Senior, and Chiao-fe Shu “IBM Smart Surveillance System (S3): Event Based Video Surveillance System with an Open and Extensible Framework” Machine Vision and Applications Journal 2017
- [5] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, Xiaogang Wang “Joint Detection and Identification Feature Learning for Person Search” The Chinese University of Hong Kong 2017
- [6] Weihua Xu, Md Azher Uddin ID, Batjargal Dolgorsuren, Mostafijur Rahman Akhond, Kifayat Ullah Khan, Md Ibrahim Hossain and Young-Koo Lee “Similarity Estimation for Large-Scale Human Action Video Data on Spark” Data and Knowledge Engineering Lab 2018