Intelligent Video Surveillance System Based On Machine Learning
S. L. Bangare\textsuperscript{1*}, Reuben F. Tribhuwan\textsuperscript{2}, Krutika M. Shukla\textsuperscript{3}, Sheetal A. Jain\textsuperscript{4}, Rushikesh G. Vispute

\textsuperscript{1, 2, 3, 4, 5} Department of Information Technology, Sinhgad Academy of Engineering, Pune, Savitribai Phule Pune University, India
*Corresponding author’s e-mail: sunil.bangare@gmail.com

Abstract- Nowadays 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 robbery detection. 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 an important part in the security and protection of modern cities. Since Video surveillance system has become a critical part in the security and protection of modern cities, since smart monitoring cameras equipped with intelligent video analytics techniques can monitor and pre-alert system by capturing abnormal activity and events. In recent years, more and more video surveillance devices are deployed as there are increasing demands on public security and smart city.

Keywords— Intelligent and Cost effective Video Surveillance, Fire Detection, Deep neural network.

1. INTRODUCTION
The intensive task of monitoring surveillance regions as well as exploring 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 parse 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 has 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 systems which can do early event detection and pre alert generation systems.

2. LITERATURE SURVEY
1. “Cost-Effective Vehicle Type Recognition in Surveillance Images With Deep Active Learning and Web Data”

Yue Huang et.al [1] proposed a method for vehicle type recognition. The main aim of this paper was to reduce the manual labeling of surveillance images as annotating large scale images from surveillance in ages is time consuming. According to this proposed method, firstly, a memory space is constructed using a large scale fully labeled auxiliary dataset collected from the internet. Then, the similarity measurement in memory space and the entropy help in emphasizing the diversity and uncertainty in the query strategy. The proposed method was evaluated on the Comprehensive Cars dataset and it was concluded that the proposed method could reduce the annotation cost by 40%.
2. “Multi-Object Tracking with Quadruplet Convolutional Neural Networks”

Jeany Son et.al [2] proposed Quadruplet Convolutional Neural Networks (Quad-CNN) for multi-object tracking which figures out how to partner object discoveries crosswise over edges utilizing quadruple misfortunes. The proposed system considers 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. This method was experimented on open Witticism Challenge datasets.


Patel Parin and Gayatri Pandi [3] state Traffic Monitoring is a difficult undertaking on jam-packed streets. Traffic Monitoring methods are usually costly, manual, time expending and also involve human administrators. Huge scale stockpiling and investigation of video streams were unrealistic due to constrained accessibility. Now with the advancement in technology it is possible to carry out object discovery and tracking, conduct examination of traffic designs, number plate acknowledgment and reconnaissance on video streams created by traffic checking. 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.

4. “Detection of fire using image processing techniques with LUV color space”

Divya Pritam [4] introducing the detection of fire using various image processing techniques with LUV color space. The basic technique in identification of fire in an image is color detection. However, the color of fire varies from red, yellow and white. Also, there are non-fire objects with fire like color. Hence in order to improve the accuracy of the system we combine the color detection technique with other fire detection techniques. Edge detection, motion detection, area covered by flames, existence of smoke, growth of fire and background segmentation are some techniques which are combined by various researchers and used to correctly classify the fire images and fire-like non fire images in a video. This paper presents the comparative analysis of five recent vision based fire detection system.

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. 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.

6. “Smart Monitoring Cameras Driven Intelligent Processing to Big Surveillance Video Data”

Zhenfeng Shao et.al [6] states that in order to make better use of surveillance video data, presenting a methodology to maximize the role of smart cameras in surveillance system is a meaningful approach. In the system mentioned in this paper, first of all, a database of abnormal behavior is established for smart cameras. This is done in order to store and manage the warnings in emergency. Secondly, surveillance videos are stored selectively according to the warning information, which in turn, substantially reduces storage space. At last, when video evidences are needed in some cases, those associated with abnormal behaviors are traced and accessed preferentially. For Medical
Image related work we have referred S. L. Bangare et al research work [7] [8] [9] [10].

3. EXISTING SYSTEM APPROACH

In existing systems using machine learning and surveillance techniques, we have come on a conclusion that there is no promising solution for pre-event identification and alert generation. All alert systems are based on sensors and hardware devices which are very expensive. Existing work on video surveillance uses recorded video sequences for crime investigation which is a post investigation process. In which chances to 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. Hence, we are working on smart surveillance system for giving most promising solution over existing post prevention methods.

4. ALGORITHM AND MATHEMATICAL MODEL

1] Convolutional Neural Network (CNN):

In proposed work we are using CNN which takes video frames as an input. After getting frames from video it will be processed using image processing techniques for feature evaluation. We extract different features from those images regardless of the events they consist. 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 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.

- Step 1- Input event and normal event videos
- 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.

b) Pooling

The expression of pooling is to constantly decrease the dimensionality to limit the number of factors and calculations in the network. This limits the time of training and maintains an 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.
a. **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)

![Flattening Matrix](image)

**Fig. 3 Flattening Matrix**

b. **Fully Connection**

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

![Fully Connected Layer](image)

**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. $s = \text{start of program}$
2. $H = \{h_1, h_2, h_3 \ldots , h_n\}$ where $H$ is the set of video frames/images of testing dataset.
3. $I = i_1, i_2, i_3$ testing images submitted by the User, i.e. fire sequences, parking vehicle sequences
4. $O = o_1, o_2, o_3$ Set of voice outputs from the function sets, Output of desired event detection, i.e. fire detection
5. $V = \text{alert voice as an output}$
6. $e = \text{End of the program}$

b] **Training Model 2:**

**Set Theory**

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

Where,

$s = \text{Start of the program}$.

$T = \text{Train set}$

$D = \text{video Dataset contains } \{L_1, L_2, L_3, \ldots , L_n\}$ where $L$ is label sets of total dataset images

$L_1 = \text{set of label one fire data}$

$L_2 = \text{set of label two vehicle parking}$

$L_3 = \text{set of label three robbery videos}$

$M = \text{Trained model}$

5. **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.
In our 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(a) Fire recognition (No Fire)
   Fig. 6(b) Fire recognition (fire)

2) Abnormal activity result:

   Fig. 7(a) Robbery detection
   Fig. 7(b) No Robbery detection

3) Smart parking result:

   Fig. 8(a) No Person recognition
   Fig. 8(b) Person recognition
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.

6. COMPARITIVE 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 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.

Table I : Classification of Frames

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Category</th>
<th>Number of Frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Positive Frames</td>
<td>750</td>
</tr>
<tr>
<td>2</td>
<td>Negative Frames</td>
<td>250</td>
</tr>
</tbody>
</table>

7. 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.

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 192 found Fire detected and 222 found No Fire respectively. We classified video data into fire and no fire categories based on accuracy factor which is our main motive.

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.

REFERENCES

[2] 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.


[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


