International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

SMART SURVEILLANCE SYSTEM BASED ON MACHINE LEARNING

Anamika Kulkarni[1], Varsha Molawade[2], Nakshatra Mungase[3], Deepak Khachane[4]

[1,2,3]Student, Computer Engineering, New Horizon Institute of Technology & Management (Thane), Maharashtra, India
[4]Professor, Computer Engineering, New Horizon Institute of Technology & Management (Thane), Maharashtra, India

---

Abstract-*We propose a quick and robust approach to the detection and tracking of moving objects. Our method is based on using area computed by image comparison showing the motion. While it is known among researchers that image comparison are compatible for accurate computation of movement, not much attention is paid to making systems for detecting and tracking objects using this feature. In our method, extracted area by using image comparison is compared with the same area of previous frame. The area is above the thresholding criteria then it is marked as area where motion has been taking place. Input images are given by using webcam. Detected objects are tracked. The experimental results on indoor-scenes show fast and robust performance of this method.*

Key Words: object tracking, motion detection, Face detection, Haar classifier, CNN, background subtraction, face recognition.

1. INTRODUCTION

Detecting and tracking moving objects are widely used as low-level tasks of computer vision applications, like video surveillance, robotics, authentication systems, user interfaces by gestures. A recent report by Kyle Goehner [1] illustrates the concept that examines the utilization of three different back- ground subtraction algorithms — Mixture of Gaussians (MOG),Visual Background Extractor (ViBe), and Pixel-Based Adaptive Segmentation (PBAS) they have used this algorithms for observing wildlife . Out of which we are using back-ground subtraction algorithm for security purpose in the Organizations. The drawback of this existing system it requires a person to monitor the system.. Back-ground Subtraction detects the moving object from static camera and also compare the pixel between the current frame andthe previous frame. So in our project we are using this algorithm for observing the unauthorized movement around the distinct area.The other algorithms that we are using in this project are Convolutional Neural Network (CNN) for face recognition Haar Cascading Classifier which takes the eigen values for face calculation and it uses Color recognition for fire detection. The input in our project is that if unauthorize person for activity takes place it will alert the admin by buzzing the alarm and it will buzz critical alarm for higher risks. The advantage in our project is that it doesn't require the person to monitor.

2. REVIEW OF LITERATURE

Several surveillance systems are developed to monitor a large area , due to which one can find difficulty technical solutions. Compared with traditional cctv surveillance security methods. Smart surveillance based security method has shown far better results.

2.1 Limitation of Existing System

In the existing system of CCTV Surveillance system, the observer / security guard would have to continuously monitor the screen for any unwanted activity. Cameras enable users to record footage for viewing later on , and to help nab criminals, and receive justice from the law. They cannot, however, stop a criminal offense when it's ongoing. They do not alert neighbors or the police like an alarm would. Previous system is very costly. Requires man power for monitoring. Will not give any notification when crime is happened

3. SYSTEM DESIGN

3.1 Overall System Overview

The whole method of developing the machine learning based smart surveillance system is explained in detail below. The developing procedure is divided into several important stages, including obtaining the training dataset and augmentation, training the model and finally buzz alarm and send the notifications (messages, e-mails) to the admin.

3.2 Dataset Preparation

The first step is to prepare the dataset of authorized people in the organisation. In this, the individuals were photographed using a webcam. Approximately 10-15 images were taken of each person with changes in movements and expressions. The images of each person was then saved in a different folder labelled with the student's name as shown in the figure below.

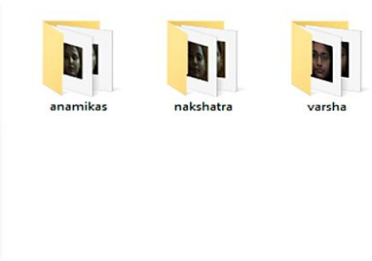

Fig -1: Dataset preparation

3.3 Motion Detection

We continuously capture image frames using webcam. It detects every motion taking place in the frame and highlights it by bounding box as shown in the figure. We use Background Subtraction algorithm for the same.

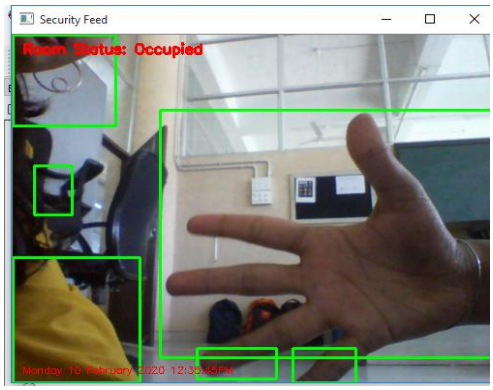

Fig-2: Result of motion detection

3.4 Face Detection

If any motion is detected it will go for the face detection. This is done using Haar Cascade classifier which uses Eigen values. These eigen values are points on our face like Eyes, Nose, Mouth. If the face is detected it shows "**user**" on top of the bounding box.

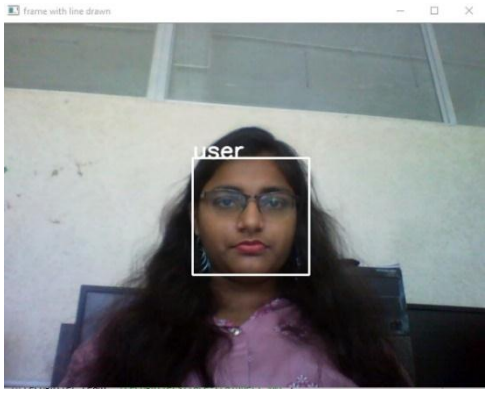

Fig -3: Result of Face detection

3.5 Face Recognition

After a face is detected, the face recognition module will be implemented. Here the system will compare the detected face with the dataset, and if it is an authorized user then it will display the name of the person on the box instead of user.

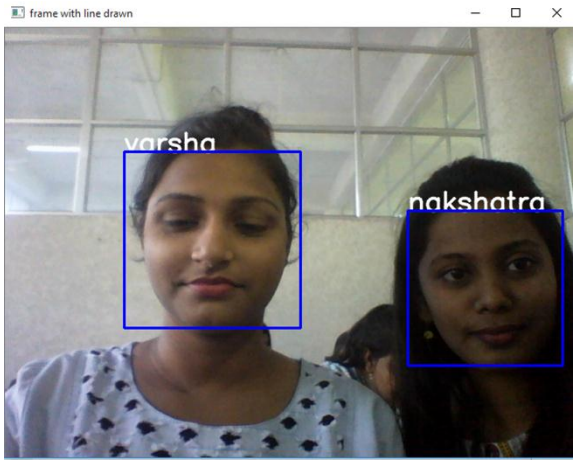

Fig-4 : Name Displayed for recognized faces

3.6 Proposed System/Architecture

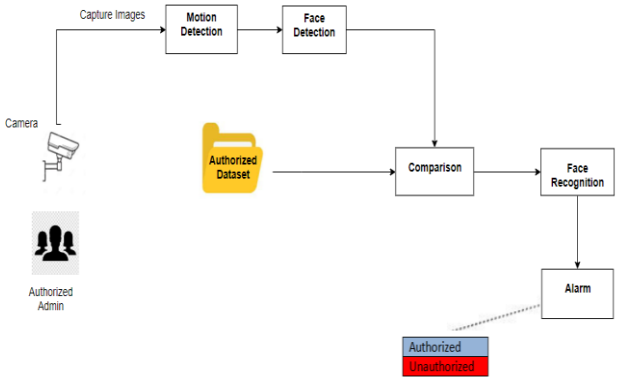

Fig- 5: Block Diagram

4. IMPLEMENTATION

4.1 Haar Cascade Classifier

A Haar Cascade is basically a classifier which is used to detect the object for which it has been trained for, from the source. The Haar Cascade is trained by superimposing the positive image over a group of negative images. Calculating a feature through Cascaded Classifiers Cascaded is extremely efficient and fast.

Calculating all 180,000 features contained within a 24 × 24 sub-image is impractical. Fortunately, only a small fraction of these features are needed to determine if a sub-image potentially contains the specified object. In order to eliminate as many sub-images as possible, only a couple of the features that outline an object are used when analyzing sub-images. The goal is to eliminate a considerable amount, around 50%, of the sub-images that don't contain the object. This process continues, increasing the number of features used to analyze the sub-image at each stage.

In our project, using a 2 GHz computer, a Haar classifier cascade could detect human faces at a rate of at least five frames per second.[5]

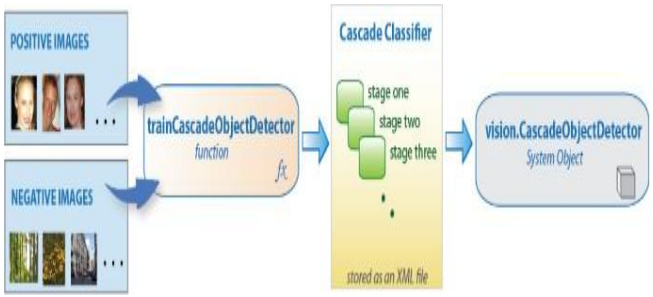

Fig -6: Haar Classifier

4.2 Convolution Neural Network (CNN)

A convolutional neural network (CNN) is a specific sort of artificial neural network that uses perceptrons, a machine learning unit algorithm, for supervised learning, to analyze data. CNNs apply to image processing, natural language processing and other kinds of cognitive tasks.[6]

In neural networks, Convolutional neural network (ConvNets or CNNs) is one of the main categories to do image recognition, image classifications. Objects detection, recognizing faces etc., are some of the areas where CNNs are widely used.

Technically, deep learning CNN models to train and test. Each input image will pass it through a series of convolution layers with filters (Kernals), Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1. The below figure is a complete flow of CNN to process an input image and classifies the objects based on values.

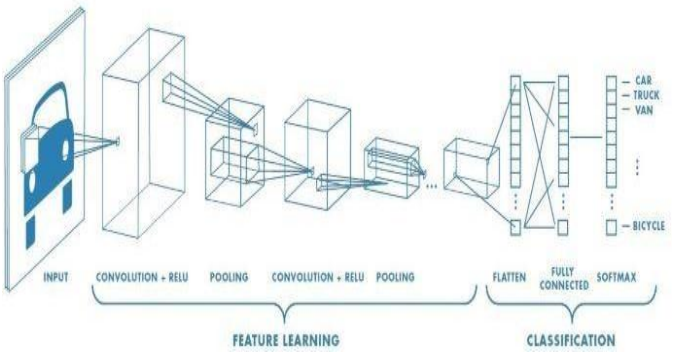

Fig -7: Neural network with many convolution layers

1. Convolution Layer:

Convolution is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data.

2. Pooling Layer

Pooling layers section would reduce the number of parameters when the images are too large. Spatial pooling also called subsampling or down sampling which reduces the dimensionality of each map but retains the important information. Spatial pooling can be of different types:

- Max Pooling
- Average Pooling
- Sum Pooling

3. Fully Connected Layer

The fully connected layers, we combined these features together to create a model. Finally, we have an activation function such as softmax or sigmoid to classify the outputs as cat, dog, car, truck etc.,

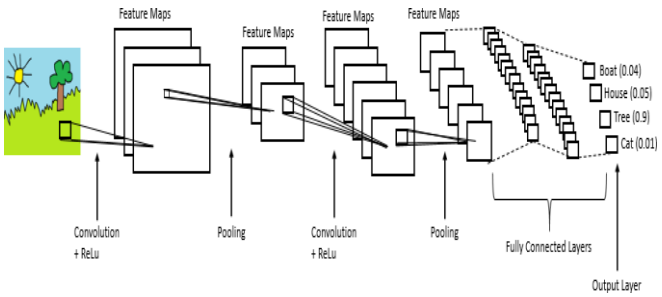

Fig-8: CNN Architecture

4.3 Background Subtraction Algorithm

Background subtraction is essentially detecting moving objects in videos using static cameras. The basic idea in the approach is detecting the moving objects from the difference between the current frame and a reference frame, which is named "background image" or "background model". The background image must be good enough to represent the scene with no moving objects and be regularly updated in order that it is adapted to the varying luminance conditions and geometry settings.

Frame difference is an uncomplicated form of background subtraction. The current frame is simply subtracted from the previous frame, and if the difference in pixel values for a given pixel is higher than a threshold h T then the pixel is considered part of the foreground.

i i h frame - frame > T -1

The estimated background is simply the previous frame and it is very sensitive to the threshold *h T.*

The result shows "the frame difference" method as very low computationally intensive and efficient method. It also subtracts out background noise (such as waving trees), far

better than the more complex approximate median and **Mixture of Gaussians** methods (high computation methods). But the important challenge in this method is the determination of appropriate threshold, since the result solely depends on the threshold used.

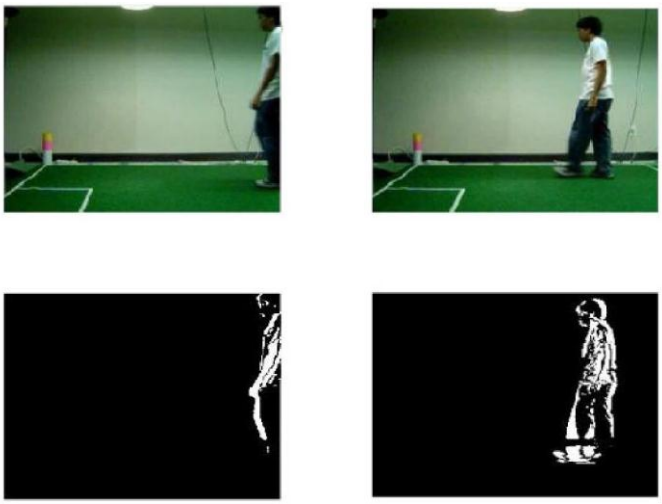

Fig -9: Result of background Subtraction

5. RESULT

As discussed above, the system is suitable for real-time usage, as the performance is sufficiently high and required computational resources are low. Compared with traditional cctv surveillance security methods. Smart surveillance based security method has shown far better results. The algorithm improves the detection and location of moving objects in the video images. Also face recognition feature makes it even more useful in places where there are many people working like organizations. The system is able to effectively detect all the motions occurring in the video frame.

It is also able to detect more than one faces effectively as shown in the figure below.

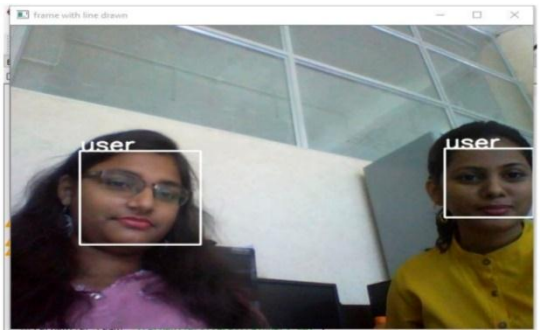

Fig -10: Detection of more than one face

Incase there is one authorized person and one unauthorized person, the system is able to detect both authorized and unauthorized face. The known face is bounded by a blue box with his name on top of the box and the unknown face is bounded by a green box.

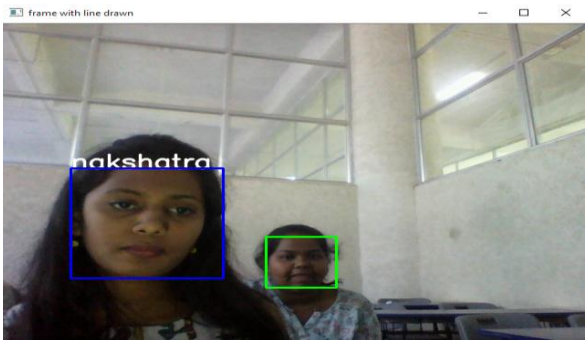

Fig-11: Both authorized and unauthorized face detected

Whenever motion is detected an alert message is sent to the admin that an unknown face is detected in the room.

Further an image of the face is also sent to the admin's registered email.

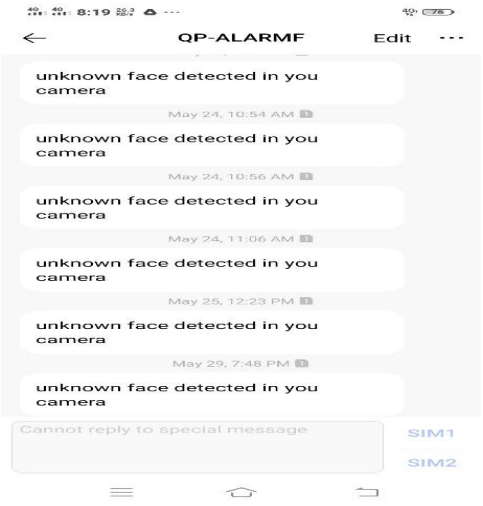

Fig-12: Alert message sent to admin

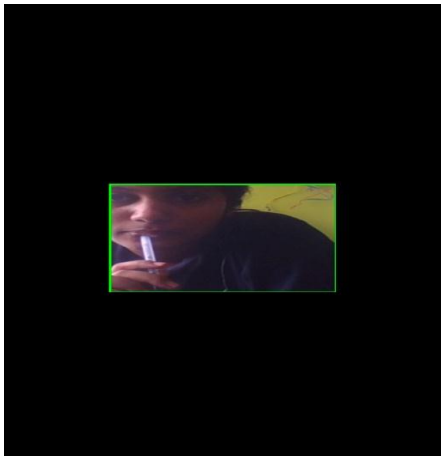

Fig-13: Image sent to the admin's email

6. CONCLUSION

There is a huge interest in the market to make technical equipment smart and self-learning. An important component in such systems is the ability for a computer to track and identify moving objects. The problem of tracking the movement of a desired object that is captured by a real time video stream is of interest because of the many applications that can be derived from it. The object is marked using a green square Superimposed on the object which makes the position of the object clear to the observer. The algorithm improves the detection and location of moving objects in the video images. Also face recognition feature makes it even more useful in places where there are many people working like organizations. Thus it is a very useful application for security purposes.

REFERENCES

1. Kyle Goehner, Travis Desell, "A comparison of Background Subtraction Algorithms for Detecting Avian Nesting Events in Uncontrolled outdoor Video", in 11th International Conference on eScience , 2015.
2. A.F. Bobick, and J.W. Davis, "The recognition of human movement using temporal templates," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 257-267, 2001.
3. M.Turk, A.Pentland, "Eigenfaces for recognition", *Journal of cognitive neuroscience*,vol.3,no.1,pp.71-86,1991.
4. Facial feature detection: https://pdfs.semanticscholar.org/9263/c9d03b990e3f263b538f85dcc0b1d7aa7528.pdf
5. Understanding of Convolutional Neural Network (CNN) — Deep Learning : https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning