Real time Object Detection: A Path Towards Video Surveillance As a Service

Dr. Shridevi Soma[1], Sunita[2]
[1, 2]Dept Of CSE
[1, 2, 3] PDACE, Kalaburgi Karnataka, India

Abstract- *Video surveillance is the largest segment in the security system. Object detection from the real time video under surveillance is challenging in terms of space and time to detect the object. The cloud storage service helps to fulfil the storage requirement of video under surveillance. This research paper aims to develop a solution for Real Time Object Detection: A Path Towards Video Surveillance as a Service is proposed to provide optimal solution for detecting objects in real time. The design approach of RTOD-VSaaS used for intruder detection that reinforces surveillance technology to provide essential security to the life and associated control and alert operations. This work contains various modules such as Admin, Video Capture, Preprocessing, Email Generating Module and Report Generating Module. Camera sensor captures video under surveillance with a speed of 20 frames per second. The Input video is streamed and stored in Cloud server using twitch and drop box respectively. Preprocessing techniques such as Frame generation and Sobel Edge Detection is considered in the experimentation. After the detection of the object using Kalman Filter an alert message is sent to the authorized person through mail. The cloud video streaming reveals, cloud-based video surveillance as a service.*

Keywords- Video surveillance, Object detection, Cloud, Internet of Things

I. INTRODUCTION

The demands on video surveillance systems are rapidly increasing day by day. One of the primary things individuals can need to understand regarding their security system is whether or not they have ability to connect it over the internet for remote viewing. Remote access to live and recorded video using smart devices is an important feature in VSaaS. Video surveillance as a service have made it possible to view your remote security camera from internet-enabled PC or phone from anywhere in the world.

This technology related to security provides a secure and safe environment. The system captures the video from the camera, at the same time live feeds from the camera can be seen directly on server and simultaneously video is stored. Then apply the image processing techniques such as sobel edge detection and Kalman filtering for object detection. The term image processing indicates the processing on image or video frame which is captured as an input and the result set of processing is a set of related parameters of an image or video. The important part of object detection and motion analysis is movement of the object; here the aim of real time object detection system is to detect the motions of object from the background image. Object tracking is a process of segmenting an area of interest from a video frames and keeping track of object motion and its position.The various objectives of the proposed system are to detect an intruder(motion), capture an image of the intruder and convey an alert message to the owner.

II. RELATED WORK

T Forsyth [8] proposed the most popular way to implement motion detection algorithm using image segmentation and image subtraction techniques of computer vision.Most of the algorithm first segments the foreground moving objects from the background image. To do this, they would have to take a sequence of images with no motion by default to initialize the background image. This background image would be updated subsequently to provide a real-time environment where changes to the background are taken into consideration. Foreground objects can be acquired by using simple arithmetic using image subtraction. The result of the subtraction techniques where pixels belonging to the current image are subtracted by the corresponding pixels in the background image or vice versa would obtain the foreground moving objects. After obtaining the foreground objects, the focus for region of interest is set to these foreground objects instead of the whole image. Therefore, further image processing is performed only on these regions of interest. One of the important steps here is to extract important features of the moving objects to recognize the object. As mentioned, the first step of a motion detection algorithm would be the foreground-background segmentation step. **Ming and Tim [5]** presented a way to improve the usual image subtraction techniques which only uses the pixels intensity values by adding also the Colour illumination invariance into the algorithm. As a result, their work can also be implemented effectively on motion detection areas where the illumination changes rapidly. Implementing edge detectors helps in obtaining a clearer picture of the moving objects. **Ramesh, Rangachar and Brian [7] & Milan,Vaclav, Roger [16]** stated in their text that using edge detectors combining with the different pictures will give a better result for motion detection by helping to overcome several limitations. There are many edge detectors algorithms being introduced in the image processing field. Mostly all text on image processing does cover the topics on edge detection.

III. METHODOLOGY

Performance of real time object detection in video surveillance depends on its ability to detect moving objects in the environment. An action, such as analyzing the objects motion, requires an extraction of the foreground objects, making moving object detection a crucial part of the system. RTOD-VSaaS system captures video from camera, simultaneously it uploads the file to the cloud and live feeds from camera can be seen directly server. Storage, streaming, and security are main factors of VSaaS. The cloud video streaming reveals, cloud-based video surveillance as a service. The captured video then it converts to the number of frames. On each frame applying pre-processing which removes the noise and makes the image sharpen. Once the object is detected then it sends alert to the owner by sending video and audio clips to email and gives alarm alert to the authorized person.

Object Detection

Steps Involved in object detection are as follows:

Step 1: Capture the input video frames from camera

Step 2: Captured video frames are converted to greyscale frames

Step 3: Gaussian blur is applied for noise removal from greyscale frames

Step 4: Considering the first grayscale frame as the reference frame for generating a difference frame and a threshold frame

Step 5: Dilating the Threshold frame and finding pixel contours in that frame

Step 6: Filtering the current frame and finding the contour area and bounding the end-points in a rectangle

Step 7: Capturing the time-stamp when objects enters and exits the frame
Step 8: Generating the Time data

Sobel Edge Detection (Sobel operator)

The Sobel operator is used for edge detection. It works by calculating the gradient of image intensity at each pixel within image. It finds the direction of the largest increase from light to dark and the rate of change in the direction. The result shows how the image changes at each pixel, and therefore how it is that pixel represents an edge. It also shows how that edge is to be oriented. The filter applied to a pixel in a region of constant intensity is a zero vector. The result of applying that filter to a pixel on an edge is a vector that points across the edge from darker to brighter values.

Kalman Filter Algorithm

Kalman filtering is used for filtering noisy signals, generating non-observable states, and predicting future states of the object. It is necessary to calculate the object position and speed in each instant. In proposed work input is considered as a sequence of images captured by a camera containing the object. Then using an image processing method, the object is segmented and later calculated their position in the image. Therefore, consider system state as x_k the position x and y of the object in the instant k. Kalman filter is adopted to make more efficient the localization method of the object to be detected using search window centered in the predicted value $\hat{x}_k^-$ of the filter.

Steps Involved in vision tracking are:

Step 1: Initialization (k = 0). Kalman filter searches for the object in the whole image because the previous position of the object is not defined.

Step 2: Prediction (k > 0). Kalman filter predict the relative position of the object, and that position $\hat{x}_k^-$ is considered as search center to finding the object.

Step 3: Correction (k > 0). Kalman filter locate the object (which is in the neighborhood point predicted in the previous stage $\hat{x}_k^-$) and use its real position (measurement) to carry out the state correction.

Following section provides the details of the main phases of proposed work

A. Image acquisition

Input Video image is captured from a camera with a resolution of 640x480.After video recording of a scène, the video frames were created. All video camera has the feature "frames per second‖ (FPS)".20 frames per second are generated for this surveillance system.

Color conversion

An image captured here is in form of color image. A grayscale image is used on entire process instead of the color image. The grayscale image only has one color channel that consists of 8 bits while RGB image has three channels.

B. Image Preprocessing

Obtained Grayscale images can contain a lot of noise or random variation in brightness or hue among pixels. The pixels in these images have a high standard deviation, which just means there's a lot of variation groups of pixels. Because the image captured by the camera is in two-dimensional, Gaussian blur uses two mathematical functions x-axis and y-axis to create a third function, also known as convolution. The formula of a Gaussian function in two dimensions is the product of two such Gaussian functions, one in each dimension:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad \text{……….. (1)}$$

In equation (1), where x is the horizontal axis and y is the vertical axis, and σ is the standard deviation. The third function creates a normal distribution of those pixel values, smoothing out some of the randomness. Values from this distribution are used to build a convolution matrix which is applied to the original image. Each pixel's new value is set to a weighted average of that pixel's neighborhood. The original pixel's value having the highest Gaussian value and neighboring pixels receive smaller pixel values as their distance to the original pixel increases.

C. Frame differencing

Aim of proposed system is to find the difference between two video frames. This technique starts from the assumption that the background is static and compares the pixels of the frames captured in very small-time intervals (Δt). If the difference value is above a pre-defined threshold value (Γ), it means that the value of the pixel has changed and movement occurred. This method can be described by the following Equation (2):

$$I_i = \begin{cases} 1, & if\ |f_t - f_{t-1}| \geq \Gamma \\ 0, & if\ |f_t - f_{t-1}| < \Gamma \end{cases} \quad \text{……… (2)}$$

In equation (2) where is the binary result of the absolute difference of a pixel between two consecutive frames $(f_t\ and\ f_{t-1})$.The differences of all the frame pixels give us a collection of binary numbers that ultimately translate the movement. This method allows a good movement recognition, it is easy to implement and require less memory space.

D. Background subtraction

To identify the moving objects in the video background subtraction method is used, where each pixel from the video frame is compared against a reference image of the background. When pixels in the current frame differ significantly from the reference image, it is considered that movement occurred. A very plain approach to this method is to detect the foreground objects as the difference between the current frame and an image of the scene with static background, similar to the technique addressed before (Equation 3):

$$I_i = \begin{cases} 1, & if\ |f_t - f_{background}| \geq \Gamma \\ 0, & if\ |f_t - f_{background}| < \Gamma \end{cases} \quad \text{……….. (3)}$$

E. Thresholding

After background subtraction, using an optimal threshold a grayscale input image is converted into a binary image. Apply thresholding methods on binary image. Each pixel in an image is replaced with a black pixel if the image intensity $I_{i,j}$ is smaller than fixed constant T (that is, $I_{i,j}$<T) or if the image intensity is greater than that constant then it is considered as white pixel. A Threshold value is set in proposed system as 30.

F. Motion list

It includes the start and end time of object in motion. Its main function is to store the time values during object detection and its movement.

IV. EXPERIMENTAL RESULTS

Camera at the input end of the acts as a sensor of IoT system. After sensing/capturing input video it is sent as a video stream to the cloud storage as shown in Figure-1. In traditional computing data is stored on PC's local hard drive, in case of cloud computing the data is stored on physical and virtual servers that are hosted by a third-party service provider. Cloud dropbox is used to store the video stream in this work. Dropbox files can be accessed from any device via the Internet.

The Image is captured from the camera can be broadcasted using "Twitch" channel as shown in Figure-2. Once the object is detected in the surveillance system an alert notification is sent to the authorized person as shown in Figure-3 on laptop with voice message. The Figure-4 shows the video file sent through the program so that the owner can able to get the alert video form anywhere through this mail.

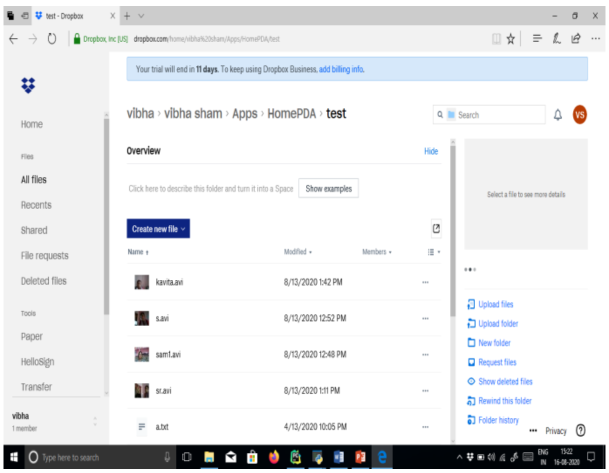

Figure-1: Video database stored in the dropbox

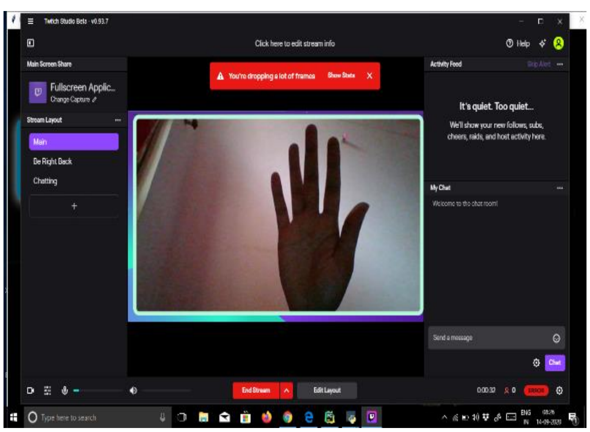

Figure-2: Streaming of input video

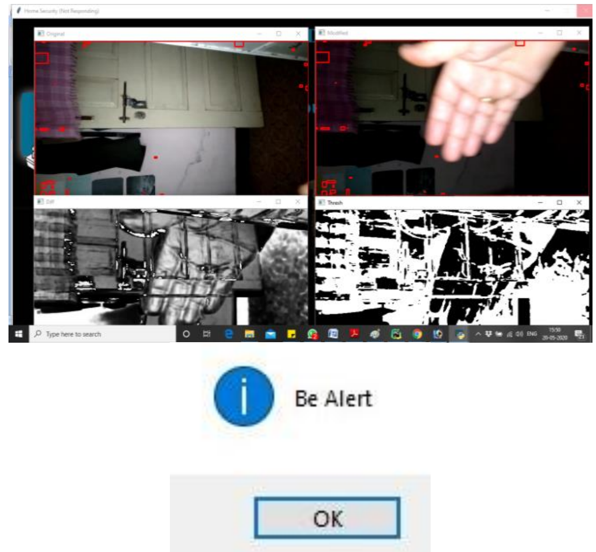

Figure3: Alert notification after object detection

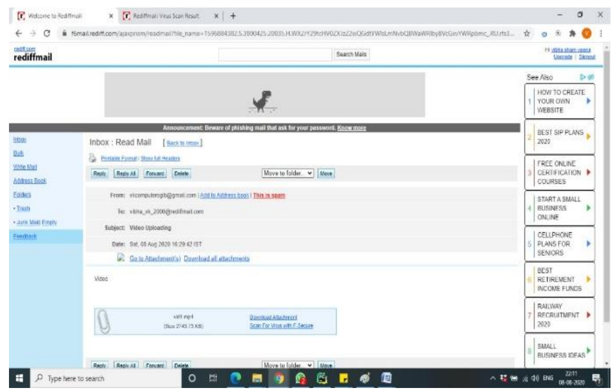

Figure-4: Video file mailed to the authorized person

V. CONCLUSION

It is an active surveillance system which will alert the user when the object is detected. Live video streaming is an additional advantage of the system. Web servers which help the user to view the live video provided by VSaaS. This system also sends the captured video to the owner by email. The user can get alerts anytime and anywhere through email on smartphone/desktop/laptop. Whenever any unknown or suspicious movement is detected, it gives an alarm and voice message. In this work live video is stored in cloud drop box and object detection is carried out by using image processing techniques. Twitch channel is used for live streaming of video under surveillance. Once the Object is detected authorized person get an alert message along with Object detected video by mail. This system is designed with an aim that it can be used for all kind of people since security of every one's home should not be left behind.

REFERENCES

[1] David Moore, "A real-world system for human motion detection and tracking", California Institute of Technology, June 2003.

[2] Yang Song, Luis Goncalves, Pietro Perona, "Learning Probabilistic Structure for Human Motion Detection", California Institute of Technology.

[3] Yang Song, Xiaolin Feng, Pietro Perona, "Towards Detection of Human Motion", California Institute of Technology.

[4] Randal C. Nelson, "Qualitative Detection of Motion by a Moving Observer", University of Rochester.

[5] Ming Xu, Tim Ellis, "Illumination-Invariant Motion Detection Using Colour Mixture Models", City University, London.

[6] Lina J. Karam, David Rice, "Image Convolution Concepts and Applications online tutorial", http://www.eas.asu.edu/~karam/2dconvolution/, Arizona State University.

[7] Jain, R., Kasturi R., Schunck G., "Machine Vision", McGraw Hill, 1995

[8] Forsyth,D.A., Ponce,J., "Computer Vision: A Modern Approach", Pearson Education, Upper Saddle River, NJ, 2003

[9] G. Johansson, "Visual perception of biological motion and a model for its analysis.", Perception and Psychophysics, 14:201-211, 1973.

[10] S. Birchfield, "Derivation of Kanade-Lucas-Tomasi tracking equation", http://robotics.stanford.edu/˜birch/klt/derivation.ps , 1997.

[11] Zoran Zivkovic, Ferdinand van der Heijden, "Improving the selection of feature points for tracking", Pattern Analysis and Recognition vol 7, no.2, 2004.

[12] Robert Pallbo, "Motion Detection: A Neural Model and its Implementation", Lund University Cognitive Science.

[13] Ying-Li Tian, Arun Hampapur, "Robust Salient Motion Detection with Complex Background for Real-time Video Surveillance", IBM T.J Watson Research Centre.

[14] Mike Smart, Barry Keepence, "Understanding Motion Detection", Indigo Vision VB8000 motion detection enabled transmitter documentation.

[15] Andrew Blake, Kentaro Toyama, "Probabilistic Tracking in a Metric Space", Microsoft Research Ltd. Cambridge U.K & Microsoft Research Redmond, WA, U.S.A.

[16] Milan Sonka, Vaclav Hlavac, Roger Boyle, "Image Processing, Analysis, and Machine Vision", PWS Publishing, 1999.

[17] Linda G. Shapiro, George C. Stockman, "Computer Vision", Prentice Hall, 2001.

[18] Thomas B. Moesland, Erik Granum, "A Survey of Computer Vision-Based Human Motion Capture", Computer Vision and Image Understanding, 81:231-268, 2001.

[19] Bob Davies, Rainer Lainhart, etc. , "Intel Open Source Computer Vision Libraries", http://www.intel.com/research/mrl/research/opencv/ , Intel Corporation.

[20] Ronan Collobert, "Torch Tutorial", Institut Dalle Molled' Intelligence Artificielle Perceptive Institute, 2 October 2002