Bachelor of Science in Computer Science & Engineering

An Approach to Detect Emergency Vehicle Using Haar-Cascade Classifier

by

Md. Naimul Ahasan

ID: 1504086

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

August, 2021

An Approach to Detect Emergency Vehicle Using Haar-Cascade Classifier

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

Md. Naimul Ahasan
ID: 1504086

Supervised by

Dr. Kaushik Deb
Professor
Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)
Chattogram-4349, Bangladesh.

The thesis titled '**An Approach to Detect Emergency Vehicle Using Haar-Cascade Classifier '** submitted by ID: 1504086, Session 2019-2020 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

Board of Examiners

________________________________ Chairman

Dr. Kaushik Deb
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

________________________________ Member (Ex-Officio)

Dr. Md. Mokammel Haque
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

________________________________ Member (External)

Dr. Mohammad Ibrahim Khan
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

Signature of the candidate

Date:

Acknowledgements

The success and final conclusion of this thesis necessitates a great deal of advice and assistance from many people, and I consider myself quite lucky to have gotten it through the completion of this thesis. Everything I've done has been made possible by such oversight and direction. Professionally and personally, it's been a worthwhile experience. Above all, my gratitude to THE ALMIGHTY is beyond expression.

I am indebted to my esteemed supervisor, Dr. Kaushik Deb, Professor, Department of Computer Science Engineering, Chittagong University of Engineering Technology (CUET), for his advice, encouragement, and unwavering support during my thesis effort. I recognize the full worth of his numerous probing questions that forced me to consider things from new angles, as well as his relentless support throughout the process.

At last I would like to recognize the contribution and invigoration I have received from my family and fellow classmates throughout this thesis work.

Abstract

The number of vehicle is increasing on a daily basis and so is traffic jam, for which emergency vehicles such as ambulance, police vehicle , fire service vehicle can hardly reach the anticipated places on time. This study addresses an approach to detect emergency vehicles on roads or highways. Here, Gaussian Mixture Model is used to subtract background to fasten the detection of emergency vehicle by using Haar features. Meanwhile, YCbCr shadow removal method has been used to improve the detection and edge detection technique Canny method has also been utilized to enhance the precision of vehicle detection. However, taking into account the lighting conditions, darker images are restored performing Histogram Equalization mechanism. Finally Cascade classifier has been performed the verification of the emergency vehicles.

Further studies may add feature like alerting the next traffic junctions to clear a dedicated lane as emergency vehicle is approaching.

Keywords— H istogram Equalization , Gaussian Mixture Model, YCbCr model, Shadow Removal, Canny Edge Detection, Haar-Cascade Classifier

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Digital image processing deals with manipulation of digital images through a digital computer. It is a subfield of signals and systems but focuses particularly on image. DIP focuses on developing a computer system that is able to perform processing on an image. In recent years, we have been using digital image processing to reduce our daily life problems such as vehicle detection, human authentication, object detection, edge detection, machine perception and license plate recognition and many more. In an over populated country like Bangladesh, successful implementation of Intelligent Transport System(ITS) is difficult. Roads are full of static vehicle and in that case, emergency vehicles passing require much more time. That's why every year a large number of people face death or accidents. So a proposition requires to identify the moving emergency vehicles on the road. Here, this method proposed an approach where digital image processing and machine learning have coagulated together to detect emergency vehicles. We understand emergency vehicle as ambulance, fire service vehicles and sometimes police vehicles. This method enables the way of detecting these emergency vehicles so that unwanted occlusion, traffic jam could be avoided by the cars. Due to occlusion, lightning change and other factors the task becomes more laborious. A video based proposal is presented in this paper to eliminate these problems. Initially, foreground objects are separated using Gaussian Mixture based background subtraction model.Meanwhile,median filtering is used to remove the salt-pepper noise. A simple shadow removal technique based on YCbCr color model improves the precision of the vehicle detection. Then, edge detection was performed using Canny method for precise tracking. Finally, for verification of the emergency vehicles a machine learning mechanism known as 'Haar-Cascade' classifier has been used.

1.2 Framework/Design Overview

The procedure is separated into three modules in this suggested technique. Background Segmentation Module is the first of these modules. The picture is captured and then improved from various input sequences such as video frames. Prior to background model adaption, many pre-processing steps are used. Gaussian Mixture Model is one of the finest options for separating foreground and background models. Object detection is the second module. To avoid salt-pepper noise, the foreground picture must be filtered using the Median method. As a result, the identification and elimination of shadows is carried out. The Canny technique will be utilized to detect the objects' edges in this case. As a result, an item in a bounding box will be discovered. The Classifier training module is the third module. The classifier will learn to distinguish between positive and negative images from a huge number of datasets. As a result, a testing image collection will guarantee that the system's accuracy in recognizing emergency vehicles based on characteristics is further enhanced.

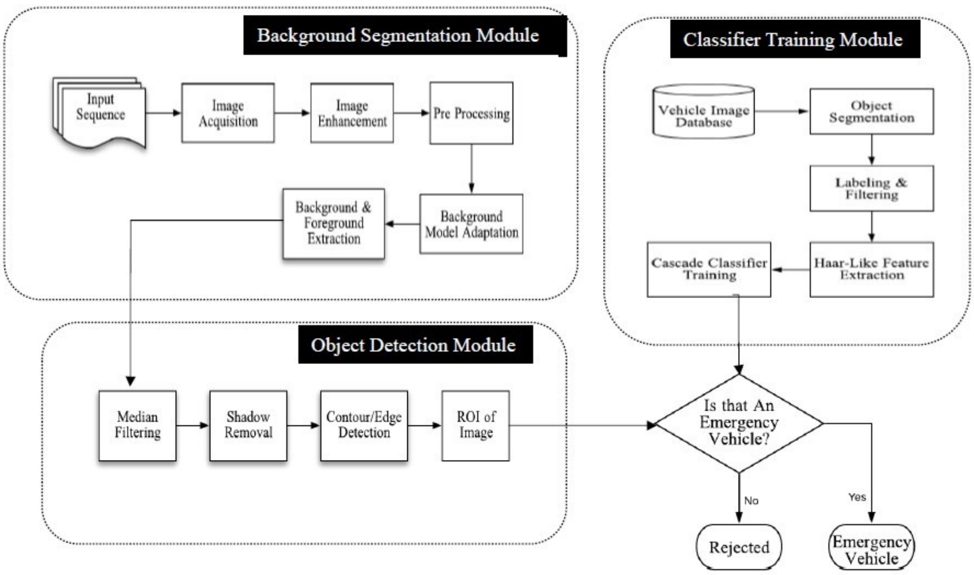


Figure 1.1: Design Overview

1.3 Difficulties

These are some difficulties whose may interrupt the procedure:

- Dataset Collection
- Obscurity in image
- Indifference in Vehicle Color
- Illuminantion variation
- Occlusion on roads
- Shadows
- Noisy images
- Low resolution CCTV footage

1.4 Applications

The most interesting part of the application of this project is to aid the traffic control system in order to avoid waste of time of the emergency vehicles.

- Intelligent Transport System
- Vehicle Recognition
- Overcoming life-threatening situation
- Emergency Vehicle (EV) Identification
- Specify EV from occluded road

1.5 Motivation

Our main theme was to construct a system which can keep pace with the existing systems. Since machine learning provides a vast research field, image processing boost up the entire task along with solving real-life complex problems. Also to be added that, in an overcrowded land of living like Bangladesh, it requires a lot to control a huge number of growing vehicles. In order to maintain the proper impact of the emergency vehicles, this study of differentiating emergency vehicles from others (vehicles existing on the pitch) can be useful at a great manner.

- Overcome wasting of time for traffic jam
- Opportunity to reduce accident rate
- Faster moving of emergency vehicle

1.6 Thesis Organization

The following is how the rest of the report is structured. Background segmentation is described in Sect.3.3.8. Then, in Sect.3.3.10, we'll talk about object extraction. The vehicle verification is discussed in Sect.3.3.11. The outcomes of the experiments are described in Sect. 4. Finally, Sect.5 summarizes the study's findings and potential future directions.

Chapter 2

Literature Review

2.1 Introduction

In recent years the number of vehicle is increasing beyond limit which results in extensive traffic jam in almost all the roads. The most pathetic victims of these traffic jams are the people who are in need of help or urgency. Emergency vehicle can not reach to them on time. In urban areas, vehicle detection is more complex task due to occlusions,lighting changes and much more factors. So far several researchers around the world have worked on vehicle detection and came up with various techniques and methods to overcome those issues.

2.2 Emergency Vehicle Detection Related Literature Review

In this paper [1] a new framework to detect illegally parked vehicle using dual background model subtraction is presented. In this adaptive background model system, the background is generated based on statistical information of pixel intensity that robust against lighting condition. Amidst several collected reference frames , reference background model is generated. Hence Foreground is analysed using geometrical properties, which is then applied in order to filter out false region. Each frame is compared with the previous background model and updates Max and Min buffer comparing pixel intensities. If difference of Min and Max greater than the threshold, it is said to be foreground. Another task is applied here which is Candidate Object Region. It is generated by calculating area, aspect ratio and occupation ration. This method is quite faster than performing image subtraction between current frame and reference background model against illumination changing.

Another approach of car detection is proposed in this paper [2], where Scalable Histogram of Oriented Gradient(SHOG) is used to extract feature. This approach can extract features from image without even resizing it.SHOG divides the gradients in several layer based on orientation. That's why at any fixed dimensional region, it can extract high-discriminated features. Using SHOG instead of Histogram of Oriented Gradient (HOG) results in 3-4% more accurate system. One more thing,instead of sliding window technique , here in [2] , Laser Range Finder (LRF) is used. LRF is associated with camera information which is mounted on car.

In [3] background is modeled using Gaussian Mixture Model (GMM) not like dual modeling in [1]. GMM is highly effective in places with varieties in luminance and high density of traffic as it uses background subtraction model. A median filtering is then applied to remove salt-pepper noise. After segmenting background, shadow in image needs to be removed. [4] To begin, a method for identifying shadows based on statistics of intensity in the YCbCr color space is proposed. A shadow density model is used once the shadows have been discovered. The image is divided into multiple sections with the same density according to the shadow density model. Finally, by relighting each pixel in the YCbCr color space and restoring the coherence, the shadows are eliminated. To detect the contour of object [5] used Canny Edge detection procedure.

Detecting objects based on features necessitates a machine learning-based technique. As a result, given a big dataset, the framework must be trained in order to categorize any samples. In this paper [6] a method called Integral Image is utilized to learn pictures. This method allows to memorize the entire image in a short amount of time as the system needs not to compute the whole image rather combine intensities of previous row and column. A paper [7] describes an algorithm named Adaboost which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers.

[8] In this paper, presented an accurate and robust license plate localization

approach based on the Support Vector Machine (SVM) classifier.Support Vector Machine (SVM) classifier is a simple discriminating supervised learning algorithm. It outputs a hyperplane which classifies new examples into several classes.The hyperplane keeps a margin between the classes.

Next,cascading [9] process discards non-objects very quickly. It is a sliding window technique in which each cascaded stage is used to determine whether a sub window contains vehicle or non-vehicle.A given sub-window is immediately discarded if it fails in any of the stages.

In [10] they have proposed a method of emergency vehicle detection considering only one feature which is the light bar used in the emergency vehicle for alerting while approaching on the road. They took the HSV model of that region and applied SVM classifier to identify the desired vehicle.

2.3 Conclusion

In a nutshell the contribution proposing in this research can be summarized as following- after adapting image from dataset, image preprocessing is required, hence segmented and then Haar feature based cascade classifier can robustly verify different Bangladeshi emergency vehicles including ambulance,fire services etc. within minimal computational cost.

2.3.1 Implementation Challenges

As stated earlier, detecting emergency vehicle faces several challenges such as -

- Illuminantion variation
- Occlusion on roads
- Shadows
- Noisy images

Chapter 3

Methodology

3.1 Introduction

To get higher accuracy on emergency vehicle detection, it necessitates a proper training with lots of image containing emergency vehicles. Different states or countries define several types of vehicle as Emergency Vehicle. Obtaining frames with precisely segmented background and enhancing histogram to smoothly detect emergency vehicle is the main purpose of this study.

3.2 Diagram of Framework

In this proposed method total process is divided into three stages. First one is Background Segmentation Module.Here from several input sequence such as video frame, image is acquired and then enhanced. Several pre processing is happened before background model adaptation.To extract foreground from background model,Gaussian Mixture Model is one of the best preference. Second stage is object detection. Foreground image requires to undergo Median filtering to avoid salt-pepper noise. Hence shadow detection and removal is performed. Here Canny method will be used to detect edge of the objects. Thus object will be detected in a bounding box. Third stage is Classifier training module. From a large number of dataset , classifier will learn identifying positive image and negative image. Hence a testing image set will further ensure the system's accuracy of detecting emergency vehicles based on features.

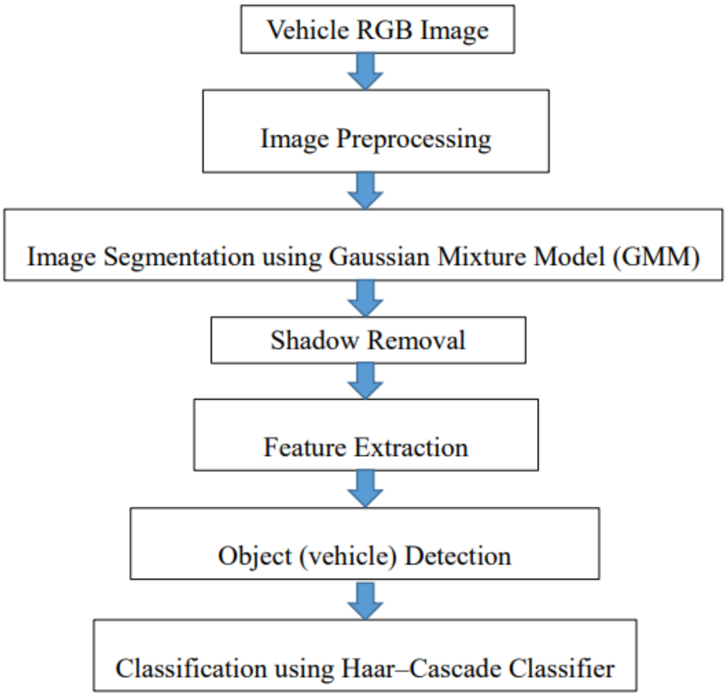


Figure 3.1: Proposed Framework

3.3 Implementation

3.3.1 Input Images

We have considered three classes of emergency vehicle. Ambulance, Police Vehicle and Fire service vehicle. Dataset used in this study, are mostly generated manually. Those images may differ with each other in various properties like– size, color, orientation, lighting condition, number of object in each frame, shadow, noise etc. If we proceed further with such diversities, it will take a great computation and training time which will led classifier to poorly performance. So overcoming those issues is a must.

Detecting object necessitates good imageset. In this regard collected input images may have to process further. As mentioned early, this study we are considering three classes of vehicle. Here is some of the input images-

(a) Ambulance Image

(b) Ambulance Image

(c) Ambulance Image

(d) Ambulance Image

(e) Ambulance Image

(f) Fire Service Vehicle

Figure 3.2: Input Images

3.3.2 Image Preprocessing

3.3.2.1 Image Resizing

RGB images were firstly resized.

- To obtain better efficiency rate for classifier
- Low resolution takes less amount of space
- Better visual perception

High resolution may serve better quality , however taking long processing, training and execution time. In very low resolution, due to lower quality some features become hard to detect. In this case we were needed to trade-off between these two cases and that's why we chose 224×224 size.

(a) Ambulance Image

(b) Ambulance Image

(c) Police Image

(d) Police Image

(e) Fire Image

(f) Fire Service Vehicle

Figure 3.3: Resized Images

3.3.3 Grayscale Conversion

After resizing, RGB images were converted into Grayscale image. Grayscale image is a 8 bit image which is able to express the object details fully without color.

- Its representational and computational complexity is higher than binary image (bitmap or 2 bit image) but lesser than color image (RGB)
- A bit of exquisite
- White = 100% brightness , 0% gray
- Black = 0% brightness , 100% gray

Here is the grayscale images–

(a) Ambulance Image

(b) Ambulance Image

(c) Police Image

(d) Police Image

(e) Fire Image

Figure 3.4: Grayscale Image

3.3.4 Histogram Equalization

Histogram equalization (HE) is a technique for adjusting image intensity to enhance contrast, to improve quality of image and to deemphasize noises. Image enhancement results in improved contrast, enhancing definitions of edge. Histogram equalization can be implemented both globally or locally. It is performed by plotting pixel intensities vs frequency of the pixel intensity.

Image enhancement is required:

- to obtain uniform distribution of intensity
- to improve contrast
- to emphasize on the boundary of the object
- to de-emphasize noise.

Global will focus on the entire image where CLAHE improves local contrast instead of focusing on the entire image.

In road or highway, during daytime, many shadowed region can be found due to sunlight or obstacle by another vehicle or any large tree. In that case, we won't be able to detect the vehicle that remain in that shadowed place. Again, due to sunlight, some portion of the image may shine greatly, there Global HE is not suitable.

In this study, we have tried to use both types of histogram equalization technique. However at first checking pixels' mean intensity of each frame is required, if it is lesser than 127, then in this study CLAHE is applied instead of Global HE.

3.3.5 Global Histogram Equalization

- Firstly take a matrix of a grayscale image
- Count the total number of pixel associated with each pixel intensity
- Then we need to compute (for image with discrete gray values) the probability of intensity

$$Pm(rk) = \begin{pmatrix} nk \\ n \end{pmatrix} \tag{3.1}$$

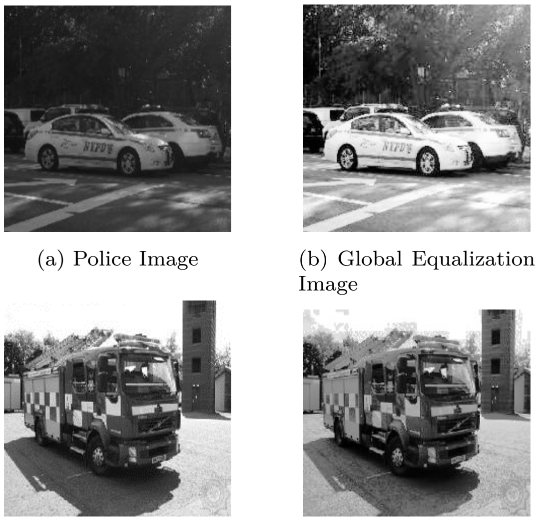
  Where L = Total number of gray level , nk= number of pixel with gray value rk, n = total number of pixels

- Compute cumulative density function

$$sk = T(rk) = \sum_{j=0}^{k} Pm(rj) \tag{3.2}$$

- Multiply cumulative probability by desired image intensity

- Finally apply Floor rounding (round the decimal values to lower integer values)



(a) Police Image
(b) Global Equalization Image
(c) Fire Service Image
(d) Global Equalization Image

Figure 3.5: Global Histogram Equalized Image

3.3.6 CLAHE

To bring out features in small portion of the image local histogram equalization is applied. Adaptive histogram equalization is an approach of local histogram equalization which computes several histograms and enhance local contrast along definition of edge. However it overamplifies the noise in relatively homogeneous region. In this study we have considered CLAHE to meet this issue. CLAHE is Contrast Limited Adaptive Histogram Equalization. It limits the amplification by clipping histogram. Here is how it performs-

- At first divides the entire image into tile / block of (M×M) size
- Normalize each tile
- Determine clip limit from the normalized value
- Pads the image before splitting into regions (if necessary)
- In noisy region, clip the pixel that rises above the specified limit and hence distributes uniformly to the other bins
- To remove neighboring artifacts , use bilinear interpolation

Tile Grid size = Number of rows and columns in each tile

In addition, by applying CLAHE, we may see a great variation in mean intensity. Here tile grid size is 16×16 as it provides optimal result.

(a) Police Image (b) Local Equalization Image

Figure 3.6: Local Histogram Equalized Image

(a) Fire Service Image (b) Local Equalization Image

Figure 3.7: Local Histogram Equalized Image

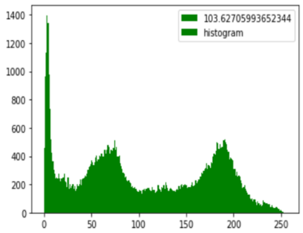
3.3.7 Histogram Comparison

Histogram is the mapping of number of pixels per intensity. It can be generated by plotting pixel intensity versus frequency of pixel intensity or probability of pixel intensity.

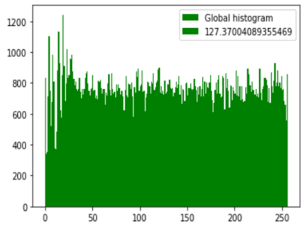
Here, if we have a look in the histograms of the actual image, globally and locally histogram equalized images, we can see there is no change in mean intensity value during graylevel and globally equalized image. However after applying CLAHE mean intensity greatly changed.

The reason is CLAHE improves local contrast while Global HE takes entire image to enhance. These led the drastic change in mean intensity. As stated earlier,in this study, apply CLAHE when mean intensity is less than 127.

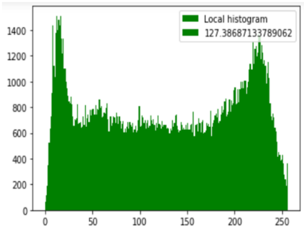
(a) Ambulance Image



(b) Histogram



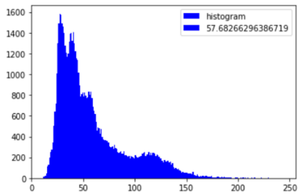
(c) Global HE



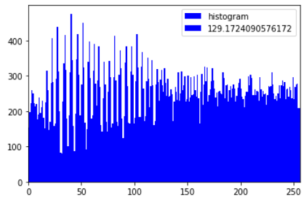
(d) Local Histogram

Figure 3.8: Histogram Comparison of Image

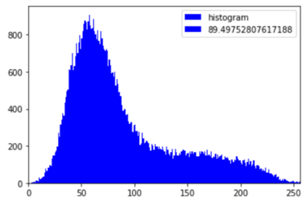
(a) Police Vehicle



(b) Histogram



(c) Global HE



(d) Local Histogram

Figure 3.9: Histogram Comparison of Image

3.3.8 Background Segmentation

Background segmentation is required to extract foreground objects from the scene. Several technique could be applied for background subtraction. Methods such as: Adaptive dual background model, Gaussian Mixture Model (GMM) , K-means algorithm etc are available for background segmentation. However before doing so, we need to understand our requirements explicitly. Before constructing any system for emergency vehicle detection, there are some cases that should be considered, such as:

- Real time analysis of data
- Object overlapping in the visual field
- Shadows & lighting change conditions
- Slow-moving objects

Since a non-adaptive background model requires manual initialization, such techniques are not suitable for this particular project.

We have implemented GMM to subtract background.

3.3.8.1 Gaussian Mixture Model

Gaussian Mixture Model (GMM) is a mixture of K Gaussians describing the distribution of a random variable x. Gaussian Distribution is a probability distribution that explains a listing of outcomes and probabilities associated with each outcome of an experiment. It is represented by a bell shaped curve and the probability density function is–

$$\eta = (2\pi)^{-0.5}\sigma e^{-0.5(xt-\mu)\Sigma^{-1}(xt-\mu)} \tag{3.3}$$

Here, µ = mean value and sigma = standard deviation and

$$sigma^2 = variance \tag{3.4}$$

For a Gaussian mixture model with k components the probability of a random variable is given by-

$$P(xt) = \sum_{i=1}^{k} \eta\omega i, t \tag{3.5}$$

GMM involves two process.

1. Bakground Modeling: Firstly, it is assumed that there is no foreground information in the initial frame. The background is modeled using the initial frames.Then each pixels are modeled as the mixture of Gaussians rather than a particular type of distribution.Based on the variance of each of the Gaussian of the mixture, it is determined which Gaussians may correspond to background colors. Pixel values that do not fix the background distributions are considered foreground until there is a Gaussian that includes them with sufficient consistent evidence supporting it. The background is modeled based on the probability of observing a particular pixel value. The Gaussian mixture components for a pixel have normalized weights calculated from the past observations. Modeling Process-

- Considering, all image points are mutually independent.
- The Intensity distribution of every point at any time t is modeled as a mixture of K . K is determined by on-line K-means approximation.
- Then we have to model a Gaussian evaluating the parameters such as weight (w), mean (µ), standard deviation of initial frame.
- Matched Gaussians having a value greater than an appointed threshold(T) is defined as background

$$B = argminb(\sum_{i=1}^{b} \omega i, t > T) \tag{3.6}$$

Updating

- If a pixel matches with one of the k Gaussian then the value of w, µ and standard deviation is updated. Here are the equations:

$$\omega i, t+1 = (1-\alpha) \cdot \omega i, t + \alpha \tag{3.7}$$

$$\sigma^2 i, t+1 = (1-\rho)\sigma^2 i, t + \rho(Xi, t+1 - \mu i, t+1)(Xt+1 - \mu i, t+1)^T \tag{3.8}$$

$$\mu i, t+1 = (1-\beta)\mu i, t + \rho Xt + 1 \tag{3.9}$$

- If there is k number of Gaussians that do not match then only weight is updated as follows:

$$\omega i, t+1 = (1-\alpha)\omega i, t \tag{3.10}$$

3.3.9 Shadow Removal

After finding the foreground median filtering is applied to remove the salt-peeper noise. Due to various reason, luminance can be differed in different images which

(a) Ambulance Image (b) Police Image (c) Fire Image

Figure 3.10: Background Segmented Image using Gaussian Mixture Model

create shadows in images. Shadow is lighter than real black portion and remain in different plane from the object. In that case, we have to perform shadow detection and removal technique to obtain actual edge of the objects in the image. In this study we have implemented [4] technique of removing shadows based on YCbCr color model. Steps are as follows:

- Convert RGB image to YCbCr image
- Compute mean and standard deviation of the Y-plane
- Detecting shadow/non-shadow If current pixel's intensity is

$$X_t$$

$$Xt < (\mu - \frac{\sigma}{3}) \tag{3.11}$$

  then the region is detected as shadow and paint white otherwise it is non-shadow or black region.

- Perform morphological operation Closing to reject misclassified pixels
- Find the average pixel intensity in the lit areas and the shadow areas.
- Then derive their difference and ratio.
- Add these differences and ratios with the YCbCr plane.
- Hence by converting the YCbCr image into RGB image, we will obtain shadow removed image

(a) Ambulance Image With Shadow

(b) Shadow Removed Image

Figure 3.11: Shadow Detection and Removal

3.3.10 Canny Edge Detection

After performing shadow removal technique we need to detect the edge of the object.Canny Edge Detector is a widely-used edge detection algorithm by John F. Canny. It's a multi-stage algorithm. The processes are describing below:

- **Noise Reduction** – At first it performs smoothening the image by a 5×5 Gaussian filter.

- **Intensity Gradient & Orientation** – Smoothened image is then filtered with a Sobel kernel in both horizontal and vertical directions to get 1st derivative in both axes respectively Gx & Gy.

$$G = \sqrt{(Gx)^2 + Gy)^2} \tag{3.12}$$

$$\theta = \tan^{-1} \frac{Gy}{Gx} \tag{3.13}$$

- **Non-Maximum Suppression** – Scan full image and reject pixels that contain no edges. At each pixel, pixel is checked if it's a local maximum in its neighborhood in the direction of gradient.

- **Hysteresis** – In this stage edges are ensured. 2 level threshold is used. MinVal, MaxVal. If

  - edge intensity gradient maxVal = Sure Edge
  - edge intensity gradient < minVal = Non Edge
  - edge intensity gradient between minVal maxVal, then check connectivity with any sure edge

The main reason of using Canny edge detection is that, it's stage of Hysteresis gives almost pure edges. Considering this, weak edges are rejected if they are not connected with any sure edges. In the above image, Edge A is above the maxVal,

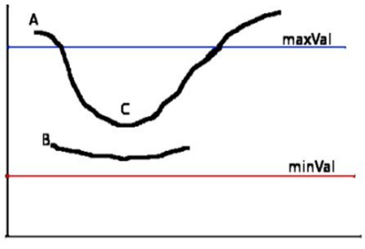
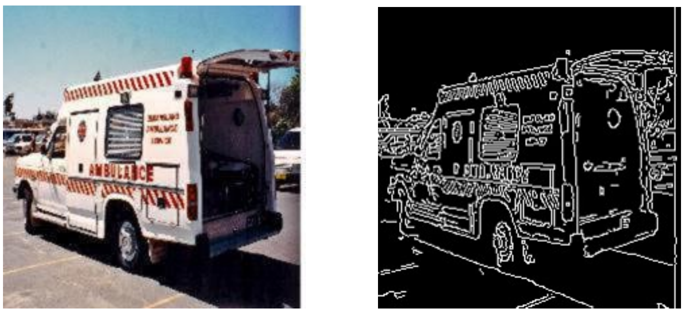


Figure 3.12: Hysteresis in Canny Method

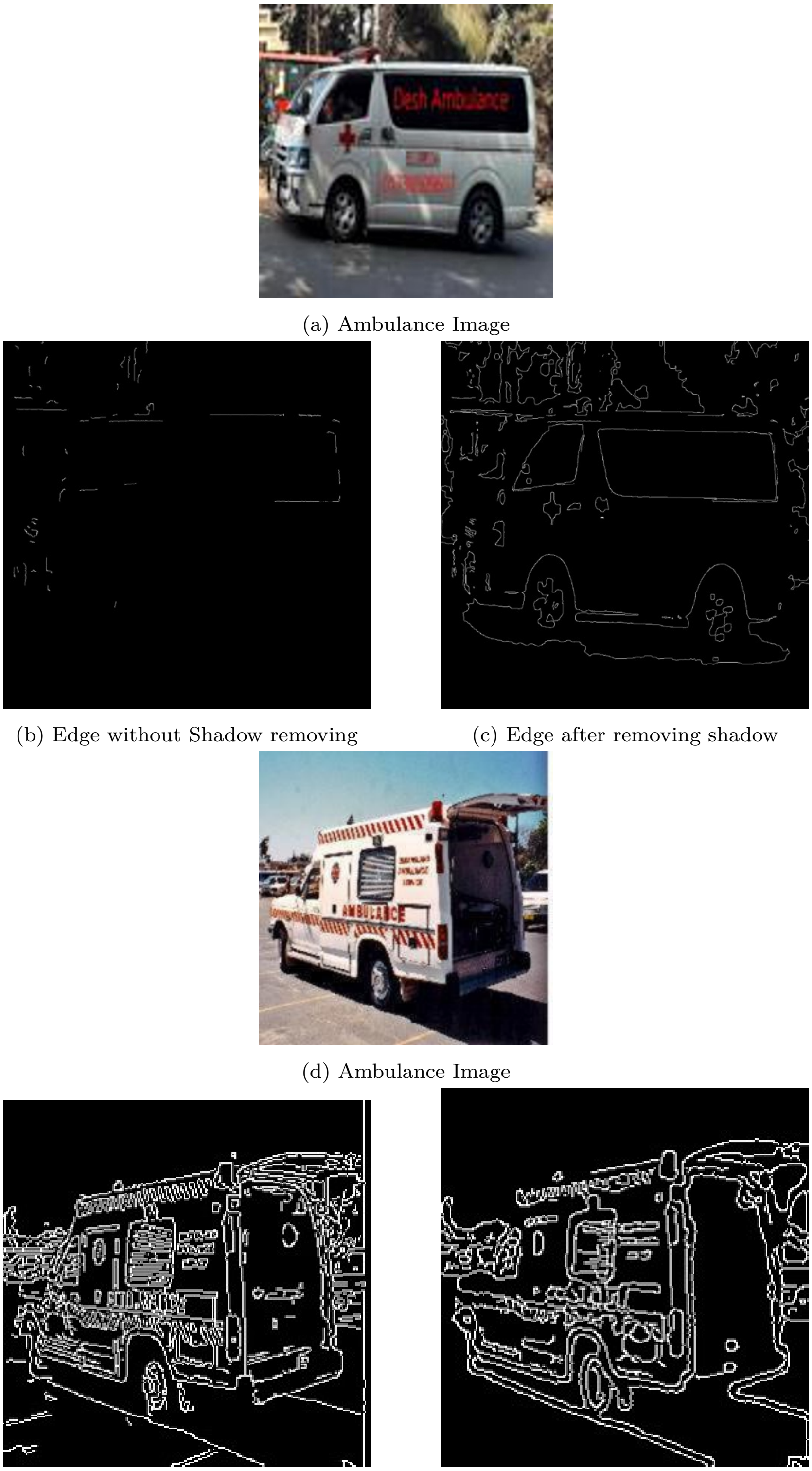
so it is "sure-edge". Although edge C is below maxVal, it is connected to edge A, so that also considered as valid edge and a full curve is formed. But edge B, although it is above minVal but it is not connected to any "sure-edge", so it is discarded. This stage also removes small pixels noises on the assumption that edges are long lines.



(a) Ambulance Image (b) Detected Edge

Figure 3.13: Example of edge detection

Extracting Region of Interest These examples mentioned below, are a demonstration of why shadow removal is important. It results in precisely finding the object region as well as edge detection. Shadow removal technique helps to extract features from the foreground. This also assists the classifier to smoothly identify the object as emergency vehicle.



(a) Ambulance Image

(b) Edge without Shadow removing

(c) Edge after removing shadow

(d) Ambulance Image

(e) Edge without Shadow removing

(f) Edge after removing shadow

Figure 3.14: Example of edge detection after removing shadow

3.3.11 Vehicle Verification Using Haar Features Based Cascade Classifier

Haar Cascade Classifier is a fast, widely used object detection method. It was first mentioned by Viola & Jones. It has 4 subtopics whose are –

- Haar Features
- Integral Image
- Adaboost Training
- Cascade Classifier

Haar Features

Haar Features are sequence of rescaled square shapes function that are masked over an image. Each rectangle has a number of pixels. So, firstly, sum of the pixels in each rectangle should be calculated. Then difference between Black region and white region has to be determined. At first feeding the system with lots of positive and negative image. System will generate Haar features from the positive images. In a typical 24X24 window, almost 200K features can remain, although everyone is not useful. To get a faster output along higher accuracy the system will get rid of less useful features. These are some Haar features:

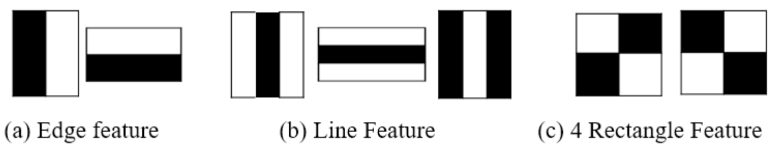


Figure 3.15: Haar Features

Integral Image

To avoid such a huge number of non useful features and detect Haar features in a more fast manner, the idea Integral Image was introduced. Here, integral images form by adding intensities of previous row and column. By this approach only the four corner pixel values of a rectangle need to calculate instead of entire image. Here is a sample:

Here in this image, we are considering a simple edge feature is applied over the sample image. At first an integral image will generate by adding each pixels with it's previous row and column values successively. The first pixel value is 10 , so at the second pixel it will increment and become 15. Third pixel value will be updated as 40. Then second row first column pixel value will be updated as 98 because 88 will be added with its previous value which is 10. But second row,

10	5	25	82	42	13	63	46
88	64	35	39	66	21	15	44
2	45	78	24	42	54	90	62
4	12	46	73	36	25	85	46
35	66	3	56	42	73	85	25
56	98	24	1	50	53	42	14
78	46	62	85	53	27	26	58
34	52	26	52	29	47	24	34

(a)

(b)

Figure 3.16: (a) Sample Image with an Edge feature, (b) Sample Edge Feature

second column value 64 will be updated as 167 by adding 98 (updated previous row pixel value), 5 (previous column pixel value), 64 (previous pixel value) Thus an integral image will generate.

10	15	40	122	164	177	240	286
98	167	227	348	456	490	568	658
100	214	352	497	647	735	903	1055
104	230	414	632	818	931	1184	1382
139	331	518	792	1020	1206	1544	1767
195	485	696	971	1249	1488	1868	2105
273	609	882	1242	1573	1839	2245	2540
307	695	994	1406	1766	2079	2509	2832

Figure 3.17: Integral Image

In a 100X100 image, instead of computing 10,000 pixels, it computes only four operations. Thus it lessens the computational time. Here, in this integral image instead of computing overall pixels, this approach will compute only the pixels of the masked areas previous row and column. This process will reduce a lot of weak features in a manner to fasten the computational process. This approach only scales the feature instead of the image. At first it will

- Calculate sum of pixels of dark and white regions
- Calculate the difference of dark and white region of the edge feature

Here sum of the lower or white region pixels is 419. Now subtracting last pixel value of dark region from white region and again for their previous row values.

12	46	73	36
66	3	56	42
98	24	1	50
46	62	85	53

(a)

10	15	40	122	164
98	167	227	348	456
100	214	352	497	647
104	230	414	632	818
139	331	518	792	1020
195	485	696	971	1249
273	609	882	1242	1573

(b)

Figure 3.18: (a) Pixels of Edge feature, (b) Four corners of the Edge Feature

1573-1020 = 553 ,

139-273 = -134.

Adding these two value we obtain the sum of pixels of bright region.

553 + (-134) = 419

We can consider another example, this time with another Haar feature Line feature applied over the same sample image.

10	5	25	82	42	13	63	46
88	64	35	39	66	21	15	44
2	45	78	24	42	54	90	62
4	12	46	73	36	25	85	46
35	66	3	56	42	73	85	25
56	98	24	1	50	53	42	14
78	46	62	85	53	27	26	58
34	52	26	52	29	47	24	34

(a)

10	15	40	122	164	177	240	286
98	167	227	348	456	490	568	658
100	214	352	497	647	735	903	1055
104	230	414	632	818	931	1184	1382
139	331	518	792	1020	1206	1544	1767
195	485	696	971	1249	1488	1868	2105
273	609	882	1242	1573	1839	2245	2540
307	695	994	1406	1766	2079	2509	2832

(b)

Figure 3.19: (a) Pixels of Line feature, (b) Four corners of the Line Feature

The line feature applied above has shown later this paragraph. In the above sample image, at first an integral image will generate like the previous example. Hence, sum of the indifferent region, more accurately sum of the middle black region pixels need to be calculated. Which is 280. Now calculating four corner values. Black region's last value 818, white region's 414. Previous column value of black and white portion is respectively 164 and 40. So, 818+40-164-414 = 280

By this method, integral image helps to understand strong and weak features.

Subsides weak features, produce an integral image for adaptive boosting method of image training.

Strong feature is the combination of weak features multiplied by their respective weights.

- At first all images are given same weight. Hence we have to normalize the weight.

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{N} w_{t,j}}; N is the number of image. \tag{3.14}$$

- Then we need to detect error.

$$\in_t = \sum_i w_{t,i} |f_t(x_i) - y_i| \tag{3.15}$$

A feature is applied over all images. If it can detect correctly, then error e = 0. If not error, e = 1.

- Chose the lowest error by multiplying error and significance.
- Then weights are updated.

$$w_{t+1,i} = w_t \times \beta^{1-e_t} \tag{3.16}$$

- $$\beta_t = \alpha_t \times \frac{e_t}{1 - e_t} \tag{3.17}$$

  , where alpha is the significance

Finally, the strong feature F(x) will be 1 if sum of the weighted features is half of the sum of the weights, otherwise zero.

F(x) = 1 , if

$$\sum_{i=1}^{t} \alpha_t \times f_t(x) \geq 0.5 \times \sum_{i=1}^{t} \alpha_t \tag{3.18}$$

AdaBoost Training

Integral image quickly computes detector features. A learning system called AdaBoost [6] selects critical visual elements from a wider collection. This approach produces classifiers that are computationally efficient.

Cascading

This is the last step of object detection method used in this study. Cascading is the process of checking if a window is not an object region.A cascade of classifier consists of multiple stages of filters. This process discards non-objects rapidly. Here,

- A window will generate to check the object region.
- Firstly the first classifier with highest weight found earlier will perform. It will directly reject the non-vehicle features.
- If first feature is approved, then it will shift to the second classifier
- Like this, if all the features are approved, the object will be detected.

10	15	40	122	164	177	240	286
98	167	227	348	456	490	568	658
100	214	352	497	647	735	903	1055
104	230	414	632	818	931	1184	1382
139	331	518	792	1020	1206	1544	1767
195	485	696	971	1249	1488	1868	2105
273	609	882	1242	1573	1839	2245	2540
307	695	994	1406	1766	2079	2509	2832

Figure 3.20: Cascading of Classifier for finding region of interest

Two things to be added, we needed to set minimum neighbor and scale factor value. A threshold is set for the number of true outputs required to detect an object. If the computed value of minimum neighbor is lower than the threshold than the framework will detect as many objects as it can. It will ultimately give more false result. On the contrary, If scale factor is higher, then it will require more computational time however provides a better result.

Chapter 4

Results and Discussions

4.1 Introduction

The procedure for detecting emergency vehicles was elaborated in the previous chapter. In this chapter, the performance of the recommended method is briefly assessed. This framework is built using Opencv, a computer vision technology. The method was completed using a PC with 4GB of RAM and a Core i5 processor. The total run time of pre-processing and training time of the process was around 89 minutes. It took the system about 1.052 second to detect the output. Several machine learning algorithms are also discussed and contrasted in this chapter.

4.2 Dataset Description

The resolution of the images in the dataset collection ranges from 44x98 to 3600x3050. They were reduced in size to 224x224 pixels, as described in the previous chapter. 75% of the data is used for training and 25% that is for testing and validation in this situation. The dataset is imbalanced. We were focused to ensure a good outcome through our method. Almost 1600 images were taken as dataset where 650 are ambulance image, 455 images of Bangladesh Police Vehicle along with 495 images of fire service vehicles.

4.3 Impact Analysis

4.3.1 Social and Environmental Impact

To ensure a country's digitization, intelligent transport system is inevitable. Lack of proper road safety, road distribution policy and digitized roads has caused so many accidents in recent years. Bangladesh has not introduced with emergency lane yet, which causes emergency vehicles to stuck in usual traffic jam etc. This project can be used to detect emergency vehicles from occluded road scene and hence notify next crossing station to make way for the approaching emergency vehicle.

4.3.2 Ethical Impact

When a choice, event, or action violates a society's moral norms, ethical dilemmas arise. Individuals and corporations alike could be interested in these disputes developed as a result of the possibility that all of their operations would be questioned on ethical grounds. Our technology wishes the best for traffic control police by teaching them how to spot emergency vehicles early, make room for them and save their time.. It is in their best interests as well as the advancement of our country's modernization.

4.4 Evaluation of Metrics

4.4.1 Confusion Matrix

A confusion matrix is a metric that demonstrates how well a classification model works on a set of test data for which the true values are known. It's great for figuring out Recall, Precision ; specifically Accuracy.

The terms used in Confusion Matrix are as follows:

- **True Positive (TP):** The model predicted positive outcomes which is validated by the final label.
- **True Negative (TN):** The model predicted negative result but which is verified by the final label.
- **False Positive (FP):** The model predicted a positive result, while the actual outcome is negative.
- **False Negative (FN):** The model anticipated a negative outcome, while the final result is a positive one.

Some of the other assessment measures we use in our study are as follows:

Accuracy: Overall, what percentage of the time does the classifier get it right is measured by accuracy. This takes into consideration both TP (True Positive) and TN (True Negative). To determine accuracy, apply the following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision: Shows how many of the precisely anticipated situations turned out to be positive. Precision may be calculated using the following formula:

$$Precision = \frac{TP}{TP + FP}$$

Recall: This number indicates how many real positive cases our model correctly predicted. Recall may be calculated using the formula below:

$$Recall = \frac{TP}{TP + FN}$$

F1 score: Aids in the evaluation of recall and precision at the same time. The F1 score may be calculated using the formula below:

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

4.5 Evaluation of Framework

The main objective of this proposed method is to recognize emergency vehicle from road images, CCTV footage. Throughout the method, Haar-Cascade classifier was employed. Our framework successfully identified the vehicle class with an accuracy of 94.56%. Where other frameworks only identified if its static or not, emergency or not; however we experimented with both. In comparison to previous research, the suggested framework is obviously the superior option for a combined strategy.

4.5.1 Confusion Matrix for Emergency Vehicle Detection Framework

The confusion matrix has been used to measure the outcome of our emergency vehicle detection and it's noticeable that certain classes were misclassified. The confusion matrix used to evaluate the data, as well as the number of correctly identified and incorrectly identified vehicles are shown in Figure. The vehicles that was successfully recognized is shown by the diagonal region among the total number of images in each class.

		Predicted		
		Ambulance	**Police Vehicle**	**Fire Service Vehicle**
Actual	**Ambulance**	619	7	24
	Police Vehicle	4	427	24
	Fire Service Vehicle	8	18	469

Table 4.1: Confusion Matrix for Emergency Vehicle Detection Framework.

4.6 Evaluation of Performance

After deciding on the Haar-Cascade framework for the goal, we have compare our findings to those of other techniques. In this part, we'll compare our strategy against a number of machine learning approaches as well as previous research. Finally, the entire architecture's outcomes have been demonstrated. All of the preceding procedures are carried out using video sequences shot in various lighting conditions (day, night, uneven), and in various vehicles. We've included examples of picture sequences that meet all of the aforementioned requirements.

Emergency Vehicle Class Name	Precision	Recall	F1 Score	Accuracy
Ambulance	0.981	0.952	0.966	0.952
Fire Service Vehicle	0.945	0.938	0.941	0.938
Police Vehicle	0.907	0.947	0.927	0.947
Average (%)	94.43%	94.61%	94.47%	**94.56%**

Table 4.2: Performance Parameter Values for the proposed Haar-Cascade Classifier.

The results of detecting emergency vehicles are displayed in the table above. Calculating recall and precision helped to determine the sample's detection accuracy. For the custom dataset that we provided, the suggested system had an average precision of 94.43 % and a recall of 94.61%. Most of the false positive and false negative results were generated due to vehicle misidentification and inadequate backdrop modeling in congested situations.

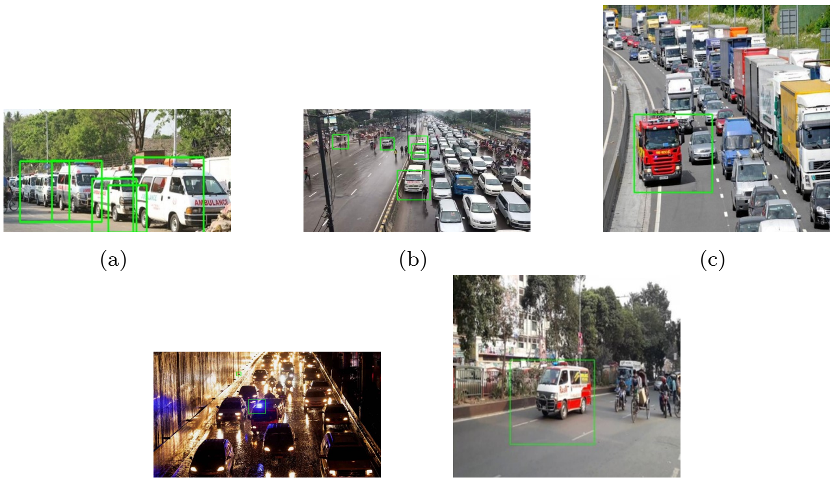
4.6.1 Time Complexity Analysis

Average time complexity of the proposed framework is shown in , where it can be seen that the proposed method can perform its steps within limited time.

Background Subtraction (second)	Shadow detection (second)	Vehicle detection (second)	Classification (second)	Average computational time (second)
0.07	0.47	0.399	0.113	1.052

4.6.2 Output of our proposed framework

We tested with various images after we have trained our model. Classification of predicted output of the emergency vehicle classes are given. As discussed the arrays represent detection of ambulance, police vehicle and fire service vehicle from an image.



(a) (b) (c)

(d) (e)

Figure 4.1: Classification Output results of test images.

4.7 Conclusion

A demonstration of our framework is included in this chapter. This chapter contains the results of the emergency vehicle detection using our chosen Haar-Cascade Classifier. In addition, we compared among various methodologies. The thesis project comes to an end in the following chapter.

Chapter 5

Conclusion

5.1 Conclusion

We provide a framework for recognizing emergency vehicles in the following study to improve existing traffic monitoring systems. The key benefit of this dissertation work is the ability to recognize emergency vehicles and analyze them in real time. We used the Gaussian Mixture Model to separate foreground items from the background in a more smooth way, which improved visual perception. This approach is adaptable to changes in lighting or appearance, which led in a few erroneous predictions for a large dataset. To eliminate the salt-pepper noise from the foreground items, median filtering is used. Following that, a shadow detection procedure is utilized to eliminate shadows from foreground signals in order to avoid object misclassification. As a result, Canny edge detection is used to locate the edges of objects. For vehicle verification we have trained Haar feature based cascade classifier, which is well known for robustness. Hence finishing the project we faced lots of challenges, eventually which led to improper accuracy of this work.

5.2 Future Work

The proposed framework begins by modeling the background using the Gaussian Mixture model, with all objects assumed to be the background at first. The algorithm will not be able to detect an emergency vehicle if it remains in the scene until the background is modeled. Again, the system is not highly capable enough of distinguishing other vehicles whose are adjacent in an image of occluded roads.

This problem can be solved by cumulative background modeling. Vehicle identification can be improvised by implementing Convolutional Neural Network,which may deliver greater accuracy in busy areas with numerous people.

References

[1] A. Filonenko, K.-H. Jo *et al.*, 'Illegally parked vehicle detection using adaptive dual background model,' in *IECON 2015-41st Annual Conference of the IEEE Industrial Electronics Society*, IEEE, 2015, pp. 002 225–002 228 (cit. on p. 5).

[2] V.-D. Hoang, L. Kurnianggoro, K.-H. Jo *et al.*, 'Scalable histogram of oriented gradients for multi-size car detection,' in *2014 10th France-Japan/8th Europe-Asia Congress on Mecatronics (MECATRONICS2014-Tokyo)*, IEEE, 2014, pp. 228–231 (cit. on p. 6).

[3] A. Mutsuddy, K. Deb, T. Khanam and K.-H. Jo, 'Illegally parked vehicle detection based on haar-cascade classifier,' in *International Conference on Intelligent Computing*, Springer, 2019, pp. 602–614 (cit. on p. 6).

[4] K. Deb and A. H. Suny, 'Shadow detection and removal based on ycbcr color space,' *SmartCR*, vol. 4, no. 1, pp. 23–33, 2014 (cit. on pp. 6, 18).

[5] X. Wang and J.-Q. Jin, 'An edge detection algorithm based on improved canny operator,' in *Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)*, IEEE, 2007, pp. 623–628 (cit. on p. 6).

[6] P. Viola and M. Jones, 'Rapid object detection using a boosted cascade of simple features,' in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, Ieee, vol. 1, 2001, pp. I–I (cit. on pp. 6, 26).

[7] H. Fleyeh, R. Biswas and E. Davami, 'Traffic sign detection based on adaboost color segmentation and svm classification,' in *Eurocon 2013*, IEEE, 2013, pp. 2005–2010 (cit. on p. 6).

[8] D. Yang, J. Kong, N. Du, X. Li and X. Che, 'A novel approach for license plate localization based on svm classifier,' in *2010 2nd IEEE International Conference on Information Management and Engineering*, IEEE, 2010, pp. 655–660 (cit. on p. 6).

[9] B. Ashwini, B. Yuvaraju, A. Y. Pai and B. A. Baliga, 'Real time detection and classification of vehicles and pedestrians using haar cascade classifier with background subtraction,' in *2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, IEEE, 2017, pp. 1–5 (cit. on p. 7).

[10] H. Razalli, R. Ramli and M. H. Alkawaz, 'Emergency vehicle recognition and classification method using hsv color segmentation,' in *2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, IEEE, 2020, pp. 284–289 (cit. on p. 7).