Driver Assistance System

Project report submitted in partial fulfillment
of the requirements for the degree of

Bachelor of Technology
in
Electronics and Communication Engineering

by

Aakash Negi - 15UEC001
Abhimanyu Singhal - 15UEC002
Chaitanya Maheshwari - 15UEC016

Under Guidance of
Dr. Joyeeta Singha

Department of Electronics and Communication Engineering
The LNM Institute of Information Technology, Jaipur

April 2018

The LNM Institute of Information Technology
Jaipur, India

CERTIFICATE

This is to certify that the project entitled Driver Assistance System , submitted by Aakash Negi (15UEC001), Abhimanyu Singhal (15UEC002) and Chaitanya Maheshwari (15UEC016) in partial fulfillment of the requirement of degree in Bachelor of Technology (B. Tech), is a bonafide record of work carried out by them at the Department of Electronics and Communication Engineering, The LNM Institute of Information Technology, Jaipur, (Rajasthan,) India, during the academic session 2018-2019 under my supervision and guidance and the same has not been submitted elsewhere for award of any other degree. In my/our opinion, this thesis is of standard required for the award of the degree of Bachelor of Technology (B. Tech).

Date

Adviser: Dr. Joyeeta Singha

Dedicated to Our Families, Friends and Teachers

Acknowledgments

We would like to thank The LNMIIT for providing us with ample knowledge, opportunities and facilities which are playing substantial role in helping us overcoming the challenges faced by us in this project. We would like to wholeheartedly thank our project guide, Dr. Joyeeta Singha for her humongous support and guidance in the B.Tech. Project and also in our lives in college, which helped us a lot in overcoming a lot of hurdles. Sincere thanks to all the persons who knowingly or unknowingly helped us in completing this project. We would like to thank our parents for being the primary reasons in whatever we have gained and learned so far. Thank you for making us who we are today. Thank you for all your teachings which have proven indispensable in developing this project.

Abstract

Crediting the inefficient modes of transportation we have, wheels are always in motion towards their betterment, be it in fields of energy, pollution, or convenience. The upcoming big step in the same is self-driving cars. Albeit different companies or their subsidiaries, like Uber, Alphabet's Waymo, Tesla are making strides towards actual road trials of such supervised or unsupervised cars, task of deploying such a vehicle in a country like India is tedious and will take decades, citing our road conditions, and weird civic sense, where people don't give weight to things like zebra crossings or indicators etc. Meanwhile, number of accidents involving human errors on road are on rise. To deal with this, our system curtails drowsy driving which is one of the major causes of road accidents, by alerting the driver when he/she shows symptoms of drowsiness and helps them identify objects and road on the path by means of image processing.

Contents

Chapter | Page

1 Introduction . . . 1
- 1.1 The Area of Work . . . 1
- 1.2 Problem Addressed . . . 2

2 Driver Drowsiness Detection . . . 4
- 2.1 Existing Algorithms . . . 4
  - 2.1.1 Viola–Jones Object Detection Framework . . . 4
  - 2.1.2 Principal Component Analysis Method (Eigenfaces) . . . 4
    - 2.1.2.1 Initialization . . . 5
    - 2.1.2.2 Recognition . . . 5
  - 2.1.3 Linear Discriminant Analysis Method (Fisherfaces) . . . 5
- 2.2 Literature Survey . . . 6
  - 2.2.1 Database . . . 6
  - 2.2.2 One Millisecond Face Alignment Method . . . 6
  - 2.2.3 Eye Tracking and Detection . . . 7
  - 2.2.4 Mouth Tracking and Detection . . . 8

3 Road Detection . . . 9
- 3.1 Existing Algorithms . . . 9
  - 3.1.1 Road Detection based on Classification Algorithm . . . 9
  - 3.1.2 K-Nearest Neighbours Classifier . . . 9
  - 3.1.3 Naive Bayes Classifier . . . 9
  - 3.1.4 Road Detection Using Road Models . . . 10
  - 3.1.5 Segnet . . . 10
- 3.2 Literature Survey . . . 11
  - 3.2.1 Database . . . 11
  - 3.2.2 Model . . . 12
    - 3.2.2.1 Network Architecture . . . 12
    - 3.2.2.2 Initial Block . . . 12
    - 3.2.2.3 Bottleneck modules . . . 13

4 Simulation and Results . . . 14
- 4.1 Eye And Mouth Detection And Tracking . . . 14
  - 4.1.1 Sample Images . . . 14
  - 4.1.2 One Millisecond Face Alignment Method . . . 15

4.1.3 Facial Contours Detected Using Our Code 15
4.1.4 End Results 16
4.2 Road And Object Detection And Tracking 17
4.2.1 Sample Images 17
4.2.2 Semantic Segmentation Using ENET Method 18

5 Conclusions and Future Work 19
5.1 Conclusion 19
5.2 Proposed work for next Semester 19

Bibliography 21

Chapter *1*

Introduction

1.1 The Area of Work

Our area of work is related to computer vision, image processing and analysis, and neural networks, particularly Residual Nueral Network. Computer Vision is the art of distillifying information from images, on which operations can be performed. Algorithm development is crucial to both, image processing and computer vision, because each situation is uniquely different, and requires several design iterations.

Digital image processing makes use of computer algorithms to do image processing on digital images. A subcategory of digital signal processing, it has many merits over analog image processing. Various algorithms can be applied on our input image, and noise and signal distortion can be dealt in a more efficient manner, in case of digital image processing.Digital image processing may be modeled in the form of multidimensional systems, as images are generally in the form of two or more dimensions.

Computer vision includes techniques for acquiring, processing, analysing and understanding digital images, and extraction of real high-dimensional data to produce numerical or symbolic output, which can be further made to operate upon. A scientific discipline, computer vision deals with the theory behind artificial systems that extract information from digital images. The data can be of various forms, like video streams, different camera angles etc. A technological discipline, computer vision applies its theoritical and modular approach to design complete computer vision systems. Further sub categories include image restoration, object recognition, motion estimation, video tracking, scene reconstruction, event detection, 3D pose estimation, learning and indexing.

Neural networks, as the name suggests are connection systems that are vaguely similar to how animal brains work. Neural networks are singular, when it comes to extract patterns or trends, or give meaning to seemingly random data to humans, as well as conventional computers. Different models of neural networks based on different algorithms can be trained to analyse plethora of data and find association, correlation, patterns or trends throughout. Further, it can be thought of that the trained model 'knows' everything about the data, including analytics of data, and can be further used to provide predictions, when presented with new situations and questions. Residual Neural Network, or ResNet is a sub category of neural networks, which behaves on constitution of pyramidial cells in the cerebral cortex. ResNet makes use of short-cuts or skip connections to jump over layers, with a limitation of jumping over just a single layer.

1.2 Problem Addressed

Self-driving flying vehicles, as seen in the Jetsons are inevitable, but now we are standing at the cusp of being acquainted with self-driving cars, as our to-go mode of transport, but that is not the case in India. Although there are several startups in India which are working on self-driving vehicles, such as Flux Auto, Fisheyebox, Hi Tech Robotic Systemz, ATImotors, Netradyne, Swaayatt Robots, Auro Robotics, OmniPresent Robot, SeDrica 1.0, but the technologies are in nascent state, as they are in nascent state, as live on-road trials is an enormous challenge in India, since there are no laws for jaywalking, and people give no attention to zebra crossings, indiacators, blinding headlights. As an extension, sleep-deprived driving aka drowsy driving or fatigued driving is one of the leading causes of road accidents involving human error. Drowsy driving can be defined as driving a vehicle while being cognitively impaires by sleep deprivation. Sleep deprivation can impair the human brain to the magnitudes equivalent to alcohol. A 1998 survey states that 23% of adults have fallen asleep while driving. When a person does not get an adequate amount of sleep, his or her ability to function is affected. As listed below, their coordination is impaired, have longer reaction time, impairs judgment, and memory is impaired. Sleep deprivation has been proven to affect driving ability in four areas:

1. It impairs coordination.
2. It causes longer reaction times.
3. It impairs judgment.
4. It impairs memory and ability to retain information.

Sufficient sleep before driving improves memory. Researchers recorded activity in the hippocampus during learning, and recorded from the same locations during sleep. The results were patterns that occurred during sleep resembled those that occurred during learning, except they were more rapid during sleep. Also, the amount of hippocampal activity during sleep correlated highly with a subsequent improvement in performance. Signs that tell a driver of a need to stop and rest are as follows:

1. Difficulty focusing, frequent blinking, or heavy eyelids.
2. Daydreaming; wandering/disconnected thoughts.
3. Trouble remembering last few miles driven or missing exits and street signs.
4. Yawning repeatedly/rubbing eyes.
5. Trouble keeping head up. Drifting from lane to lane, tailgating, or hitting a shoulder or rumble strip.

6. Feeling restless and irritable.

It has been estimated that approximately 20% of vehicle accidents have sleep deprivation as a cause. Accidents related to sleep deprivation are most likely to happen in the early to midafternoon, and in the very early morning hours. The reason that accidents are mostly likely to happen during the early to mid afternoon may have to do with the biological time clock. Each person's body has its own. Most people run on a daily rhythm of approximately 24 hours, but this can vary from person to person. The reason night time driving is so risky is because sleep becomes an irresistible urge especially from about midnight until 6 a.m. A sleepy period is also ”programmed” for the afternoon which makes that a risky time.

Chapter *2*

Driver Drowsiness Detection

2.1 Existing Algorithms

Existing algorithms for the extraction of facial landmarks and features can be classified as follows:

1. Viola–Jones Object Detection Framework
2. Principal Component Analysis Method (Eigenfaces)
3. Linear Discriminant Analysis Method (Fisherfaces)

2.1.1 Viola–Jones Object Detection Framework

The fundamental principle of the Viola-Jones face detection algorithm is to scan the detector repeatedly through out the same image with a new size each time. Even if an image should contain more than one face it is obvious that a large amount of these sub-windows evaluated would still be non-faces. This realization leads to a different formulation of the problem: The algorithm should discard non-faces instead of finding faces. The idea behind this statement is that dismissing a non-face is much faster than finding a face.

$$h(x) = sgn\left(\sum_{j=1}^{M} \alpha_j h_j(x)\right) \tag{2.1}$$

2.1.2 Principal Component Analysis Method (Eigenfaces)

The so-called Eigenface approach is one of the simplest and most efficient PCA approaches used in extraction of facial landmarks and features in facial recognition systems. This approach transforms faces into a small set of essential characteristics called Eigenfaces which are the main components of the initial set which is a subset of the training set. The process comprises two steps:

1. Initialization
2. Recognition

2.1.2.1 Initialization

1. Acquisition of the initial set of images with face called as a training set.

2. Quantify the Eigenfaces from the training set and retain only the highest of eigenvalues and discard the lowest of eigenvalues.

3. Quantify the distribution in this M-Dimensional space for each person known by projecting their respective face images onto this face-space.

2.1.2.2 Recognition

1. Quantify a set of weights based on the input image by projecting each of the Eigenfaces onto the input image.

2. Determine whether the image is at all a face by checking whether it is close enough to a "free space". Update the Eigenfaces as either known or unknown.

2.1.3 Linear Discriminant Analysis Method (Fisherfaces)

The Fisherface method improves the Eigenface method by using Fishers Linear Discriminant Analysis (FLDA or LDA) to reduce dimensionality. The LDA maximizes the ratio between class dispersion to that of within-class dispersion, and therefore works better than PCA for the purpose of discrimination. The Fisherface method is particularly useful when facial images vary greatly in illumination and facial expression.

The difference between PCA and LDA can be visualised from the image below:

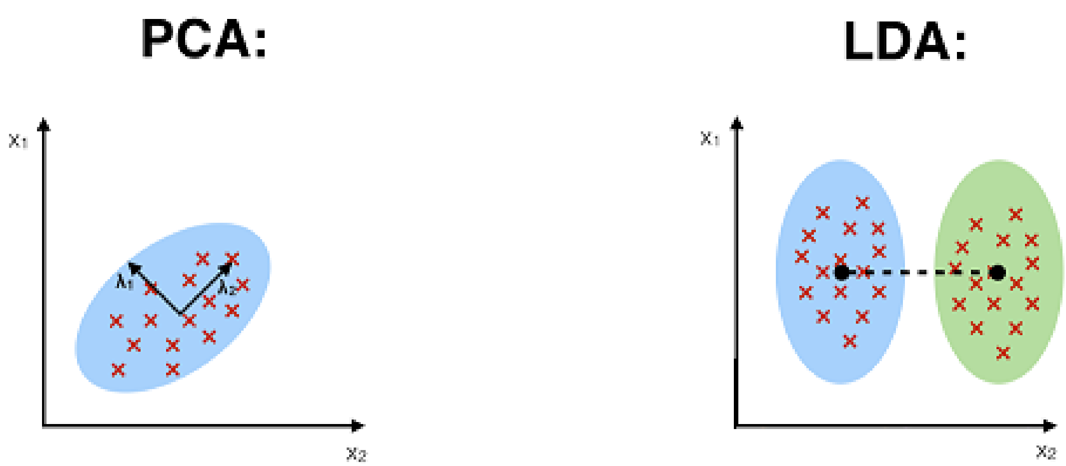


Figure 2.1: PCA vs LDA Comparison.[5]

2.2 Literature Survey

We have seen tremendous progress in the area of image processing in recent decades. Most of them are done for the automatic detection of facial point. There are a number of methodologies proposed that demonstrate great precision and efficiency.

2.2.1 Database

The main reason why many researchers of the field focus on the problem of face alignment is the plethora of publicly available annotated facial databases. These databases can be separated in two major categories:

1. The first category includes databases that are captured under controlled conditions, normally within special indoor laboratories/studios in which the camera position and the lighting source and intensity can be controlled. In most of these databases, each subject is asked to perform a posed facial expression, thus we find more than one images per subject. The most popular such databases are Multi-PIE, FRGC-V2, XM2VTS and AR.

2. The facial databases of the second major category consist of images that are captured under totally unconstrained conditions (in-the-wild). In most cases, these images are downloaded from the web by making face-related queries to various search engines. The most notable databases of this category are LFPW, HELEN, AFW, AFLW and IBUG.[3]

2.2.2 One Millisecond Face Alignment Method

This method presents an algorithm to accurately estimate the position of facial landmarks in a computationally efficient way that uses a regressors cascade.[2]

$$\hat{S}^{t+1} = \hat{S}^{(t)} + r_t(I, \hat{S}^{(t)}) \tag{2.2}$$

This method starts by using:

1. A training set of facial features labeled on an image. These images are labeled manually and specify specific (x, y) coordinates of areas around each facial structure.

2. Priors, in particular, the probability of distance between input pixels' pair.

Given this training data, a set of regression trees is trained to estimate the facial landmark positions directly from the pixel intensities themselves (i.e. there is no " feature extraction "). The end result is a facial landmark detector that can be used in real-time with high-quality predictions.

The indexes for the 68 coordinates can be visualized from the following image:

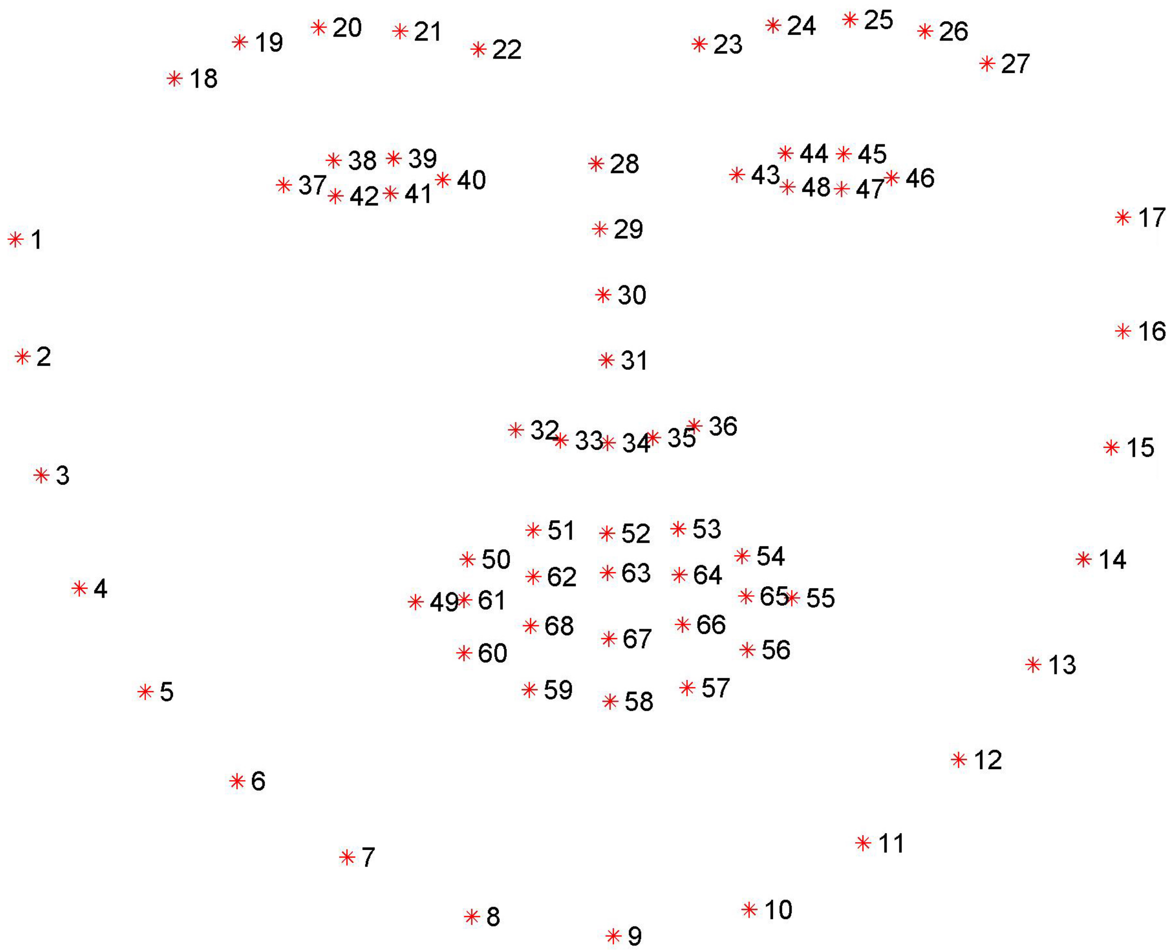


Figure 2.2: Visualizing the 68 facial landmark coordinates from the HELEN dataset.[6]

2.2.3 Eye Tracking and Detection

Detecting eye blinks is important for instance in systems that monitor a human operator vigilance, e.g. driver drowsiness. The blink of an eye is a quick closing and reopening act of the eyelid. Each person has a slightly different blink pattern. The pattern differs in the closing and opening speed, the degree of squeezing of the eye and the length of the blink. The eye blink lasts approximately 100-400 ms. We propose to use state of the art facial detectors to locate the eyes and eyelid contours. We derive the eye aspect ratio (EAR) from the landmarks detected in the image, which is used to estimate the eye opening state. Since the EAR/frame may not necessarily recognize the eye blinks correctly, a classifier that takes into account a larger time window of a frame is trained. The eye aspect ratio (EAR) between height and width of the eye is calculated by the following equation

$$EAR = \frac{||p2 - p6|| + ||p3 - p5||}{2||p1 - p4||} \tag{2.3}$$

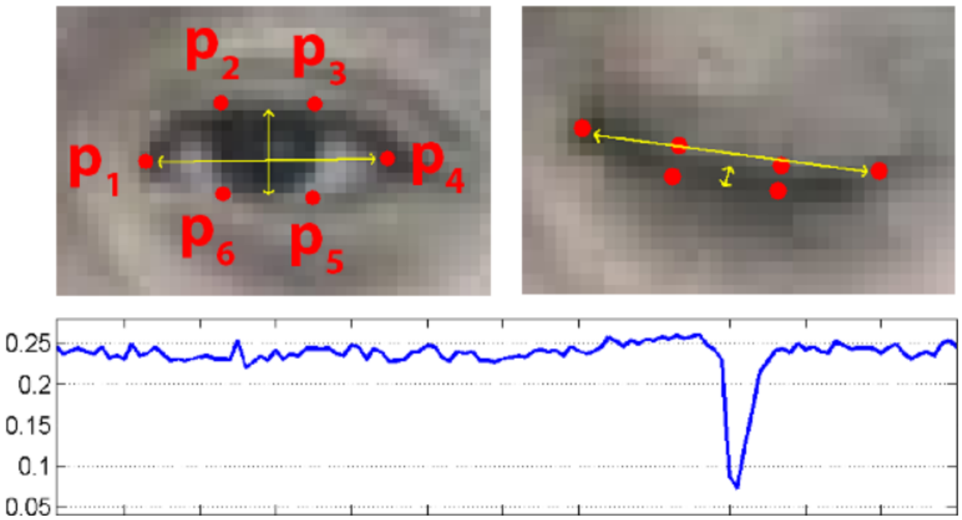


Figure 2.3: Open and closed eyes with landmarks automatically detected. The EAR plotted for several frames of a video sequence. A single blink is present.[7]

The EAR is usually constant when an eye is open and near zero when the eye is closed. It is partially insensitive to the person and the head pose. The aspect ratio of the open eye has a small variance between individuals and is completely invariant to a uniform image scaling and face rotation. Since both eyes blink synchronously, the EAR of both eyes is averaged.

2.2.4 Mouth Tracking and Detection

Similar logic in eye tracking and detection has been followed.

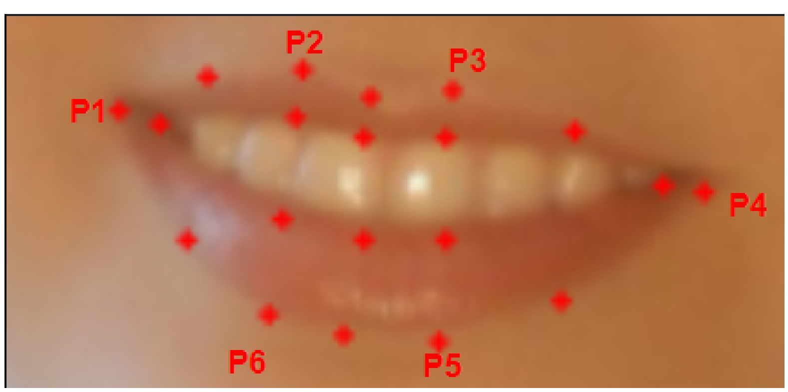


Figure 2.4: Mouth landmarks detected. MAR analogous to EAR is used.

Chapter *3*

Road Detection

3.1 Existing Algorithms

Existing Algorithms for road and scene detection can be classified as follows:

1. Road Detection based on Classification Algorithm
2. Road Detection Using Road Models
3. Segnet

3.1.1 Road Detection based on Classification Algorithm

The fundamental idea behind this method is that it is divided into two parts. In the first part, a binary image is obtained by applying greyscale transformation and thresholding processes. In the second part, K-Nearest Neighbours and Naive Bayes Classifier are being applied on the image. Because of the above two parts, road and non road regions are being determined. A town's satellite image is obtained with the help of Google Maps. From these images, little square pieces are extracted in which some contain road regions and some contain non-road regions. Size of each piece is 15*15 pixels and a total of 120 pieces are collected. Each piece's pixel colour values is divided into three range;[0,85],[86,170] and [171,255]. We obatin nine different mean values from a single piece.

3.1.2 K-Nearest Neighbours Classifier

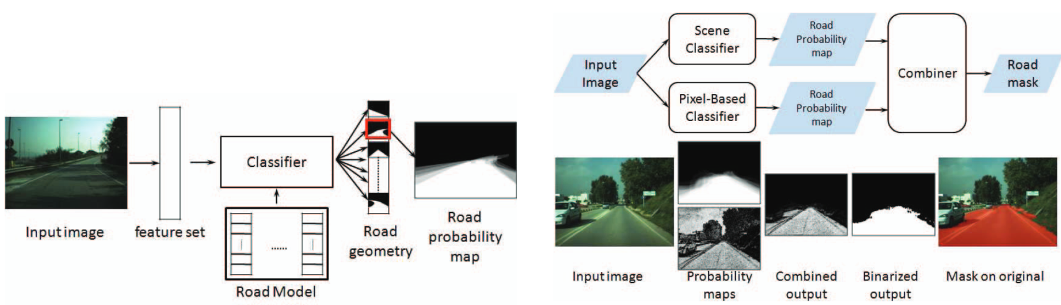
This algorithm is performed by shifting a N*N pixels block on the given satellite image. Shifting process begins at(0,0) point and the shifting distance is 5 pixels.The tow classes are road and non-road. With the help of Euclidean distance the closest seven pieces are taken and if at least 6 of them are roads, then the blocks region is classified as road.

3.1.3 Naive Bayes Classifier

Naive Bayes algorithm is performed in the same way as K-Nearest Neighbors. There is only one difference in class selection phase. We divide nine mean values into three ranges.

3.1.4 Road Detection Using Road Models

This method is based on road classification which provides the probability that an image contains certain type of road geometry. No boundary extraction and previous pixelbased road detection is used here. This allows it to cover more complex road types. When the training of the classifier takes place, a road probability map is learned for each road geometry. Then, once it determines that an input image contains certain type of road geometry, the road probability map is combined with the result of a given pixelbased road detection algorithm to improve the overall performance. Each class represents a different type of road geometry like left/right curves, straight segments, left/right intersections, etc.A set of images of each class is used for learning the codebooks, the class classifiers as well as a road probability map for each class.
Road detection consists in combining the road probability map coming from the scene classification phase with a pixelbased road detection algorithm.



(a) Block Diagram Of The Classifier (b) Combination of Information for Road Detection

Figure 3.1: Combination Of Information For Road Detection [8]

3.1.5 Segnet

This convolutional neural network is used for semantic pixel wise labeling. These generally take an RGB image as an input data, and have a label which is an n channel image, where n is the number of labels involved. Each channel corresponds to a label and each pixel in a certain channel will be 1 or 0 depending on whether that pixel belongs to the label corresponding to that channel.
Each encoder in the CNN applies convolution followed by batch normalization and a non linearity. After this, it applies max pooling on the final result. Decoders work in the similar fashion except that they they dont have a non linearity. When we reach the final decoder, its output is being fed to a softmax classifier which then gives the final prediction.
Typically, the encoder used is a pre trained network like VGG-16 ResNet50.

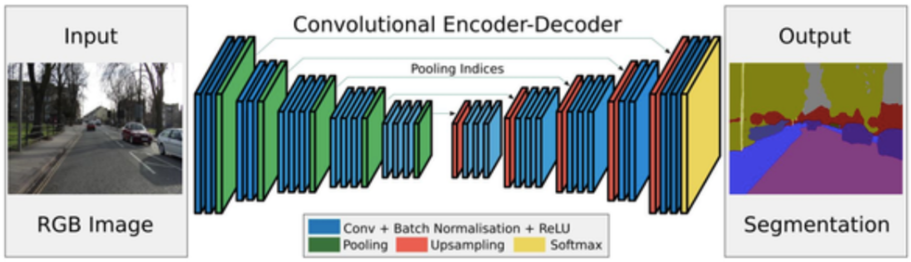


Figure 3.2: SegNet Architecture[1]

3.2 Literature Survey

We have seen tremendous growth in the field of computer vision and image processing. There are a number of methodologies proposed that demonstrate not only great precision but also great accuracy.

3.2.1 Database

The database that we have used is the Cityscape Dataset provided by — . This dataset focuses on the semantic understanding of urban street.It contains a diverse set of images with 5000 images with fine annotations and 20000 images with coarse annotations which is spread over 50 cities. The images were collected over a large timespan in daytime varying over different months(summer, spring, fall). There are overall 20-30 classes in this dataset.

This dataset is intended for:

1. Assessing the performance of vision algorithms.

2. Supporting research that aims to exploit large volumes of annotated data.

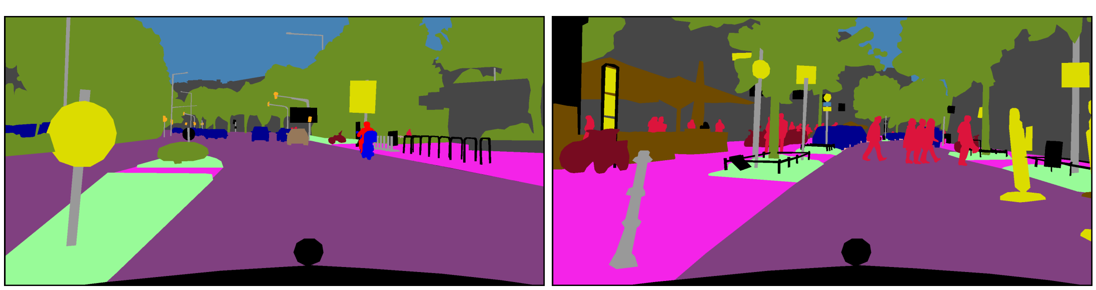


Figure 3.3: Cityscape database training set.

3.2.2 Model

ENet stands for Efficient Neural Network. It is created specially for tasks requiring low latency operations.s It is capable of running on low-power mobile devices while achieving current accuracies. When compared to the existing models, the ENet model is 18x faster, requires 75x less FLOPs(Floating Point Operations per second) and has 79x less parameters.

3.2.2.1 Network Architecture

The entire architecture of ENet is mostly based on ResNet(Residual Network). The structure is divided into one main branch and several other branches that get merged with the main branch via element-wise addition.

Name	Type	Output size
initial		$16 \times 256 \times 256$
bottleneck1.0	downsampling	$64 \times 128 \times 128$
$4\times$ bottleneck1.x		$64 \times 128 \times 128$
bottleneck2.0	downsampling	$128 \times 64 \times 64$
bottleneck2.1		$128 \times 64 \times 64$
bottleneck2.2	dilated 2	$128 \times 64 \times 64$
bottleneck2.3	asymmetric 5	$128 \times 64 \times 64$
bottleneck2.4	dilated 4	$128 \times 64 \times 64$
bottleneck2.5		$128 \times 64 \times 64$
bottleneck2.6	dilated 8	$128 \times 64 \times 64$
bottleneck2.7	asymmetric 5	$128 \times 64 \times 64$
bottleneck2.8	dilated 16	$128 \times 64 \times 64$
Repeat section 2, without bottleneck2.0		
bottleneck4.0	upsampling	$64 \times 128 \times 128$
bottleneck4.1		$64 \times 128 \times 128$
bottleneck4.2		$64 \times 128 \times 128$
bottleneck5.0	upsampling	$16 \times 256 \times 256$
bottleneck5.1		$16 \times 256 \times 256$
fullconv		$C \times 512 \times 512$

Figure 3.4: ENet Architecture[4]

3.2.2.2 Initial Block

The input image taken will have a resolution of 512x512. The input is then passed through a convolutional layer of (13 filters) and separately passed through Max pooling(2x2; without overloop). The result is then concatenated.

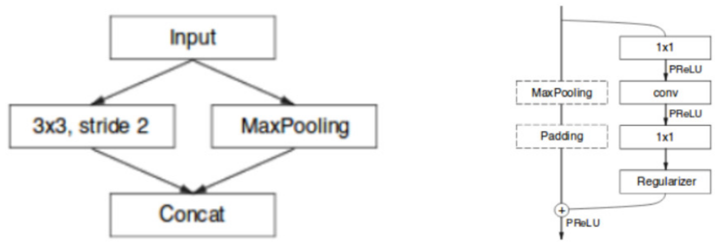


Figure 3.5: Block diagram of Intital Block and Bottleneck module[4]

3.2.2.3 Bottleneck modules

The bottleneck modules is term used for denoting branches. All the bottlenecks have the same structure.Each bottleneck consists of three convolutional layers. The first layer (1 x 1) projection reduces the dimensionality while the second layer(1 x 1) projection restores the dimensionality. In between these convolutions, a regular, dilated or full convolution is being done. Batch normalization and PReLU(Parametric ReLUs) are placed between all these convolutions. MaxPooling id done on the main branch . The first projection in the branch is replaced by a non-overlapping 2 x 2 convolution and the activations get zero padded to equal the number of feature maps.

In the decoder MaxPooling is replaced by MaxUnpooling. Obviously, the activations are not padded anymore but instead spatial convolution without bias is performed.

Chapter *4*

Simulation and Results

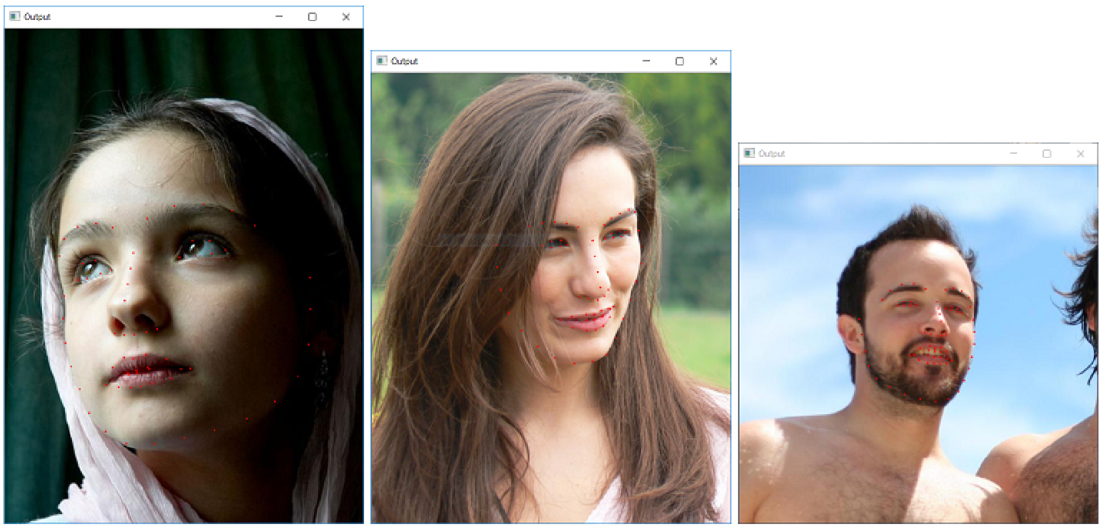
4.1 Eye And Mouth Detection And Tracking

4.1.1 Sample Images

(a) Sample Image 1 (b) Sample Image 2 (c) Sample Image 3

Figure 4.1: Sample Images

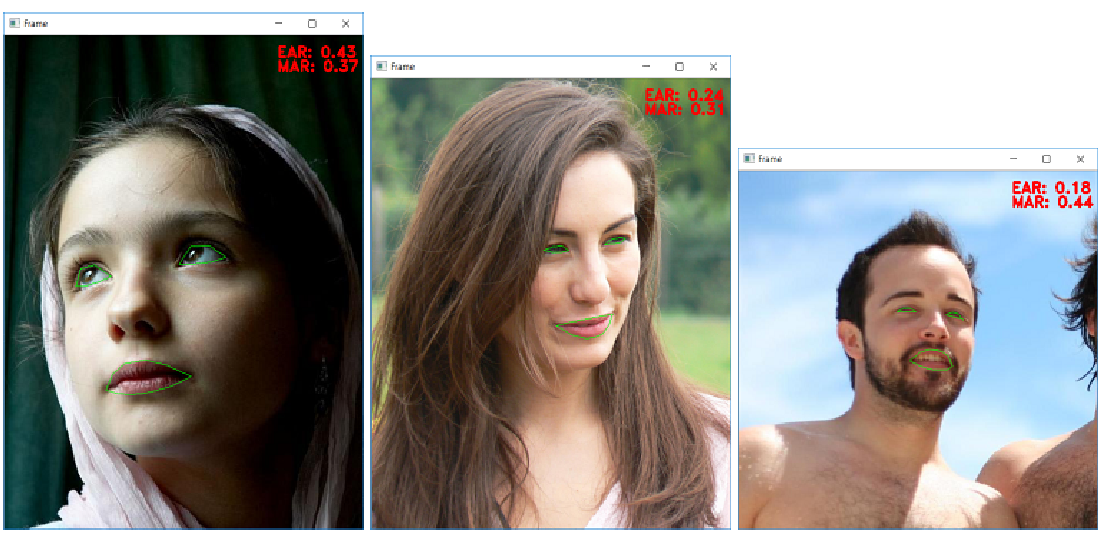
4.1.2 One Millisecond Face Alignment Method



(a) Sample Image 1 (b) Sample Image 2 (c) Sample Image 3

Figure 4.2: Algorithm Applied On Sample Images

4.1.3 Facial Contours Detected Using Our Code



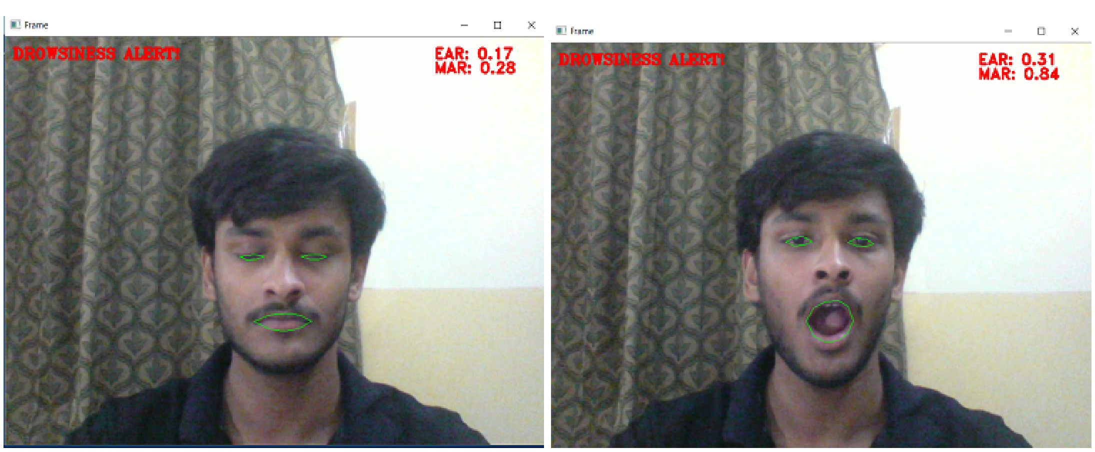
(a) Sample Image 1 (b) Sample Image 2 (c) Sample Image 3

Figure 4.3: Contours Detected On Sample Images

4.1.4 End Results

Here are the following results that we obtained:

Figure 4.4: Normal face and eye detection.



(a) Tired eyes detected.

(b) Yawn detected

Figure 4.5: Drowsiness detection.

4.2 Road And Object Detection And Tracking

4.2.1 Sample Images

(a) Sample Image 1

(b) Sample Image 2

(c) Sample Image 3

Figure 4.6: Sample Images

4.2.2 Semantic Segmentation Using ENET Method

(a) Sample Image 1

(b) Sample Image 2

(c) Sample Image 3

Figure 4.7: Algorithm Applied On Sample Images

Chapter 5

Conclusions and Future Work

5.1 Conclusion

We explored the problems faced by people due to driver drowsiness and chose to step in. We observed different databases and chose the most suitable one for our aim, which came out to be Helen dataset. We trained the database with variant alignments to the camera, using One Millisecond Face Alignment Method. Face detection was used to detect faces from our live video stream, frame by frame. On the event of detecting a face, eye and mouth were detected separately and then were passed on for drowsiness check. For checking a face for drowsiness, aspect ratios of eyes and face were measured respectively, and upon the aspect ratios (EAR, MAR) crossing a pre-calculated threshold value, driver was alerted.

Computer vision using image processing is a near-to-perfect way to use images in form of data and working out algorithms accordingly, but it comes with its own set of limitations. Camera quality should be good, which comes with a slightly high price. Also, if the system is to be used in night, then the camera must have night light, adding to the budget. The hardware where we dump should be really fast for real-time detection, as time is of essence in such cases, where fatal accidents are a possibility. One major problem that doesn't seem to be able to be rectified is if the driver is using sunglasses. In such a case, our system is unable to detect and driver's eyes and is rendered useless. Despite the above mentioned limitations, our system has proved time and again to produce desirable results, thus is ready to be used by anyone who plans to launch our executable file on a suitable hardware with an appropriate camera, and can be used to help the problem of driver drowsiness and resultant accidents to quite an extent. Our system can and should be used to reduce the number of road accidents till at least self driving cars on the roads of India become a reality, and perhaps even after that with certain adjustments and rigorous betterment of the same.

5.2 Proposed work for next Semester

In this semester, we have completed the driver detection part of our project. In the upcoming semester we will be working towards the detection of the roads in real time. For that, we will be taking the help of the deep learning model CNN through which we will train our machine to identify roads. Since our project is mainly focussed on the safety of the driver, we will also do car and pedestrian detection so that the driver may know if any object is coming close to his car or not. Also, we will be adding a distance sensor to our working model so that the drive could be notified whether any object is coming dangerously close to his car or not and the driver could avoid it.

Bibliography

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. In *IEEE*, 2017.

[2] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *CVPR*, 2014.

[3] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. Interactive facial feature localization. In *UIUC*, 2012.

[4] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. In *CoRR*, 2016.

[5] S. Raschka. Articles of python lda. In *Linear Discriminant Analysis.*, 2014.

[6] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: Database and results. In *Image and Vision Computing (IMAVIS), Special Issue on Facial Landmark Localisation "In-The-Wild".*, 2016.

[7] T. Soukupova and J. Cech. Real-time eye blink detection using facial landmarks. In *Computer Vision Winter Workshop*, 2014.

[8] J. M. lvarez, T. Gevers, and A. M. Lpez. Vision-based road detection using road models. In *IEEE*, 2009.