

A REAL-TIME PEDESTRIAN DETECTION SYSTEM FOR SAFETY APPLICATIONS

by

Rinju Alice John

Bachelor of Technology,

Kerala, India, July 2013

A project

presented to Ryerson University

in partial fulfillment of

the requirements for the degree of

Master of Engineering

in the Program of

Electrical and Computer Engineering

Toronto, Ontario, Canada, 2016

© Rinju Alice John, 2016

AUTHORS'S DECLARATION

I hereby declare that I am the sole author of this project. This is a true copy of the project, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this project to other institutions or individuals for the purpose of scholarly research

I further authorize Ryerson University to reproduce this project by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my project may be made electronically available to the public.

Ryerson University

Title: A real-time pedestrian detection system for safety applications

Degree: Master of Engineering

Year of convocation: 2016

Name: Rinju Alice John

Program: Electrical and Computer Engineering

ABSTRACT

Nowadays, People are more distracted by their vulnerable devices, whenever they enter a cross road. As a result, a fatal accident or injury will occur. This motivated the need to implement a reliable pedestrian detection system. To optimize the system, a cross road scenario is considered where the driver is taking a right turn and a smart camera is used to capture consecutive pictures of the pedestrian. The consecutive frames are studied using Region Of Interest method and the Gaussian mixture model method. Once the detected pedestrian enters region of interest in less than 2 meters, a warning and automatic brake system is initiated to prevent the accident. Finally, the results of the proposed methods are compared based on the processing speed and performance rate of the Shape based detection technique (Wei Zhang, [12]). The performance rate was above 90% and processing speed was about 1 sec for the proposed methods.

ACKNOWLEDGEMENTS

I take this opportunity to express my sincere gratitude to all those who have been instrumental in the successful completion of the project design. First of all I thank Professor Kaamran Raahemifar, who is my supervisor for his help in the selection of the project topic and for the guidance and suggestions in the design of the project.

I am also thankful to the Electrical and Computer Engineering department faculties and to the management of the Ryerson University for providing all the facilities for completing this work. Lastly, above all I thank God Almighty without whose blessings this effort would not have been a reality.

Dedicated to my friends and family

TABLE OF CONTENTS

Author's Declaration	ii
Abstract	iii
Acknowledgements	iv
Dedication	v
Table of Contents	vi
List of Figures	viii
List of Tables	ix
1. Introduction	1
1.1. Motivation	1
1.2. Survey on pedestrian and cycling safety in Toronto	1
1.3. Project objective	2
1.4. Scope of project	3
2. Review on recent technology and its related work	4
2.1. Related work	5
2.2. Recent technology	11
2.2.1. Types of sensors used for detection purposes	11
2.2.2. Common detection approaches	12
2.2.2.1. Stereo vision based detection	12
2.2.2.2. Shape based classification	14
2.2.2.3. Feature extraction method	15
2.2.2.4. Radar and Lidar sensors	17
3. Theory	20
3.1. Image acquisition and frame classification	20
3.2. RGB color model	22
3.3. Image segmentation	24
3.3.1. Thresholding	25
3.3.2. Segmenting RGB values	26
3.3.3. Background subtraction and estimation	26
3.4. Image enhancement	27
3.4.1. Edge detection	27

3.4.2. Blob detection and noise filtering	28
3.4.3. Histogram processing	29
3.5. Image extraction- morphological image processing	30
4. System implementation and Results	32
4.1. System overview	32
4.1.1. Device specifications	32
4.1.2. Software specifications	33
4.1.3. Device coordination	33
4.2. System design	33
4.2.1. Assumptions and flow diagram	33
4.2.2. Proposed design	35
4.3. Analysis of system design for consecutive frame based on ROI method	36
4.3.1. Image acquisition and frame generation	36
4.3.2. RGB segmentation	36
4.3.3. Separating background and foreground	37
4.3.3.1. Color based threshold detection technique	37
4.3.4. Calculation of blueness value	37
4.3.5. Pedestrian detection based on ROI generation	39
4.4. Analysis of system design for consecutive frame using Gaussian model	41
4.4.1. Frame classification	42
4.4.2. Foreground and background separation	42
4.4.3. Pedestrian detection based on Gaussian mixture method	42
4.5. Discussions	43
5. Conclusion and Future work	45
5.1. Conclusion	45
5.2. Future work	46
Reference	47

LIST OF TABLES

2.1: Comparison of current pedestrian results	9
2.2: Types of sensors	11
2.3: Comparison of different techniques discussed	19
4.1: Camera parameter	32
4.2: Computer hardware specifications	32
4.3: Matlab specification	33
4.4: Calculation of blueness for different color combinations	39
4.5: Comparison for both the proposed methods	44

LIST OF FIGURES

1.1: Bar graph representation of traffic fatalities by categories	1
1.2: Pedestrian fatalities across the GTA	1
2.1: Reaction time for a given distance decreases with vehicle speed	4
2.2: Overview of PROTECTOR module	13
2.3: Shape-based detection[12]	14
2.4: Shape based detection masking-Stereo verification [22]	15
2.5: Hierarchical shape detection [22]	15
2.6: Block diagram for feature extraction	16
2.7: Illustration of HOG	17
2.8: Detection based on Probabilistic approach	17
2.9: Coverage for the sensors	18
3.1: Image capturing in a pin-hole camera (Wikipedia)	20
3.2: Coordinates in pinhole camera	21
3.3: Successive frames used for background estimation [27]	22
3.4: Gray image storage representation [12]	23
3.5: RGB image storage representation	24
3.6: Thresholding at $t=128$	25
3.7: Neighborhood structure for edge detection [28]	28
3.8: Effective and non-effective noise removal	29
3.9: Histogram distribution for red pixels	30
4.1: Flow chart for pedestrian detection	34
4.2: Three stages in the proposed method	36
4.3: Histogram for RGB models	37
4.4. Colorbar for RGB plane	37
4.5: Thresholding for separating foreground and background	40
4.6: Pedestrian detected using ROI method	41
4.7: Pedestrian frame 204 at top elevation	42
4.8: Foreground separation of the detected pedestrian	43

CHAPTER 1.

INTRODUCTION

1.1. MOTIVATION

The traffic situations have undergone dramatic changes over the last few years and this lead to the development of an innovative technology to detect pedestrians or other vulnerable road users [1]. The motivation for this research is based on the survey report for the past five years from 2009 to 2014. However, 2013 has seen the highest number of road accidents. This paper is focused on designing a method for pedestrian detection which in turn can reduce the number of road accidents. Even though, many researches were being carried out for pedestrian detection, most of the methods had drawbacks on false detection (FP). False detection means objects that are incorrectly detected as pedestrians and this research paper has overcome some of the challenges on false detection [1].

1.2. SURVEY ON PEDESTRIAN AND CYCLING SAFETY IN TORONTO

In earlier times, walking and cycling were the active means of transportation for human beings and this had a number of health benefits for their day-to-day life [2]. This survey is done to ensure safety while walking and cycling. The death rate of pedestrians and cyclist as a result of collision with vehicles is the highlighted part of the survey. The collision and injury/fatality data were obtained from the 'Toronto Police Services' collision reports for the period 2008-2012 and this electronic database contains information on all police-reported vehicle collisions that occurred in the City [2]. For every collision it lists the date, time, location, age of the person involved, and the severity of any injuries that resulted.

From the bar graph (Figure 1.1) for the past four years, especially from 2009 to 2013, it is seen that the pedestrian and cycling death rate was less than 50-40 percent in the year 2011. But Toronto city had reduced its number of traffic fatalities significantly, and it seems stabilized.

During the first six months of 2013, Toronto city ties with 2011 record which contained 36 pedestrian and cyclist fatalities. However, the last six months of 2013 was worse and it contained 45 traffic fatalities. Therefore, the year 2013 had a total of 43 percent increase over the year 2012. The death rate of pedestrians shot up to 67 percent.

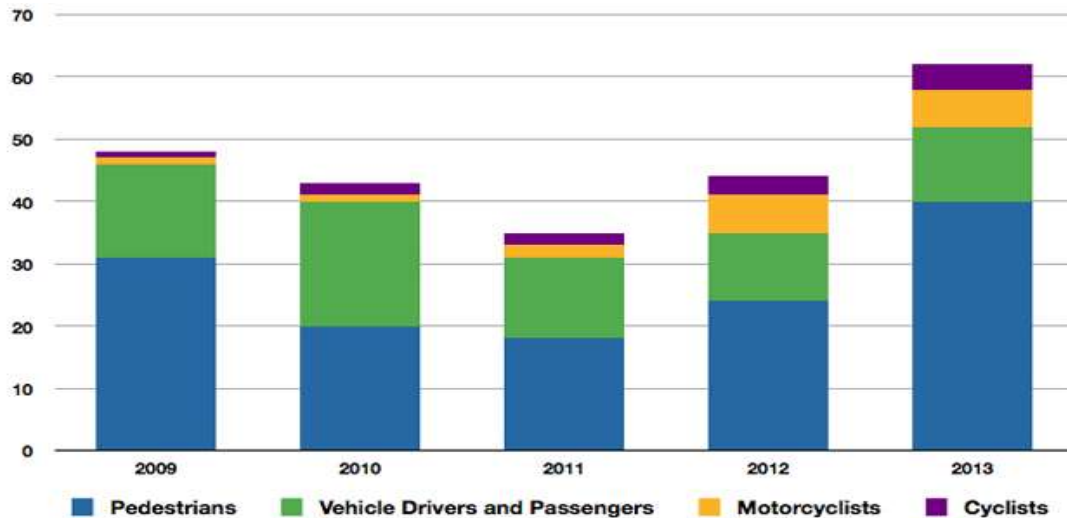


Fig. 1.1: Bar graph representation of traffic fatalities by categories [3]

It's clear from the graph (Figure 1.2) above that the Toronto city faced most of the pedestrian accidents compared with the neighboring cities.



Fig. 1.2: Pedestrian fatalities across the GTA [3]

1.3. PROJECT OBJECTIVE

The objective of the report is to detect real time pedestrian using smart phone camera attached to the vehicle while the developed software can detect the pedestrian on the roadside when the vehicle tries to turn right. The method used to detect pedestrian is very simple and thus prevent the accident occurring across road side. This research scenario suggests a typical situation when drivers want to take a right turn at an intersection and have to yield to oncoming cars on the main roads. However, drivers fail to notice pedestrian on their right hand because their attention is focused primarily on the road on their left-hand side which can create a potential hazard for pedestrians.

1.4. SCOPE OF PROJECT

This project is mainly motivated based on the increased need to protect pedestrians from road accidents and the project is discussed in four sections. Section 2 of the project focuses on the literature survey of the related technologies used for pedestrian detection. Later, a comparison is performed for recent technologies based on their processing speed and performance rate. However, Sections 3 and 4 discuss the theory behind the implementation and further discuss the system requirements, flow diagram and proposed techniques used for detecting pedestrians. The technical approach and results are discussed in section 4. Finally, section 5 concludes the project by comparing the results with the shape based detection technique and also incorporates suggestions for future work.

CHAPTER 2.

REVIEW ON RECENT TECHNOLOGY AND ITS RELATED WORK

Pedestrian safety methods are improving through different ways and these can be summarized as follows: Infrastructure improvement (vehicle design), situational awareness (throughout driving), cautionary signals, emergency alarm, auto braking and deploy collision mitigation [5]. To make it precise the infrastructure can be categorized in to speed control of the vehicle, pedestrian vehicle separation and the visibility of pedestrian by the driver [6]. Speed reduction results in less pedestrian injuries due to having a greater reaction time (Figure 2.1).

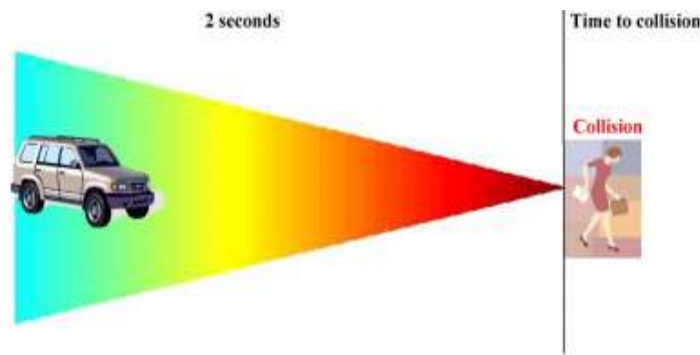


Fig. 2.1: Reaction time for a given distance decreases with vehicle speed[5]

Some popular speed control methods include speed bumps and using multi-way stop signals [6]. Separation of pedestrians and vehicle can be done by installing traffic lights and allocating time for pedestrian signals. Pedestrian visibility can be increased by improving the road lighting since most pedestrian injuries occur at night. Most accidents occur because of the visibility problem, such cases can be improved by using radar waves which can be attached to the bumper side on the vehicles. This report deals with detecting pedestrian for vehicles turning towards the right side of the intersection and the pedestrian can be detected using cameras attached to the driver side.

2.1. RELATED WORK

The related work for pedestrian detection mainly begins with the different sensors used. The most commonly used detection techniques are active and passive sensors. Passive sensors gather target data through recognition of the radiation, light, heat and vibrations occurring in the subject's domain. On the other hand, the active sensors send off signals from transmitters and data is collected from the sensors upon their reflection with the help of electrons or light waves [7]. Both sensors are classified according to the application techniques used for detection. For example, in cases where active sensors like lidar and radar cannot be used for detection, passive sensors like CCD cameras are used. Therefore, both sensors have got positive and negative feedbacks.

Steffen Heuel and Hermann Rohling developed a classification algorithm for automotive application using radar sensors (at 24GHz), which can be used for measuring velocity and distance with a band-width of 150MHz [8]. The paper proposed two systems such as single radar system, that measures the transmitted signal using a single MFSK (Multi-Frequency Shift Keying) at 39 ms. It is used to measure the radial speed, range and the level of signal [8]. In the second system named as, multiple radar system wherein measurements are modified using different speed and range of the signal. Eventually, a tracking method is used to extract features of the pedestrian. The pros of this method is that measurements were very accurate and the con is that large bandwidths are used to classify pedestrians in the urban areas.

Gavrila and Munder, [9] proposed PROTECTOR system (a real-time stereo system for pedestrian detection and tracking). The highlights of the method is that it used a texture based classification and the method used fixed cameras at 25 meters apart. The pictures are stored in frames, which resulted in 71 percent pedestrian detection and 0.1 false alarms/frame [9]. The pros of the system is that using sparse stereo the reliability increased and had higher sensor coverage. However, the con is that the datasets used were from comparatively sparse environment.

Giuseppe De Nicolao *et.al.* [10] analyzed the risk assessment of pedestrians based on certain scenario. The scenario assessed is based on PROTECTOR system (proposed by Gavrila and

Munder, [9]). As the pedestrian crosses the road the accident risk rises to 85%. The proposed solutions are based on the sensors used at different radial angles.

Cristiano Premebida *et.al.* [11] used 3D laser sensors commonly known as LIDAR in the detection and evaluation for depth perception of road crossing pedestrians [17]. The data sets are collected and evaluated by using a simple low pass up-sampling algorithm. There has been 3741 frames used for the detection on which 52% were detected successfully. The con is that the pedestrian detection algorithm resulted with less accurate values.

Wei Zhang, [12] proposed a method for pedestrian detection using stereo vision technique. The system used for pedestrian detection includes stereo vision cameras, accelerator and dual quad-core computers [12]. A layered approach is also implemented to differentiate between stationary and dynamic objects. The method processed at a frame rate between 7 to 10 Hz rates and can detect pedestrian at 30 m apart with a driving speedup to 15 - 35 mph. The results showed up 90% positive detection rate. The infrastructure used for the system is simple. The paper lacks in the image contrast and low visibility at certain lightning conditions.

Aline Senart *et.al* [13] proposed a method where a fixed sensor network is used for detecting the pedestrians crossing the road. The pedestrian detection is done by using some reflective jacket which needs to be worn by the pedestrian and these jackets can send radio waves to the sensors attached to the roadside units. These roadside units are controlled by the base station units which in turn give warning signals to the on-board computer units. Thus, this method provides continuous information about the pedestrians in VMS (variable-message sign) format even though the road conditions are beyond their limit. This method is very simple and cheap and moreover, better results are obtained with horizontal antennas that are directed towards the road. The pros of this method is that it is very promising and the pedestrian detection can be performed beyond the driver's limit. The cons are that there can occur messages congestion which can lead security issues with the messages.

M. Bertozzi *et.al.* [14], developed a system that is aimed to detect the pedestrian and to classify other objects in the surrounding by using a pedestrian detection system, such as a camera, a radar and an inertial sensor [14]. In the proposed method, three systems are used and each of them has

got different purposes. The camera can produce improved detection of the vehicle boundaries which are detected using the radar sensors. The false positives are discarded and the pedestrian is being detected at the same time by motion stereo techniques. The radar with a frequency of 24GHz is being kept at the bumper of the vehicle. The pros is that the use of the radar on the vehicle helped to determine the position and size of the human being. The cons is that the single camera which is used to detect the pedestrian is clustered with a noisy background so it needs further filtering.

Erwin Thomas Gilmore *et.al.* [15], used a low cost microcontroller system and an infrared radar system for detecting the pedestrian. This method is proposed for real time pedestrian detection and the system developed basically focuses on digital and infrared imagery. All the applications are performed on a microcontroller which is being programmed on matlab. The communication part is performed by an indicator which helps in decision making for the driver's assistance and to provide information regarding security. The pro is that the platform developed using microcontroller were successful for both on-line and off-line data processing [15]. Moreover, the method used is very unique and the design was really very successful. However, the con is that the method doesn't show any results for the complex images and this happens because the algorithm has more dependence on the aspect ratio of the image.

A. Bartsch *et.al.* [16] proposed a classification method based on radar sensors for pedestrian tracking. The detection is based on extracting data from Doppler and spatial distribution method. In this experiment, a 77GHz radar sensor is used to track the pedestrian and 95% output is achieved by the proposed classification method. The pro of the paper is that the radar sensor used with higher Doppler frequency is efficient for detection. The con is that there are possibilities for high rates of false detection.

Yang Yang *et.al.* [17] proposed a real time pedestrian detection from a moving vehicle. The proposed method was known as BHGD (Boosted Histogram of Gradient Direction) [17]. In the first detection, pixels of the window is being classified into different blocks and HOG is computed for each block. The values obtained from the HOG computation is then evaluated using BHGD. Secondly, composite classifier is built using a Adaboost algorithm [17] for selective blocks. This Adaboost algorithm is the unique feature apart from the HOG method. The

pros is that the performance is much improved compared to the HOG technique. This method has high accuracy and the detection is not affected by illumination compared to HOG technique. The con is that the size of the image used to detect is being reduced, so it can be used only for on-board pedestrian detection systems.

Javier Hernández-Aceituno *et.al.* [18] proposed a Bayesian approach in the Viola-Jones algorithm, for the pedestrian detection. This method is a statistical approach for the original algorithm [18]. The objective of this approach is to provide accuracy for the frames and to detect the pedestrian accordingly. The pros of this method is that it is more accurate for clear pedestrian images. However, the cons is that the complexity is high and the method can be affected by illumination.

Xiaofeng Chen *et.al.* [19], designed a detection method using three dimension sensing based on a Microsoft's Kinect. The method is easier and it can extract the RGB images by background separation method. Later, a region clustering method is used to detect the pedestrian. Moreover, counting and tracking method is used to design the pedestrian volume. The pros is that the occlusion is lesser compared to other proposed methods and it has got reliable detection range. The con is the lack of accuracy in pedestrian detection.

Table 2.1: Comparison of current pedestrian results

Author	Sensor type used	Test data sets	Area covered	Processing speed	Performance per frame
Steffen Heuel and Hermann Rohling, [8]	Radar sensors (24GHz)	8000 data samples includes pedestrian , non-pedestrian	-	39ms, 24GHz radar sensors	71.32%true positive for vehicle 45.20% true positive for pedestrian
Gavrila and Munder, [9]	Stereo cameras	1 hour travel to suburban - video sequences	0-25m	100ms	performance-61.0%
Giuseppe De Nicolao <i>et.al.</i> [10]	Digital camera and radar sensor	Data recorded by a on-board sensor system	10-26m	-	77.80% correct
Cristiano Premebida <i>et.al.</i> [11]	LIDAR sensors	7481 training images, 7518 testing images	-	-	Benchmark: easy-59.51 % ,Moderate-46.67 % , Hard-42.05 %
Wei Zhang, [12]	Stereo vision 3D sensors	Video Recording 100 images	0-40 m	Frame rate: 7.5 -10Hz	Detection rate is 93.75%, 6.25% false positive/frame
Aline Senart <i>et.al.</i> [13]	Wireless sensor network	Video recording	5m	500ms	More than 95%
M. Bertozzi <i>et.al.</i> [14]	Stereo cameras and radar sensors (24GHz)	Video recording	-	83 ms or 7Hz rate	Above 90% detection rate
Erwin Thomas Gilmore <i>et.al.</i> [15]	Infrared sensors	Recording on urban environment 90 images	-	-	Correct detection rate is 90% and false alarm rate is 10%

A. Bartsch <i>et.al.</i> [16]	Radar sensors	150 data frames	Maximum distance: 50m	76-77GHz	Detection rate is 29.4% false positive /frame is 9.5%
Yang Yang <i>et.al.</i> [17]	Stereo vision	10000 images from urban areas, 208 cropped images 69 video sequences	Maximum distance: 60m	-	85% pedestrian detection rate
Javier Hernández- Aceituno <i>et.al.</i> [18]	Camera images	Urban and rural images	-	-	71.2%derection rate
Xiaofeng Chen <i>et.al.</i> [19]	Microsoft Kinect- consists of video camera, four microphones, IR projector, monochrome camera	20 frames per second		50ms	93.1% positive detection rate

2.2. RECENT TECHNOLOGY

2.2.1. Types of sensors used for detection purposes

Pedestrian detection can be mainly classified into two main approaches such as vision-based approach and radar-sensor based technique. In the proposed paper a fixed camera system is used to detect the pedestrian. Usually the pedestrian detection is based on static and dynamic detection technique. For the detection technique, the motion details of the pedestrian is being used. The former technique is used in vehicles and the later approach is used for cases where cameras are used for detecting pedestrian [20]. The comparison between different sensors is being noted on the table below.

Table 2.2: Types of sensors

Sl.no	Sensors used	Function	Pros	Cons	Reference
1.	Passive sensors: Radar	Distance measurement along with pedestrian detection	Able to measure distance and speed	Not an efficient classifier and they are highly false positive due to different noises.	[8], 2010 [11], 2014
2.	Passive sensors: LIDAR	Distance measurement and environment modeling	i. Able to measure distance and speed. ii. smaller wavelength signals	i. difficult to popularize. ii. Not an efficient classifier	[11],2014
3.	Active sensors: optical camera	Detect pedestrian by feature extraction	i. Large detection capability and wider area for view. ii. Able to obtain specific information	i. Can be affected by features such as weather, illumination, etc. ii. Not accurate distance information	[17], [14]
4.	Visible light sensors	Detect pedestrian by feature	i. obtain details of the road scene	i. cannot obtain the road scene during night	[13],2008

		extraction		ii. cost is higher	
5.	Infrared radiation	Detect pedestrian by feature extraction	i. not affected by illumination ii. suitable for both day and night time. iii. not affected under high temperature condition.	i. cost is higher	[13],2008

2.2.2. Common Detection approaches

Generally, pedestrian detection and classification are based on certain characteristics such as field of view, range for detection, angular resolution, illumination and hardware cost [21]. Some of the key technologies used for detection of pedestrian is listed below.

2.2.2.1. Stereo vision based detection

This method of detection is seen in PROTECTOR system, one of the pedestrian detection technique using vision-based approach. In this method of detection, four modules are used in a multi-cue fashion such as stereo-based generation, texture detection, shape detection and stereo based classification [9]. Stereo vision based detection is a method of triangulation that finds the range between two images taken from two different places. This is possible with the help of a pair of cameras with a known separation distance. In this implementation technique, the researchers kept the cameras in parallel axes and this setup is done in order to find the similarities in the two images captured by the cameras. This similarity can be performed by image correlation and by peak calculation. Image correlation method is a local correlation-based [1] method that produces consistency of the images, which can later be converted to a range map. The correlation function is implemented as the SAD (Sum of Absolute Differences) [20]. The equation of SAD for each pixel given in the image when compared over 7 x 7 pixel regions is shown in equation 1 as [1] :

$$SAD(x, y, z) = \sum_{i=-3, j=-3}^{i=3, j=3} (|A(x + i, y + j) - B(x + i - s, y + j)|) \quad \text{equ. 1}$$

where,

A = Left image of the stereo pair.

B = Right image of the stereo pair.

x and y = Image pixel locations.

s = Number of horizontal shifts that are searched to find an image correlation.

The Figure 2.2 shows an overview of PROTECTOR module and the grey shaded portion depends on the stereo imaging. Stereo images contain color and depth information of each pixel and can easily apply any filtering algorithm to reduce the regions of interest [12]. If the horizontal coordinates of the pixels, x_l and x_r , as left and right images then the range, Z can be calculated as [1]:

$$Z = \frac{b * f}{d} \quad \text{equ. 2}$$

where,

b = the stereo camera baseline.

f = the focal length of the camera (pixels).

d = the image disparity value.

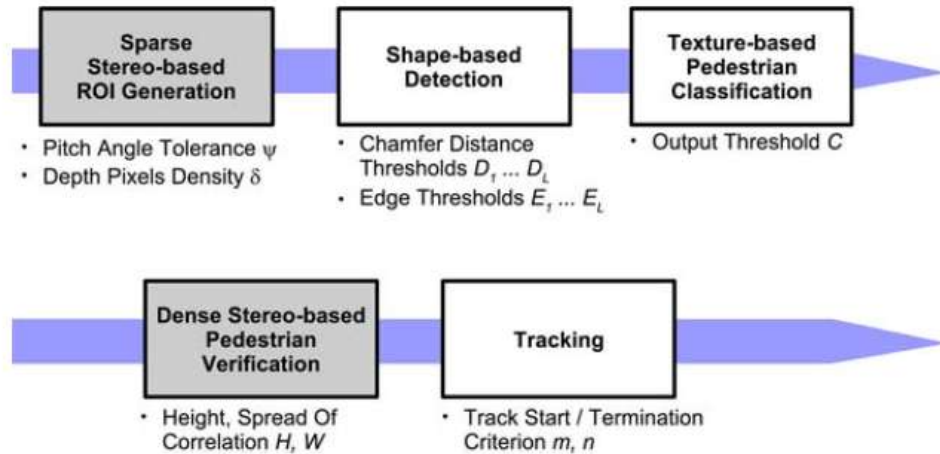


Fig. 2.2.: Overview of PROTECTOR module

2.2.2.2. Shape based classification

The shape based classification system is carried out based on model dataset which can be matched with a large set of pedestrian models. The process used in this type of detection is that the each frame of data is compared to the shape of the human being. To get better results a threshold value is being kept and if the input image and model shape based images are larger than the threshold value used then the pedestrian can be detected [20]. The shapes used for the shape matching is shown in Figure 2.3 below [12].



Fig. 2.3: Shape-based detection [12]

In these cases, a correlation approach is being used to extract missing features which can happen due to segmentation problem. These are named as occlusion or gaps at the edges and it allows the shape to fit to the correct pedestrian solution [22]. In most cases of the PROTECTOR system (Figures 2.4. and 2.5), the pedestrian templates are being matched with the distance transformed images [23]. The method can be locked into the desired object patterns and the detection can be made easier by selecting the threshold value. Therefore, by using the manually specified shape labels a shape hierarchical is considered for detecting the pedestrian automatically. Once the shape is matched the required ROI (Region Of Interest) is selected and the pedestrian is separated from the background.

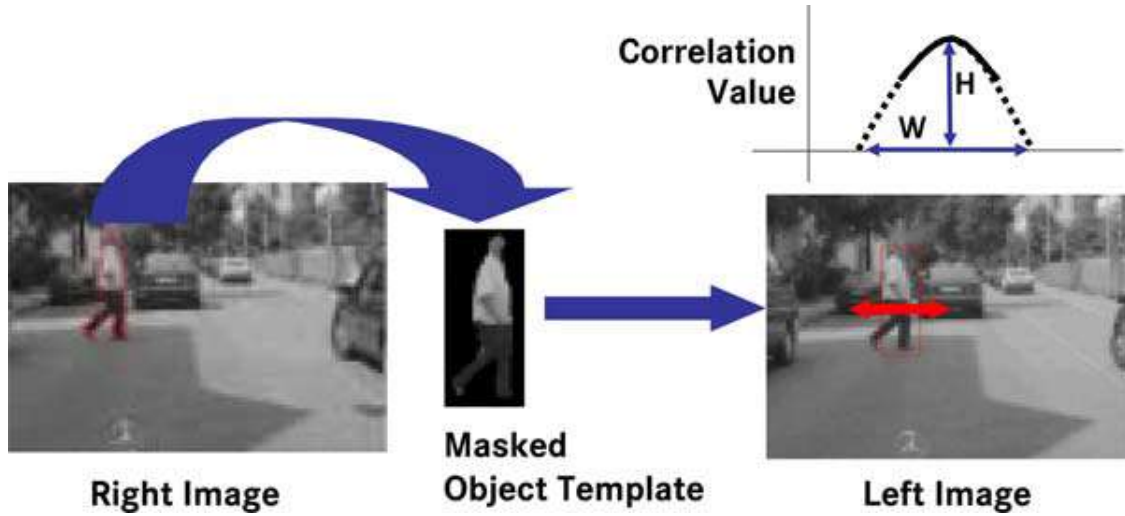


Fig. 2.4: Shape based detection masking-Stereo verification [23]

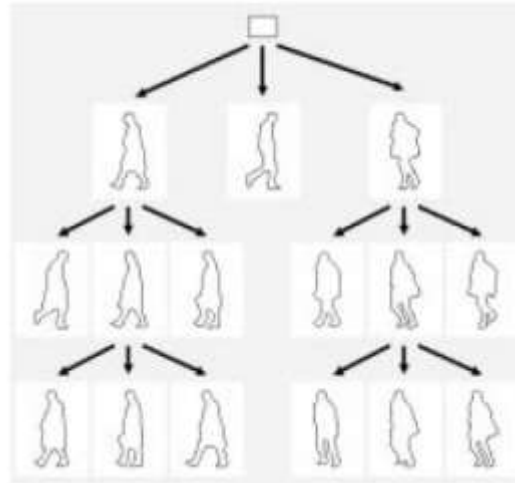


Fig. 2.5: Hierarchical shape detection [23]

Moreover, model based detection is the most common approach used for real time pedestrian detection. Even though it is the most common approach used, it has some cons also. It needs a lot of work to initialize the data set for matching the pedestrian. For example, the size, height and the postures vary for each pedestrian, which results in large datasets. Another disadvantage is that it can create accuracy issues if the contour is not properly extracted.

2.2.2.3. Feature extraction method

In this approach, the pedestrian is defined entirely based on each of his features such as size, height and angles formed while walking. Once the detection is done, a rectangle is drawn around the position of the pedestrian. The detection is done by frame classification from the given images or video clips and each frame is divided to several regions. Each division is performed based on the size of the trained features. Once the divisions are detected useful features like LBP (Local binary Pattern) features and HOG features are chosen for extraction. The block diagram for the feature extraction is shown in Figure 2.6 below.



Fig. 2.6: Block diagram for feature extraction

The LBP (Local binary Pattern) - a texture based approach was given by Ojala et al. [24,25] in order to gain texture classification. Most commonly used for face recognition, facial expression and face detection and so on. The pros of this method is that it has low computational complexities, easy and advanced, combined traditional and statistical approaches.

Another approach which can be commonly used is HOG (Histogram of Gradient). HOG is a method for encoding and matching image patches at different image orientation and scale changes [1]. It can be applied using a rectangular sampling window on the image and the gradient direction of each pixel in an image can be calculated by convolution with a Sobel mask at x and y directions [1]. The convolved ratio obtained by x and y direction gives the image feature direction. The gradient direction in each pixel can be combined in nine directions and thus forming 180 degrees. HOG can be computed by binding the pixels into the window and weighting each response by its edge strength [1]. This can result in an 8 x 1 vector which can be normalized to the range [0,1]. The illustration of HOG can be seen in the Figure 2.7. The advantages are it is less affected by illumination, it has good tolerance at different pedestrian postures such as walking, bicycling and so on [20].

Stereovision-based obstacle detection is another method where camera can extract the obstacle and Region of Interest from the scene by computing the depth of the images. In this method, the process initially adjusting the surface of the road by using the road parameter SPM (Surface

Parallax Map). Secondly, transform the occluded image using transformation equations and finally, obstacles are detected.

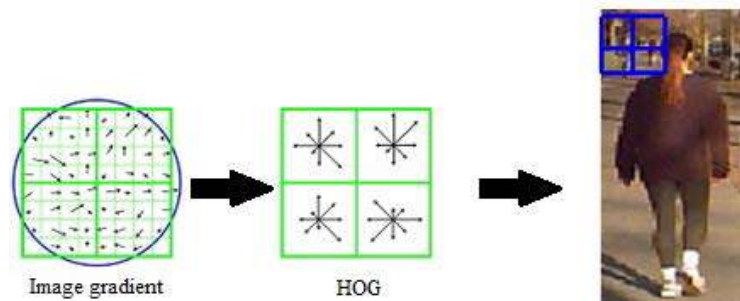


Fig. 2.7: Illustration of HOG

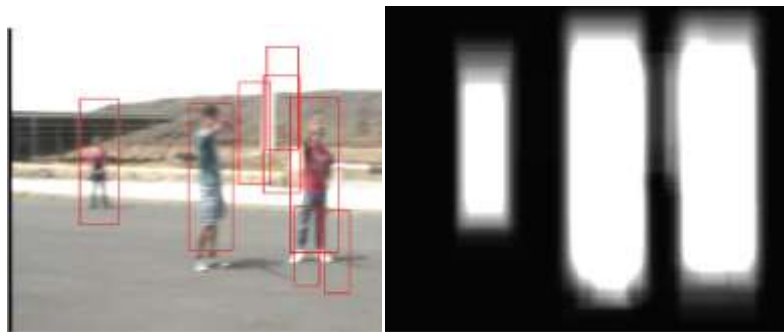


Fig. 2.8: Detection based on Probabilistic approach

Viola-Jones' method is another approach where the image features can be extracted and contain classifier cascades. To make it simpler a Bayesian approach is added to the Viola-Jones approach to detect the pedestrian in image frames. This is done by the statistical interpretation of original algorithm and the aim is to rise the precision of the autonomous vehicle to detect the pedestrians in image sequence. The Figure 2.8 shows the detection based on probabilistic approach and the advantage of this method is that it has lesser false rate.

2.2.2.4. Radar and Lidar sensors

RADAR (RADio Detection And Ranging) sensors uses FMCW (Frequency Modulated Continuous Wave) that can measure R , range of sensors and v , radial speed at various targets. It uses short radio pulses to transmit to the target and can measure the time that reflects back from the target. The reflected signal can be down converted and Fourier transformed to get the

features of the target. The range and velocity resolution, ΔR and Δv can be determined by the following equation [8]:

$$\Delta R = \frac{c}{2} * \frac{1}{f_{sweep}} \quad \text{equ.3}$$

$$\Delta v = -\frac{\lambda}{2} * \frac{1}{T_{CPI}} \quad \text{equ.4}$$

where T_{CPI} = Chirp duration and f_{sweep} = bandwidth.

Figure 2.9. below shows the coverage area for the radar sensors attached to the vehicle. Radar sensors are classified based on the carrier frequency and the maximum range covered and the use is based on the application for detecting the pedestrian.

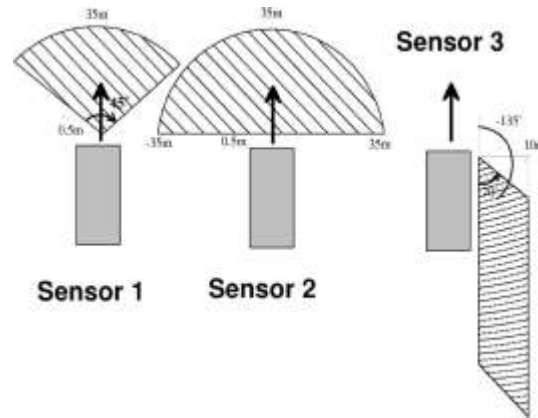


Fig. 2.9: Coverage for the sensors

LIDAR(Light Detection And Ranging) is yet another technology used in the family of Lasers. It's also a sensing technology similar to radar. LIDAR uses small wavelength as the source energy. These smaller wavelength signals can be reflected back from the objects or targets which can be measured using radar sensors. It uses series of light pulses for illuminating the target and statistical methods are used to classify the objects such as moving and non-moving objects.

2.3. COMPARISION OF DIFFERENT TECHNIQUES DISCUSSED

Sl. no	Detection Method	Advantages	Disadvantages
1.	Shape based detection	i. Hierarchical method based template matching ii. accurate method	i. large sets of template matching are used which takes time for processing
2.	Feature extraction method	i. less affected by illumination ii. good tolerance at different pedestrian postures such as walking, bicycling and so on	i. some of the features of the dataset have complex computation.
3.	Radar and Lidar sensors	i. Able to measure distance and speed	i. Not an efficient classifier and there are possibilities for high false positive in the images

CHAPTER 3.

THEORY

In this chapter, the theory related to the implementation of the system is discussed. Since the pedestrian detection is real time there will be many challenges which need to be overcome. Some of the challenges include thresholding, image segmentation, edge detection and noise filtering. Each of which is described detailed in the following sections.

3.1. IMAGE ACQUISITION AND FRAME CLASSIFICATION

Images can be acquired by camera model or sensors such as radar or lidar sensors. So, in order to establish a connection between pixels in digital cameras and the distance between the physical models such as pedestrians, objects and so on, it is necessary to acquire an image from the real world. The most commonly seen approach is using a digital camera or a 'pin-hole camera' to acquire the image. An ideal pin-hole camera consists of a very thin pinhole placed on a plane. The image reflected from the object, enters into the pinhole camera and later falls into the plane(usually a film). This leads to creation of an 'image'. The Figure 3.1 shows the creation of the image on a film [26].

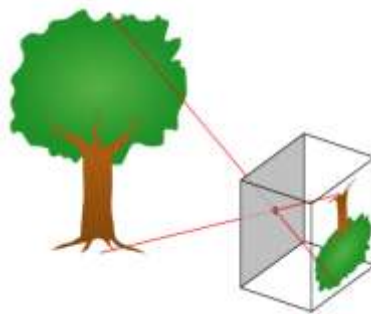


Fig. 3.1.:Image capturing in a pin-hole camera (Wikipedia)

The mapping between the image and camera can be measured by calculating the distance from the object to pinhole and from pinhole to image. Therefore in a digital camera, if Z is the distance to the lens and f is the focal length, that is the distance between the physical focal

length of the lens and the image (usually represented as millimeters). The Figure 3.2 below shows the mechanism in a pin-hole camera.

$$-y = f \cdot \frac{Y}{Z} \quad \text{equ.5}$$

$$-x = f \cdot \frac{X}{Z} \quad \text{equ.6}$$

where, the point location, $Q = (X, Y, Z)$ and the plane is represented as $P = (x, y, f)$.

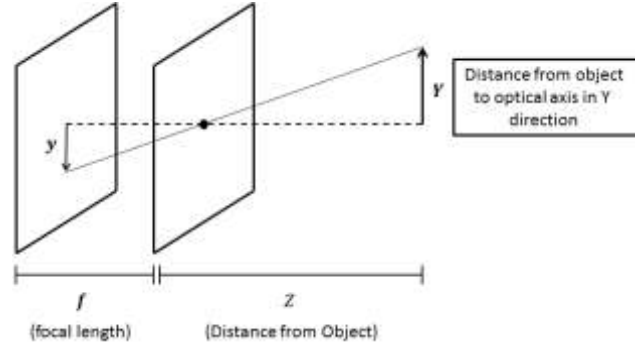


Fig. 3.2: Coordinates in pinhole camera

Even though pinhole camera is a common method, it has some drawbacks when capturing the image from the pinhole. As the pinhole can permit one image at a time it can practically slow down for videos. To overcome such situations modern camera can be used for the experimentation. In modern camera, the lens can focus to multiple rays from the light which can fall on the sensor such as CCD or CMOS. The common computer vision problems are: [26]

i. Lens displacement: It is difficult to ensure that the center of CCD imager will fall exactly on the center of the lens. Due to the slight alignment on X or Y direction relative to the optical axis [27], a correction is to be made for the distortion and equations are modified as shown below:

$$x = f_x \cdot \frac{X}{Z} + C_x \quad \text{equ.7}$$

$$y = f_y \cdot \frac{Y}{Z} + C_y \quad \text{equ.8}$$

where, f_x and f_y represents focal length in XY direction and C_x and C_y are the camera's principle point

ii. Radial distortion: In this case, lens may not be shaped perfectly and the light may bent further from the center of the lens. These radial distortions can be corrected by the following equations.

$$x_{corr} = x (1 + k_1 r^2 + k_1 r^4 + k_1 r^6) \quad \text{equ.9}$$

$$y_{corr} = y (1 + k_1 r^2 + k_1 r^4 + k_1 r^6) \quad \text{equ.10}$$

Therefore, in order to eliminate these distortions all the coefficients used in equation 9 and 10 need to be analyzed accurately.

The captured video is then converted to frames for further analysis such as image segmentation to distinguish the foreground and background image. The frames are classified according to the video captured from the digital camera and each can produce 10 frames/second. The number of the frame division depends on the time taken to capture the video and longer the time taken to capture the video more the number of frames classified. These frames are stored on bitmap files and each of them are arranged in the same sequence as the video captured. The Figure 3.3 shows the classification of frames. For example, if the video is captured for 10 seconds then the number of frames will be 100 and will be represented in RGB format. Finally, the frames are combined to form the complete video.

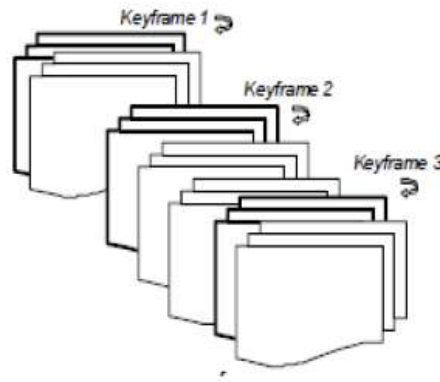


Fig. 3.3: Successive frames used for background estimation [27]

3.2. RGB COLOR MODEL

Color models represent specification of certain color in certain standard and in most cases, the models are specified by the coordinate system or a subspace where each color is represented to

specific point. In RGB model, each are specified by primary colors such as red, green and blue and each color is represented as Cartesian coordinate. The color model can be represented around as coordinates on a cube with red, green and blue located at the three corners and the secondary color cyan, magenta and yellow located at the other three corners. In this model, black is at origin, white is at the other corner farther from origin and the gray scale represents the line that expands from black to white. Normally RGB values are in the range from [0,1]. The number of bit that is used to constitute each RGB pixel is known as pixel depth. For example, if the RGB images are represented by an 8 bit image, each RGB pixel will have a depth of 24 bits and the full color image is represented as 24 bit RGB image. The total number of color in a 24 bit RGB image is $(2^8)^3 = 16777216$. Color planes are better ways to view the primary and secondary colors. This is done by fixing one of the three colors and other two colors to vary. For example, if the R value =127 is fixed, then G and B will vary from 0,1,...255. Figures 3.4 and 3.5 below represents the pixel representation for gray and RGB images.

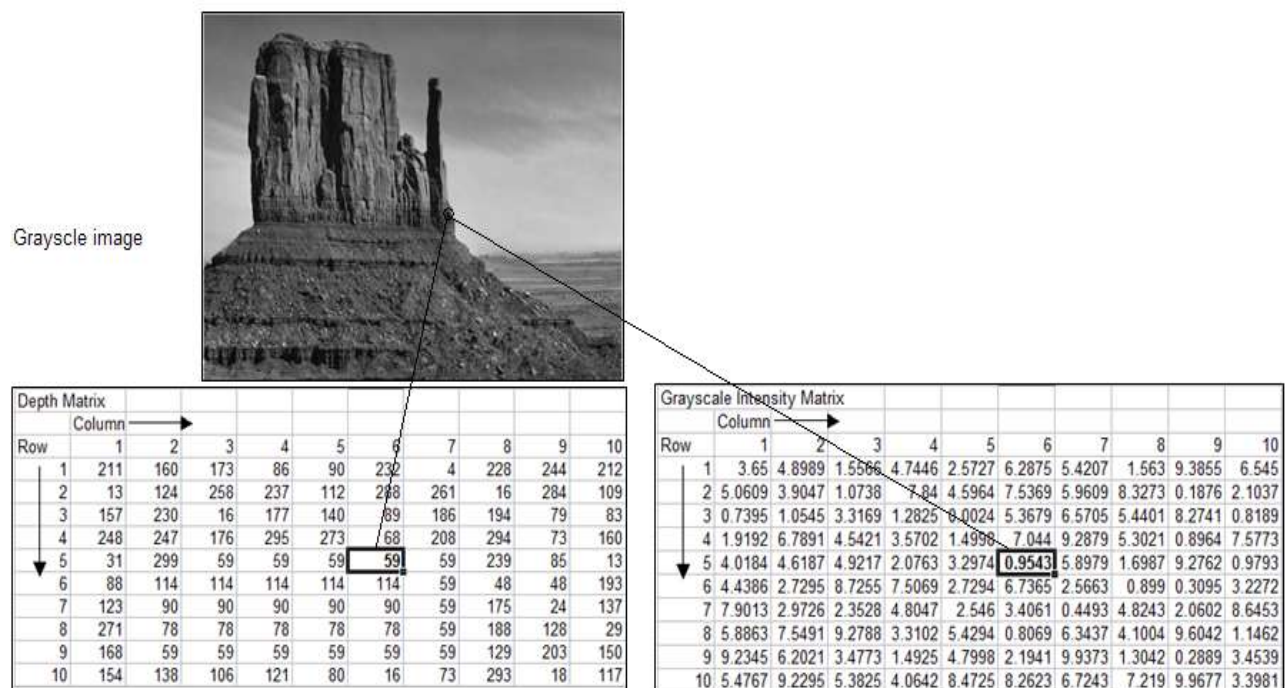


Fig. 3.4: Gray image storage representation[12]

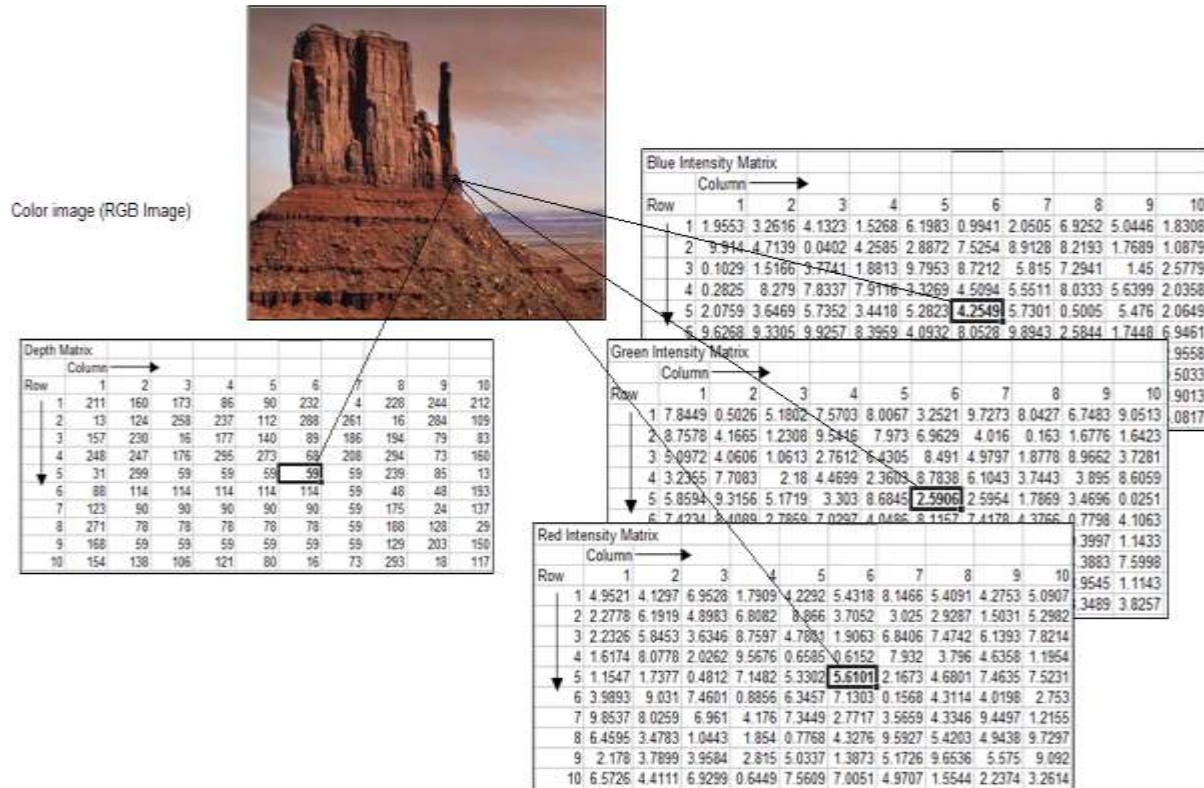


Fig. 3.5: RGB image storage representation

A color image is usually captured by using filters which are sensitive to red, green and blue respectively and if these filters are attached to the monochrome cameras then the images formed are the color image of RGB colors of the color scene. In many practical applications, the RGB color images are integrated to the devices. Therefore, the benefits of using RGB color models is that the color represents the channels differently and conserve the native values in the RGB channel.

3.3. IMAGE SEGMENTATION

The main aim in segmentation is to convert the digital image to some useful information such as detecting moving and non-moving objects. Once the frames are obtained from the video, the images are segmented to extract the pixel information from each image frame. Each pixel is represented as rectangular grids and each grid has pixel information stored in it. There are many useful information such as vehicle location, obstacles, objects, pedestrian and sign board stored on the pixels. The major challenge in image segmentation is that it is difficult to convert

undifferentiated image planes [26] into sub regions in order to measure discrete objects. Each object seen in the image is different from background due to many lighting conditions or due to variation in camera such as illumination changes for each scene under the weather conditions. To study these variations various image segmentation methods are proposed which are discussed below.

3.3.1. Thresholding

Thresholding is a common approach used for image analysis and it is a simple method used for image segmentation. It is used to construct the binary images from the gray scale images and this is being performed by initializing a threshold (cut-off value) and if the values are greater than the threshold value then the pixels are classified as background, while the others are categorized as foreground pixel. For simplicity, the pixel '1' is given to the pixels for the object or the image and these are assigned as foreground pixels. However, the pixel '0' is set for the background. The illustration of thresholding at 128 is shown in Figure 3.6. The advantage of thresholding is that the foreground and the background images can be easily separated. There are many algorithms to perform these operations such as Otsu's method (maximum variance), maximum entropy method and so on. Moreover, thresholding can be divided as bi-level and multi-level thresholding algorithm. In the bi-level thresholding, the pixels are divided into levels based on certain threshold value and the binary images are obtained from the gray level images. On the other hand, in multilevel thresholding more than one threshold value is selected. This method is just an optimum method to obtain the results.



Fig. 3.6: Thresholding at $t=128$

3.3.2. Segmenting RGB values

The objective is to segment each RGB pixel for the specified frame by using a range of values. To segment objects in a color image by using the specified RGB values, average color value estimation method needs to be performed. Let the average color value for RGB vector be denoted as 'a'. Then the RGB color segmentation can be performed by measuring the similarity factor among the colors. One of the simple methods used for segmentation is Euclidean distance method. Assume, 'z' to be an arbitrary point in the RGB plane. Then, both 'z' and 'a' are similar, if the distance between them is less than the threshold value, D_0 [29]. Therefore, the Euclidean distance method is given as:

$$\begin{aligned} D(z,a) &= \|z-a\| \\ &= [(z-a)^T (z-a)]^{1/2} \\ &= [(z_R - a_R)^2 + (z_G - a_G)^2 + (z_B - a_B)^2]^{1/2} \end{aligned} \quad \text{equ.11}$$

where, R, G and B denotes the RGB components for vector 'a' and 'z'. The locus of the points, $D(z,a) \leq D_0$ can be represented as a solid sphere of radius D_0 [29]. The points within the sphere satisfy the color criteria and vice versa. Finally, coding these sets of points in the image will produce a segmented binary image.

3.3.3. Background subtraction and estimation

In many stationary models the input frame can be classified into two frames such as stationary background and foreground. For example, in a scene there will be objects that are moving and some others are stationary for a particular interval of time. In this project, it is relevant to detect moving objects and there are challenges for color-based thresholding to detect moving objects. On the other hand, to detect other stationary objects template matching techniques are used. In this method mainly successive frames are chosen and later the successive frames are used to create motion within the frames. Over here, in this background estimation part the frames are initialized from the video clips and each frame is separated. Once the separation is done, the initial frame is set as the reference frame and the subsequent frames are compared with the reference frame to extract the foreground and background images. Here, the background images

are set to black color or 0 pixel. This method is applicable only if the background is stable. Moreover, the background can be updated according to the changes made within the time.

In background estimation method, the algorithm computes an incomplete background from the pixel that does not belong to the foreground pixel. If the foreground is moving then the background will estimate more background pixels. Moreover, once the estimation is finished, the background is subtracted from each video to separate the foreground images [27]. Then this foreground image is converted to binary images. Later, blob analysis and other morphological methods are used for each foreground image.

3.4. IMAGE ENHANCEMENT

Image enhancement is for enhancing the accountability and the recognition of pixel information within the image. The input blocks are the blocks that may contain some candidate block with some background objects and pedestrian. Image enhancement is to be used to improve the intensity of the pedestrian images in the candidate block [27]. There are many spatial domain methods that can perform enhancement on the pixels. The filtered images are then subjected to edge detection and other morphological operations.

3.4.1. Edge detection

Edge detection is the common technique used in computer vision to find the edges in the particular frame or in the image. Usually in edge detection, pedestrians or objects will be present only at a particular region on the image frame. Therefore it is not necessary to process the entire frame for edge detection. Only the particular Region of Interest (ROI) is needed for processing the image. Usually edge detection is used to segment the ROI and it can be performed by classical operations such as Sobel, Prewitt, Canny and these algorithms are applied to the binary images. For example, the equation for edge detection of 3*3 Sobel kernel at 45 degree and 135 degree are shown below in equation 12, 13 and 14 [27].

$$S_{90} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{equ.12}$$

$$S_{45} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{equ.13}$$

$$S_{135} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{equ.14}$$

At S_{45} , S_{90} and S_{135} are the kernels used to obtain the vertical and diagonal directions. These three kernels are applied to each pixel in the input frame. Once the edges are obtained, morphological operations are performed with the vertical kernels of 3 pixel height and 1 pixel width. Initially when performing the edge detection operation, there will occur many noises and these noises can be removed by filtering. However, using vertical edge detection method noises can be eliminated which occurred in the initial operation. There are other edge detection techniques like Sobel and Canny where these methods are sensitive to noise and has higher computational time. A new sequential technique was proposed by Hanene Rouabeh *et.al.*[28] to detect the edges with reduced computation time and improved accuracy. The Figure 3.7 shows the neighborhood structure for edge detection. From the proposed method the difference value between the original pixel and neighborhood average pixel values are calculated. According to the minimum value between original pixel and difference value, the new value is decided for the edge image. If the new value is higher than zero, the pixel value is set to 1 and vice versa [29].

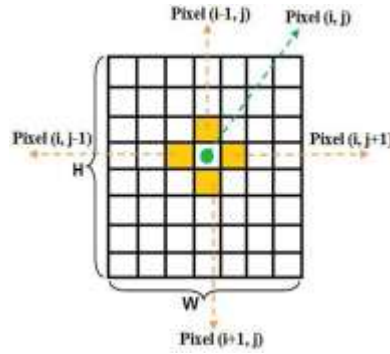


Fig. 3.7: Neighborhood structure for edge detection [28]

3.4.2. Blob detection and noise Filtering

Blob detection is the technique used for detecting objects like pedestrians for an image using component labeling. This technique consists of data structure that are used to save information on each blob. Certain blobs fall on predefined criteria such as the noise pixels which are small sized blob [27] and are removed from the image. Since the camera is placed on the vehicle, the camera captures the pedestrian at a certain distance. This can lead to lose of some pedestrian data

and in order to improve the accuracy, merging of the blob is performed. For example, if D_v , vertical and D_h , horizontal are the centroid distance between adjacent blobs, then

$$D_v < V_{\text{thresh}} \text{ and } D_h < H_{\text{thresh}} \quad \text{equ.15}$$

Once the thresholding process is completed, a slight modification is performed for the background and foreground modeling by noise filtering. This is usually performed by spatial filtering such as median or mean filtering. It can be performed the same way as the frame differencing, where the filters are applied to consecutive frames. This technique is more strong enough to change according to the background model [26]. In mean filter, the average of the image pixels in the defined area is calculated whereas median filter method is used to find median of the intensity levels of the neighboring pixels [29].

$$\text{Median filters: } B_t(x,y) = \text{median}\{I_{t-i}(x,y)\} \quad \text{equ.16}$$

$$\text{Mean filters: } B_t(x,y) = \frac{1}{n} \sum_{i=1}^{n-1} I_{t-i}(x,y), \text{ where } I_t(x,y) - B_t(x,y) > T \quad \text{equ.17}$$



Fig. 3.8: Effective and non-effective noise removal

Figure 3.8 shows stationary background segmentation for effective and non-effective noise removal filters using noise removal operation. When the objects are of the same color as that of the background, the method is ineffective because the objects occlude each other. The advantages of median and mean filters is that it can provide readymade segmentation of the boundaries and no post-processing is necessary for the all the pixels.

3.4.3. Histogram processing

Histogram is a technique used in digital image processing and it is a spatial domain method which is used for image enhancement. Histogram has an intensity level in the range $[0, L-1]$, (L is the size of the image) with a discrete function as $h(r_k) = n_k$ where r_k is the k th intensity value and

n_k is the number of pixels [29]. In practical application, normalizing histogram is performed by dividing total number of pixels in the image such as product of $M*N$, where M is the column and N is the row dimensions of the image [29].

$$\text{Normalized histogram, } p(r_k) = r_k / (M*N) \quad k=0,1,2,\dots,L-1 \quad \text{equ.18}$$

where $p(r_k)$ is also an estimate of the probability of r_k . The sum of all normalized histogram image components is equal to 1.

Moreover, histogram provides useful information of the image depending on the intensity and brightness. For example, for dark images the histogram distribution will be at the leftmost end and for light images it is the vice versa. For low contrast image the histogram distribution is on the center and the intensity scale will be wider for low contrast images. The Figure 3.9 shows the intensity distribution of red in RGB plane and its histogram image. Therefore, based on the information available from the histogram it is possible to develop a transformation function.

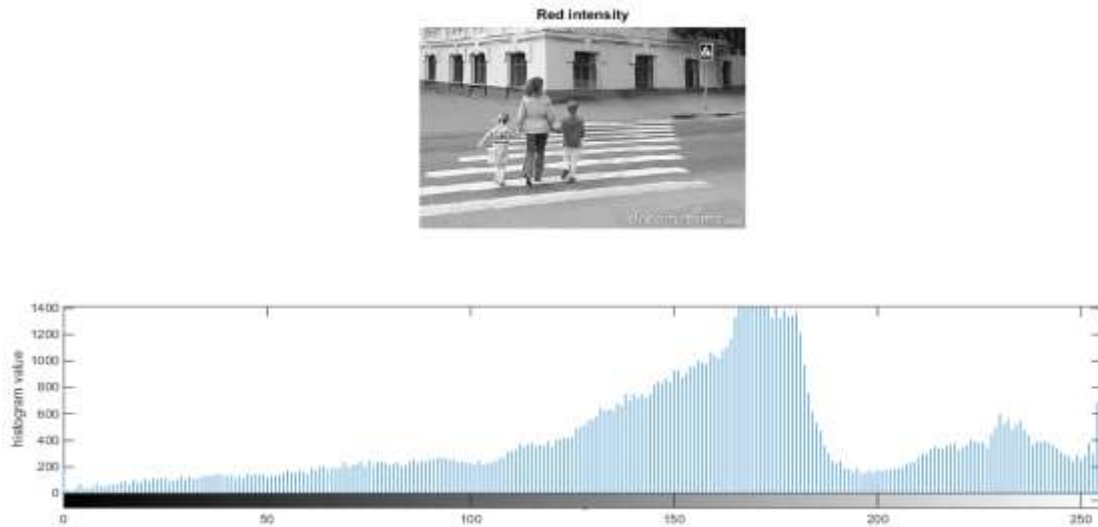


Fig. 3.9 : Histogram distribution for red pixels

3.5. IMAGE EXTRACTION - MORPHOLOGICAL IMAGE PROCESSING

Morphological operation is used for extracting image components such as image shape, boundaries, skeletons and so on. This is the final operation involved in the image extraction and the two main operations are erosion and dilation. Erosion is the operation that thins the boundaries in a binary image. Erosion is said as a morphological filtering operation in which

image details that are lesser than structuring elements are removed [29]. In general, erosion is a shrinking or thinning operation and on the other hand, dilation is a growing or thickening operation of a binary image. For example, the simple application of dilation is enhancing the written letters on an image. In this application, a structuring element acts as a filtering agent for modifying the gaps among the letters. Therefore, the advantages for the morphological operation is that it is possible to bridge the gaps.

CHAPTER 4.

SYSTEM IMPLEMENTATION AND RESULTS

In this chapter, system implementation of the proposed methods are described. This chapter will also discuss system overview of parameters used and assumptions to achieve the pedestrian detection. The system is designed for consecutive frame using Gaussian method and Region of Interest method and finally the results are compared for both cases with the shaped based detection method.

4.1. SYSTEM OVERVIEW

4.1.1. Device Specifications

A smart phone and a computer is used for processing the images and the specifications are shown below in the table 4.1 and 4.2.

Table 4.1: Camera parameter

Specification- digital camera[31]	Features
Type	Samsung Galaxy, S4
Focus adjustment	Automatic
Camera light source	Flash
Special effects	HDR
Video recording- resolution	1920*1080 (1080p)
Graphics Accelerator	Qualcomm ADRENO 320
Flash memory	16GB

The hardware specifications are shown in table 4.2:

Table 4.2 : Computer hardware specifications

Specifications	Features
Processor model	Intel core i5-4200M CPU @ 2.5 GHz
RAM memory	8 GB
System type	64 bit

4.1.2. Software Specifications

The software used for implementation is Matlab. It is a matrix based language to express computational mathematics. It is a high-level language tool for engineering and scientific applications. This software is selected because it is very user friendly and it uses graphics for focusing data values and tools for custom plots such as 2D and 3D plots. Moreover, it has additional features such as image processing toolbox, signal processing, and many more applications. Matlab also provides the traditional programming features such as control flow, error handling, source integration and testing. The features are specified in this table 4.3 [31].

Table 4.3: Matlab specification

System requirement	Features
Operating system	Any operating system such as Windows or MAC
Disk space	2GB
RAM memory	2GB, for simulink-4GB
Graphics	OpenGL 3.3 having 1GB GPU memory

The other advantages of matlab is that it can convert matlab algorithms to HDL or C/C++ language. The disadvantage is that for image processing it cannot process higher pixel values.

4.1.3. Device coordination

Since all the equipment are taken separately, the coordination of all the devices are necessary. To coordinate the smart phone and pedestrian together some issues need to be overcome such as camera calibration. There can be an issue with the focal length between the distance of the pedestrian and the smart phone. The adjustment of the camera is made in such a way that camera is fixed at a position within the vehicle. The field of view between pedestrian and camera need to be fixed before beginning the experiment. The second issue is to overcome the video transmission because the cameras cannot transmit at a fast rate to the computer and it can be lagged by 20-30 frames. To overcome this problem, a specific driver such as USB is used to help decrease frame lagging [20].

4.2. SYSTEM DESIGN

4.2.1. Assumptions and Flow diagram

This section discusses the approach to achieve the research goals. There are certain assumptions to be considered for achieving the results such as :

1. Vehicle is turning right on a cross road.
2. Speed of the vehicle is 0-30 km/h.
3. The minimum distance between the pedestrian and vehicle is 2 meters and maximum should be 8 meters.
4. The computer system need to be equipped with matlab language.

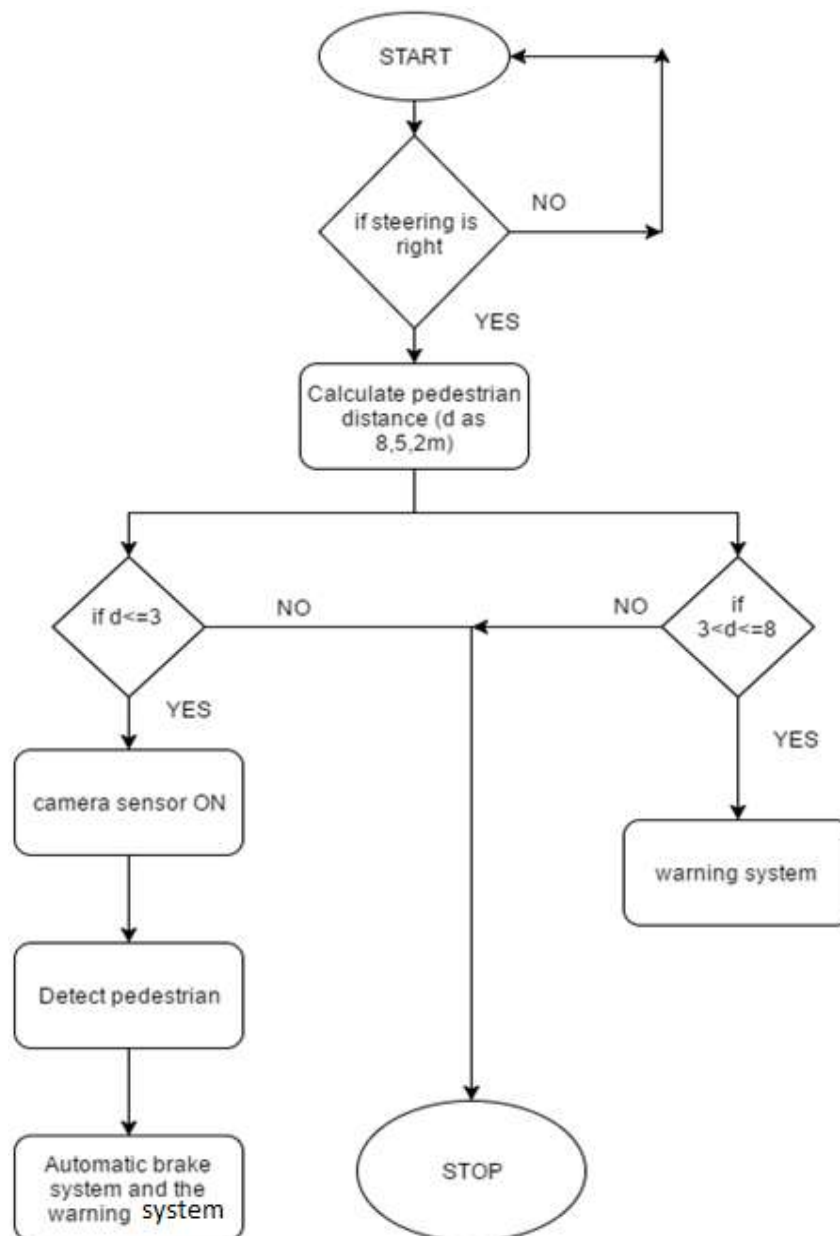


Fig. 4.1: Flow chart for pedestrian detection

Once the pedestrian approaches the cross road, the sensors (for example, smart camera) attached to the vehicle can detect the pedestrian within 2-8m of range. The necessary actions such as warning and/or automatic braking are taken according to conditions in the flowchart specified in Figure 4.1.

4.2.2. Proposed design

If the steering wheel angle sensor detect the vehicle is turning right, then the camera sensors attached to the vehicle will detect the pedestrian. The distance measurement technique is not shown in this proposed design. It is assumed that once the distance measurement between the pedestrian and the vehicle is determined the remaining steps will continue to work. If the distance is less than or equal to 8m, then the gas pedal and the head tilt are detected using the wheel speed sensors which are already attached to the vehicle. If the head tilt is left and a pedestrian is detected while the vehicle turning right and at this point an early warning pop up message will appear to slow down the vehicle which can alert the driver to avoid a possible threat.

In the second stage, if distance is less than or equal to 5m and speed wheel sensors still detect the rotation then an acute warning (audio and visual warning) will be activated and a partial auto-brake will be applied.

In the third stage, if the vehicle is still moving and the distance measured is less than or equal to 3m the proposed system will apply a strong auto-brake. Figure 4.2 explains all the three stages of the proposed method.

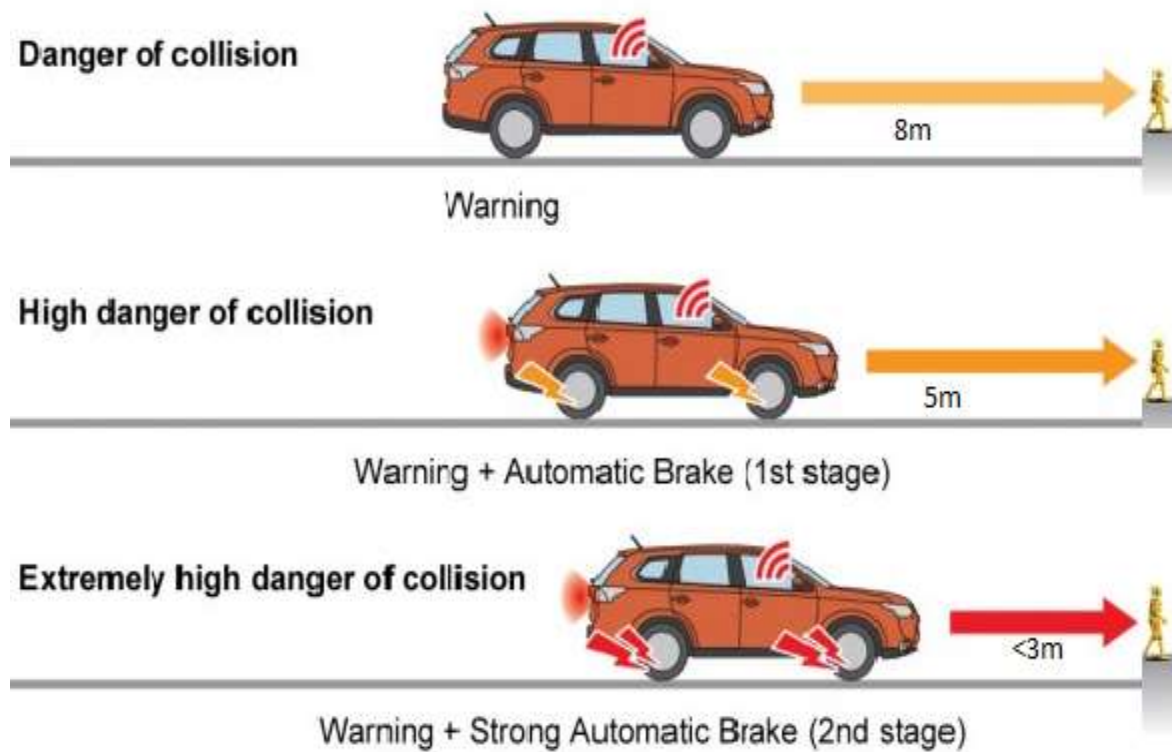


Fig. 4.2 : Three stages in the proposed method

4.3. ANALYSIS OF SYSTEM DESIGN FOR CONSECUTIVE FRAMES BASED ON ROI METHOD

4.3.1. Image acquisition and frame generation

Image capturing is done using a Samsung video camera and the captured image can contain a wide range of RGB pixels. This camera is mounted on the right side of the passenger seat and once the vehicle takes a turn at an intersection the camera turns on. Images are captured using the smart phone and five consecutive pictures/frames are taken. The resolution of the image should be at higher acquisition speed to acquire faster events. Therefore, a total of 5 consecutive datasets are taken.

4.3.2. RGB Segmentation

Once the digital image is being captured from the digital camera of the smart phone, the image is then transmitted to the computer through USB. The image size is taken as 1835x2322x3 uint8 and the image contains red, green and blue components. This section is called segmenting the

RGB values and can be represented in three planes. Initially, the image is reshaped to generate the gray image. This gray image is used for further analysis using histogram method. Histogram is used where the image intensity scale can be defined and all the image information is obtained. This histogram equalization is performed for all the three RGB planes. The Figure 4.3 shows the reshaped RGB and the histogram intensity values. As seen from the image the intensity of the blue value is distributed to a wider range than the red and green intensity values. Finally, the RGB values are extracted separately. The colorbar for all the three planes are shown in Figure 4.4.

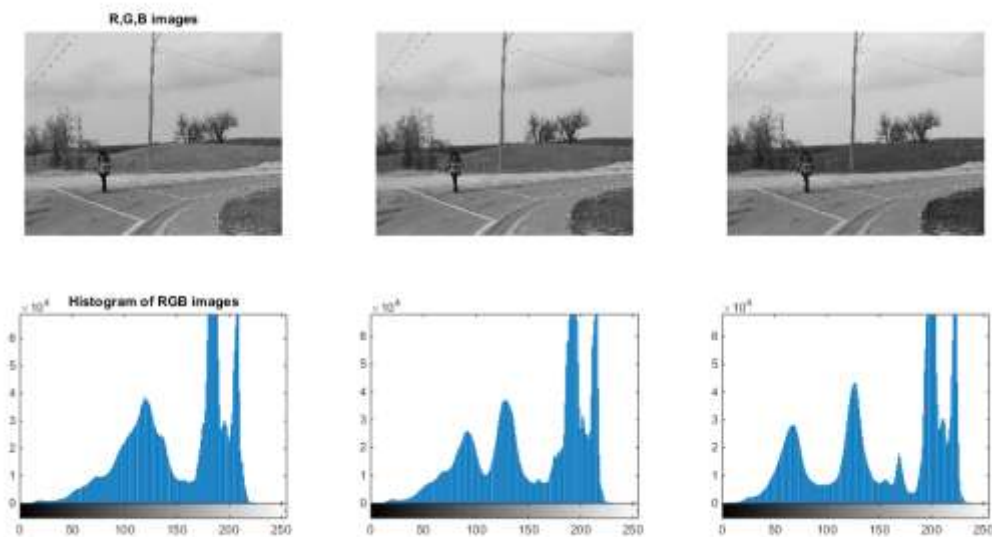


Fig. 4.3: Histogram for RGB models

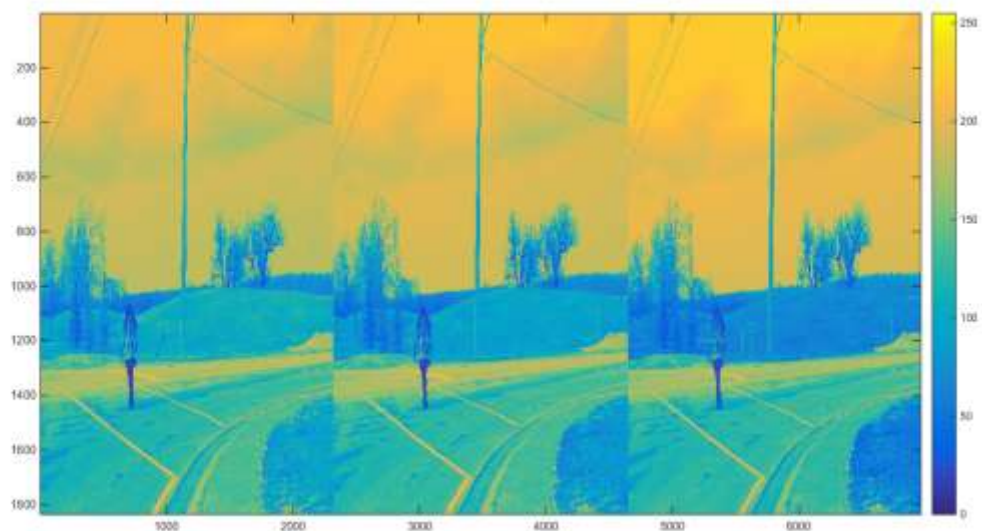


Fig. 4.4: Colorbar for RGB plane

4.3.3. Separating background and foreground

This method is used for extracting the foreground images from the background. The purpose of foreground extraction is that to reduce the number of objects in the image to attain a better accuracy. The advantage of this type of separation is that the small sized objects can be analyzed very easily. The feature extraction process is performed by calculating the threshold value and later blueness value. Thresholding is a common technique which is used for separating the foreground and the background is set to black. This thresholding technique is explained below:

4.3.3.1. Color based threshold detection technique

Thresholding is the common approach used for segmentation and is used to obtain a binary image from the given sample image. The threshold value is selected through trial and error method. This can be determined by the following equation [31]:

$$\text{Threshold, } t = (A \geq \text{low}) \ \& \ (A \leq \text{high}) \quad \text{equ.19}$$

where , low=3 and high=30 and A is the image taken for thresholding.

Once the gray images from the RGB models are extracted, then the thresholding is done for the three planes separately and compared for the different threshold values. From the analysis, choosing lower threshold value resulted in disappearance of some features and vice versa. Based on the histogram representation, the blue scale has a wide range of intensity values. Therefore, blue scale thresholding can return a binary image from the low and high values that are selected from the color map model. Thus, the foreground is separated from the background.

4.3.4. Calculation of blueness value

From the histogram analysis (Figure 4.3), the last two components such as the green and blue components have high intensity values. This project focuses on finding the blue pixels for analyzing the pedestrian, and it does not mean that all pixels in the image have high blue component. For differentiating the blue pixels, blueness value needs to be calculated such that pedestrian gets more darker and thus the background can be separated easily. As seen in the Figure 4.3 and 4.4 blue pixels have high intensity and other two components (red and green pixels) have low intensity. Therefore, the blueness pixel value is calculated as [32] :

$$b=B-\max(R,G) \quad \text{equ.20}$$

where b=blueness pixel, B= blue channel, R= red channel, G=green channel

For example, if RGB range is (255,0,0), then red pixel is 255, green and blue channel is 0 and blueness, b=-255 and the table 4.4 explained in detail for the other color components.

Table 4.4: Calculation of blueness for different color combinations

Color components	Blueness value
(255,0,0)	-255
(0,255,0)	-255
(0,0,255)	255
(127,127,255)	128
(255,0,255)	0
(0,255,255)	0

4.3.5. Pedestrian detection based on ROI generation

The final stage of the pedestrian detection is to attain full information about the shape and size of the captured image. From the blueness value calculation and thresholding technique, the pedestrian in the image gets more noise free and becomes more clearer in the frame. Later, the required ROI (Region Of Interest) matrix is formed and the pedestrian is detected. The importance of ROI method is that it is a simple method and easier to define the matrix such that it can filter out the unwanted portions in the frame. The ROI acts as a binary mask and the required ROI pixel is set to 1 and others set to 0. When the pedestrian comes within the region of interest, warnings are initiated by the onboard system. This operation can be performed easily using the *imroi* matlab function. Thus, the false negative problem is eliminated. The results of segmentation and pedestrian detection using ROI based method is shown in Figures 4.5 and 4.6 respectively. The axis in Figure 4.6 represents the colorbar or the intensity values in blue pixels. From the five consecutive images captured, all images are detected correctly and resulted in lesser false detection.



Fig. 4.5: Thresholding for separating foreground and background

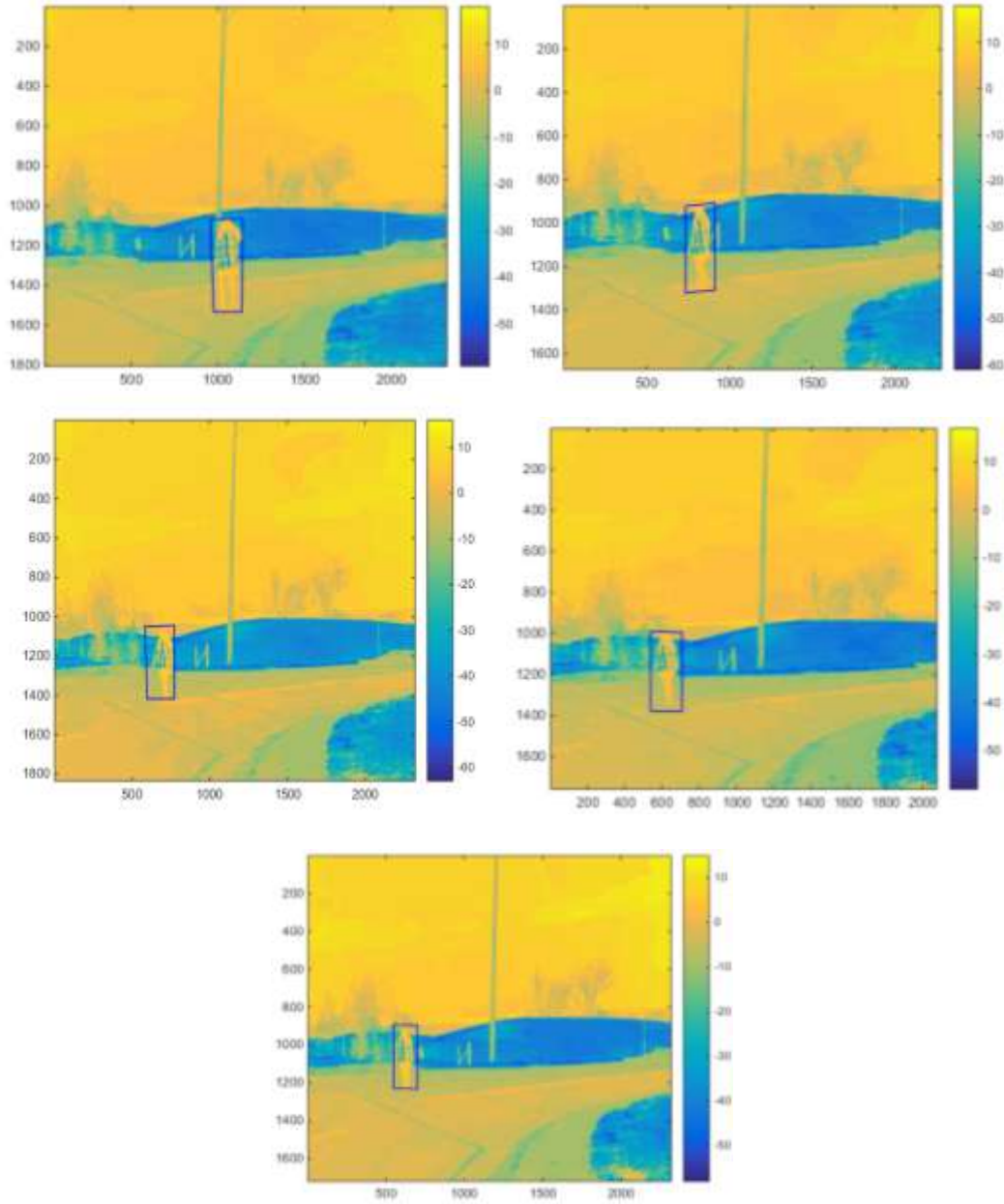


Fig. 4.6: Pedestrian detected using ROI method

4.4. ANALYSIS OF SYSTEM DESIGN FOR CONSECUTIVE FRAME USING GAUSSIAN MODEL

This is another method used for detecting the pedestrians in a crossroad. Image acquisition method is same as the method explained in section 4.3 and the camera used is the same as the Samsung smart phone. In this technique, predefined blobs are used for separating the foreground and background images and this is the main difference from the ROI method.

4.4.1. Frame classification

The video acquired from the smart phone in a top elevation is fed to the matlab and the image size is 1080 height and 1920 width. For example, if the video is captured for 10 seconds and the total number of frames will be 10 that are stored in RGB format. The frame 1 is the consecutive frame that is the base frame and set as background. The remaining frames are classified according to the values chosen for separating the pedestrian from the base frame. Thus, 10 consecutive frames are used as the data set for detection.

4.4.2. Foreground and background separation

The foreground separation is performed based on built-in functions on computer vision using Gaussian models. Initially, the frames are converted to gray scale images and the 'foreground detector' object compares the gray scale video frame to the background model to determine whether pixels belongs to foreground or background. Thus, the foreground mask can be computed using background subtraction method taken from the stationary camera. Figure 4.7 shows foreground separation.

4.4.3. Pedestrian detection based on Gaussian mixture method

Once the foreground and background separation is done, the region of interest is selected based on blob detection. This scheme establishes pedestrian-like objects using connected labels and represents in a binary image form. The blobs that have predefined dimension will be retained in the frame and other pixels removed from the frame. In these conditions, pedestrians who are closer to the camera at a distance of 10-30m can be detected efficiently.

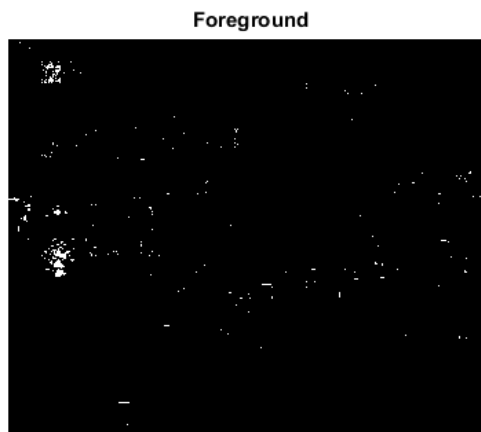


Fig. 4.7: Pedestrian frame 204 at top elevation

The advantage of this method is that it improves accuracy. Moreover, the method uses feature extraction such as statistics returned from the binary image. The Figures 4.7 and 4.8 represent foreground segmentation and pedestrian detection using Gaussian method.



Fig. 4.8: Foreground separation of the detected pedestrian

4.5. DISCUSSIONS

The algorithms applied for the pedestrian detection based on ROI method and Gaussian mixture method were effective. There are still some drawbacks which need to be overcome, but still the method has optimized in many ways:

1. The field of view was smaller because of the small camera size.
2. Algorithm used is simple.

In the ROI based method there is no blob detection used, but it is based on the calculation of blueness value. However, using Gaussian mixture method, blob detection is used to get the connected regions for detecting the pedestrian.

The pedestrian detection is tested at right turn intersection of vehicles for a set of consecutive frames. The pedestrian is detected with the help of a smart phone and further analysis is done using matlab coding. The developed system has overcome many of the cons of the other method such as shape based detection method (PROTECTOR system) of pedestrian detection. The detection method used in this project is infrastructure based system, because the detection system deals with fixed background and it will be easier to achieve the required results. Even though the method is simpler, it has got its own drawback as it is difficult to classify pedestrian in a crowded situation. When both the proposed pedestrian detection methods are compared with the

template matching method, the contour or the object mask required for detecting the pedestrian is no longer used. As seen from the table 4.5, the performance rate of ROI method and the Gaussian method are 99.9% and 90.2% respectively, while the template matching method has 93.75% performance rate. Moreover, the processing speed for ROI method and Gaussian method were 30sec and 1-2 sec. Thus, on conclusion the results of the three methods were almost similar while considering the performance rate and processing speed.

Table 4.5: Comparison for both the proposed methods

Sl no	Method	Sensor type used	Frames/ data set	Contour used	Field of view	Processing speed	Performance rate
1.	ROI method based detection	Smartphone camera	5 frames/ images	No	small	30sec	Detection rate is 99.9%
2.	Gaussian model based detection	Smart phone camera	10 frames	No	small	1-2 sec	Detection rate is 90.2%, 9.8% false positive/frame
3.	Template matching- PROTECTOR system- [Wei Zhang, [12]]	Stereo vision 3D sensors	100 frames	Yes	large	1-2sec	Detection rate is 93.75%, 6.25% false positive/frame

Therefore, from the table 4.5, the advantages in Gaussian mixture and ROI based method are lesser false rate when compared to the PROTECTOR system. Moreover, there is no need for large datasets for matching the shape of the pedestrian, which increased the proposed methods speed and accuracy.

CHAPTER 5.

CONCLUSION AND FUTURE WORK

5.1. CONCLUSION

The proposed project is based on improving the safety of pedestrian crossing the road at an intersection. From the past five year pedestrian death rate analysis, it can be seen that there is a high increase in the accidents occurring at the intersections. Overall, Toronto city has faced a high pedestrian death rate of 67% in 2013, within a span of 10 years. Thus, the proposed project can lower the pedestrian death rate by alerting the drivers of pedestrian on the road.

The first chapter in this project, describes background survey of the accidents happened in Toronto and this became the motivation for this pedestrian detection proposal. This section discusses about the objectives and scope of the project and the proposed technique focused on one scenario, where the vehicle turns right at an intersection. The second chapter, describes about the literature survey related to pedestrian detection. It also discusses about some shortcomings of the existing pedestrian detection methods such as contour matching and use of large datasets. Moreover, there are many comparisons being discussed between the existing methods for pedestrian detection.

The main topics are described on chapters three and four, where the pedestrian detection technique is being specified. This is implemented using matlab programming language and using a smart phone camera attached to the vehicle. Once the camera is attached to the vehicle, it takes pictures from the phone and it is then transferred to the computer for processing. The new technique used for pedestrian detection studied for consecutive frames. This was designed by calculating the blueness value and finding the pedestrian from the specified region of interest. For example, if the pedestrian is within 2 meter distance from the vehicle, drivers are given warnings to stop the vehicle. However, for the consecutive frame using Gaussian mixture model method, uses blobs for the pedestrian detection. Finally, both the methods are compared with

feature based pedestrian detection technique. Overall, all the discussed methods have similar processing speed and performance rate.

5.2. FUTURE WORK

In future the entire system compatibility will be improved in such a way that a hardware implementation will be done. A radar sensor will be attached to the bumper of the vehicle for measuring distance between the pedestrian and vehicle. Once the distance is measured, it will be easier to measure the pedestrian crossing the road and an auto-brake function and a sound warning will be incorporated to stop the vehicle. Thus, the traffic scenario can be improved significantly.

REFERENCE

- [1] Jayan Eledath, Bogdan Matei, Mayank Bansal, Sang-Hack Jung, and Harpreet Sawhney, "Layered Object Recognition System for Pedestrian Sensing," US Department of Transportation in October 2012. doi:[FHWA-HRT-11-056].
- [2] David Hains, "Toronto's Bad Year For Traffic Fatalities," *Cityscape* (16 January 2014) [Online]. Available: <http://torontoist.com/2014/01/torontos-bad-year-for-traffic-fatalities/>
- [3] Tim Alamenciak, "Pedestrian Fatalities In Toronto Hit 10-Year High," *Toronto Star* (22 July 2015).[Online].Available: http://www.thestar.com/news/gta/2013/12/06/pedestrian_fatalities_in_toronto_hit_10year_high.html.
- [4] Julia Werneke, and Mark Vollrath. "Analysis of intersection accidents-An accident causation and prevention perspective, *Proc. HFES*,Chalmers University of Technology, Goteborg, Sweden, 2012, pp.363-372.
- [5] Constant, Aymery, and Emmanuel Lagarde."Protecting Vulnerable Road Users From Injury," *PLoS Med*, vol. 7, no. 3 p. e1000228, 2010.
- [6] R. A. Retting, S. A. Ferguson, and A. T. McCartt, "A review of evidence based traffic engineering measures designed to reduce pedestrian–motor vehicle crashes," *Amer. J. Public Health*, vol. 93, no. 9, pp. 1456–1463, Sep. 2003.
- [7] Margaret Rouse,(2005-2016) "Passive Sensor," WhatIs.com., [Online]. Available: <http://internetofthingsagenda.techtarget.com/definition/passive-sensor>.
- [8] Steffen Heuel and Hermann Rohling, " Pedestrian Recognition Based on 24 GHz Radar Sensors, " *IEEE Radar Symposium (IRS)*, Vilnius, Lithuania, June 2010, pp.1-6.

- [9] Gavrilă D.M. and Munder S. "Multi-cue pedestrian detection and tracking from a moving vehicle," *Int. J. Comput Vision*, vol. 73, no. 1, pp. 41–59, July 2006.
- [10] Giuseppe De Nicolao, Antonella Ferrara, Luisa Giacomini, "Onboard Sensor-Based Collision Risk Assessment to Improve Pedestrians' Safety," *IEEE Trans. Veh. Technol*, vol. 56, no. 5, pp. 2405-2413, Sept. 2007.
- [11] Cristiano Pretebida, Joao Carreira, Jorge Batista and Urbano Nunes, "Pedestrian Detection Combining RGB and Dense LIDAR Data," *IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, pp. 4112 - 4117, Sept. 2014.
- [12] Wei Zhang, "In-Vehicle Pedestrian Detection using Stereo Vision Technology," *3rd Int. Conf. on Road Safety and Simulation, Indianapolis, USA*, Sept. 2011, pp. 14-16.
- [13] Aline Senart et al., "Using Sensor Networks for Pedestrian Detection," *5th IEEE Consumer Comm. and Network Conf.*, Las Vegas, NV, 2008, pp. 697-701.
- [14] M. Bertozzi et al., "Obstacle Detection and Classification fusing Radar and Vision," *2008 IEEE Intell. Veh. Symposium*, Eindhoven, Netherlands, 2008, pp. 608 - 613.
- [15] Erwin Thomas Gilmore, Chukwunweike Ugbome, and Charles Kim, "An IR-based Pedestrian Detection System Implemented with Matlab-Equipped Laptop and Low-Cost Microcontroller," *Int. J. Comput. Sci. & Inf. Tech. (IJCSIT)*, vol. 56, no. 5, pp.79-87, Oct 2011.
- [16] A. Bartsch, F.Fitzek and R.H. Raschofer, "Pedestrian recognition using automotive radar," *14th Int. Radar Symposium*, Dresden, 2013, pp.732-739.
- [17] Yang Yang, Ren Mingwu and Yang Jingyu, "Obstacles and Pedestrian Detection on a Moving Vehicle," *Int. J. Adv Robot Syst*, in 2014 © Author. doi: [10.5772/58315].

- [18] Javier Hernández-Aceituno, Leopoldo Acosta, and José D. Piñeiro, "Pedestrian Detection in Crowded Environments through Bayesian Prediction of Sequential Probability Matrices," *J. of Sensors* in 2015 © 2016 Javier Hernández-Aceituno et al. doi: [10.1155/2016/4697260].
- [19] Xiaofeng Chen, Kristian Henrickson &Yinhai Wang, " Kinect-Based Pedestrian Detection for Crowded Scenes," *Comput.- Aided Civil and infrastructure Eng.*vol.31, no. 3, pp.229-240, 2015.
- [20] Tianyu Zuo, " An Efficient Vision-Based Pedestrian Detection and Tracking System for ITS Applications," M.S. thesis, Dept. Elect. & Comput. Eng., University of Ottawa, CA, 2014.
- [21] Nedeveschi. S, Bota.S, Tomiuc.C, "Stereo Based Pedestrian Detection for Collision Avoidance Applications", *IEEE Intelligent Transport System*. Vol.10,pp, 380 - 391, February 2009.
- [22] D. M. Gavrila, S. Munder, " Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle, *Int. J. Comput. Vision*, vol.73, no.1.pp. 41-59, 2007.
- [23] D.M. Gavrila, J. Giebel and S. Munder, " Vision-Based Pedestrian Detection: The PROTECTOR System, *IEEE Intell. Veh. Symposium*, Parma, Italy, 2004, pp. 13-18.
- [24] T. Ojala, M. Pietikainen, and D. Harwood." A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp.51–59, 1996.
- [25] Vrushali B. Ghule, S.S. Katariya, " Pedestrian Detection Technique's – A Review," *Communications on Applied Electronics -Foundation of Computer Science(FCS)*, vol. 3, no.6. pp.10-12, 2015

- [26] John Valentino, " A System for Scene Understanding and Classification of Objects in the Road Environment with Application to Autonomous Vehicles," M.S. thesis , Dept. Operations Research and Financial Eng., Princeton University.
- [27] Swaraj Preet Swain, Srilokanath Dalai, " Real time pedestrian detection and tracking for driver assistance systems," Dept. of Elect. Eng. National Institute of Technology, Rourkela, 2013.
- [28] Hanene Rouabeh, Chokri Abdelmoula, Mohamed Masmoudi, " VHDL based Hardware Architecture of a High Performance Image Edge Detection Algorithm," *Int. J. Comput. Applications*, vol.9, no.12, pp. 00975-8887, April 2014.
- [29] Rafael C. Gonzalez, Richard E. Woods, " Digital Image Processing," 3rd edition, Upper Saddle River, NJ, 2008, ch. 2, 3, 6, 9, pp. 46-48, 120-122, 322, 445,402,631-633.
- [30] CBS Interactive Inc, Samsung Galaxy S4 Specifications [Online]. Available: <http://www.cnet.com/products/samsung-galaxy-s4/specs/>
- [31] The Matlab Work, Inc., (1994-2016), MATLAB Major Update [Online]. Available: <http://www.mathworks.com/products/matlab/>
- [32] Zoltán Fegyver (23 Feb, 2014), Matlab tricks.com [Online]. Available: <http://matlabtricks.com/post-35/a-simple-image-segmentation-example-in-matlab---image%20segmentation>