SHIBAURA INSTITUTE OF TECHNOLOGY

Study on Object Detection using Computer Vision by Artificial Neural Network

by

Jittima Varagul

A dissertation submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Division of Functional Control System
Graduate School of Engineering and Science

September 2018

Knowledge is great because its a factor that originates wit and progress. Humans, therefore, never stop learning.

The Royal Guidance of Majesty King Bhumibol Adulyadej, 1981
The Great king of Thailand

To my parents who give me unwavering love and always believe in me.

To Varagul family for always encouraging me in all of my pursuits to follow my dreams.

To my friends for listening, offering me advice, and supporting me in everything I do.

To teachers who kindly advise and motivate me to accomplish my goal and fulfill my dream. . .. . .

Acknowledgements

Though only my name appears on the cover of this dissertation, many great people have supported to the contribution until this dissertation has been succeeded. I would like to express my sincere gratitude to all people who has made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever.

I have gained a professional experience and valuable lifeskills during my time at Shibaura Institute of Technology (SIT) and the experience opened my eyes to other possibilities. Definitely, this dissertation would not have been accomplished without the scholarship of Shibaura Institute of Technology and Japanese Government (Monbukagakusho or MEXT) Scholarship.

My deepest gratitude is to my supervisor, Prof. Toshio Ito. He has been warmly welcome and supported me since the first day I attended his laboratory. I have been amazingly fortunate to have a supervisor who gave me the freedom to learn and researched, and always helped me when I faced trouble. His enthusiasm and immense knowledge have been an inspiration and motivation to me. He has made me believe that if I try hard enough I can achieve whatever I want in life. In the future, I would follow his step and become a good supervisor to my students as he has shown to me.

I would like to extend my gratitude to my dissertation committees: Prof. Hiroyuki Arai, Prof. Masanobu Takahashi, Prof. Hiroshi Hasegawa, and Assoc. Prof. Toshiya Hirose, for precious comments, including the questions which intended me to widen my research from various perspectives. I am gratefully indebted to them for their very valuable comments on this dissertation.

Dozens of people have helped and encouraged me immensely. I am grateful to SIT friends and my friends in Suranaree University of Technology for their encouragement, many supports, and for all the great time we have had.

Finally, I must express my very profound gratitude to my parents and to my sister for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Thank you very much.

Jittima Varagul

SHIBAURA INSTITUTE OF TECHNOLOGY

Abstract

Graduate School of Engineering and Science
Division of Functional Control System

Doctor of Philosophy

by Jittima Varagul

A collision avoidance system, which is an automobile safety system designed to reduce the severity of a collision, namely to detect objects and obstacle avoidance. In the actual vehicle driving situation on the road, it is desirable to be able to recognize the preceding obstacles. In order to prevent a collision of the vehicles and the obstacles - of which we do not know the exact shape, size or color- it uses various sensors to detect the obstacles, such as optical sensors, RADAR, SONAR, LIDAR, and camera. The camera is a master of classification and texture interpretation, which has the lowest price. As a result, The author would like to reduce the cost of the detector and improve the performance of the vehicle by making the vehicle has the ability to see and recognize the obstacles like human beings by computer vision system.

Such real-time obstacle detection by computer vision was crucial in that we often found fake obstacles such as a text, sign, or painting on the road. Thus, one of the most important obstacle avoidance is the vehicle has to be able to detect, recognize and classify the obstacles that are real obstacles or fake obstacles. So far, have been developed using computer vision with ANN applied to obstacles recognition and classification, which ANN is mathematical model for a computer that can imitate the function of human brain. Hence, it can improve the performance of the vehicle has the ability to see and recognize like a human, which is an important task in an automotive safety application.

The main objectives of this research is to study on object detection using computer vision by ANN as well as to design algorithm to feature extraction, object recognition and classification, from the actual video images taken by an on-board camera. Due to this, TDNN was used in object classification and detection, where TDNN has the potential to work on sequential data.

In this dissertation, the author proposes a novel general object detection system by computer vision with ANN, which only one camera is used. Whereas the previous work, stereo vision is widely used in advanced driver assistance systems (ADAS). This is to reduce the number of cameras, but still be able to detect objects. To obtain the effective object detection, the author also introduce several systems: object analysis, feature extraction and identification based on histogram of oriented gradients (HOG) descriptor, object classification, and TDNN. After their processes were completed, this system can effectively detect general objects, which can classify the obstacles that are real obstacles or fake obstacles.

All systems had been tested and presented results, including new findings. The results showed that the performance of each system proposed in this dissertation was highly effective. It can detect general objects, and is not restricted to vehicles, objects or pedestrians - which most of the previous work was to detect specific types of objects such as pedestrians, cars, trucks, etc. It has provided good results along with high accuracy and reliability. Therefore, this system can be applied to provide a warning to the driver when there is an imminent collision in order to prevent an accident and reduce the severity of a collision.

In my future research, the author intend to concentrate on improving the systems to detect the obstacles in bad weather such as fog or rain by deep learning. Along with the principle of deep learning neural networks to be design the algorithm to improve the detection of objects. In particular, a detection of moving objects is available to realize safer path planning, a form of informative support for the driver. The author expect this research will lead to more control over vehicles to avoid oncoming obstacles automatically and efficiently.

Contents

List of Figures

List of Tables

Abbreviations

ADAS Advanced Driver Assistance Systems

AGV Automated Guide Vehicle

ANN Artificial Neural Network

BRISK Binary Robust Invariant Scalable Keypoints

CNN Convolutional Neural Networks

FREAK Fast Retina Keypoint

HOG Histogram of Oriented Gradients

K-NN k nearest neighbor

LIDAR Light Detection and Ranging

MLF-ANN Multi-Layer Feed-forward Artificial Neural Network

MLFANN Multi-Layer Feed-forward Artificial Neural Network

MLP Multi-Layer Perceptrons

RADAR Radio Detection and Ranging

RBF radial basis function

ROI Region of Interest

SIFT Scale Invariant Feature Transformation

SONAR Sound Navigation and Ranging

SURF Speeded-Up Robust Features

SVMs Support Vector Machines

TDNN Time Delay Neural Network

VP Vanishing Point

Chapter 1

Introduction

The aim of this chapter is to introduce background information, existing problem and research objective that lead to comprehend an importance of this dissertation, which entitled Study on object detection using computer vision by artificial neural network. The proposed solution and research methodology will be presented in this chapter. Finally, describe a structure of this dissertation.

1.1 Background

Currently, there are many accidents that occur on the road. The most common cause road traffic accidents is rear-end collisions and in several instances it is driver error such as fatigue, discomfort, use of a phone while driving, or senior driver's mistake. These accidents can be reduced if these driver errors are eliminated. Furthermore, autonomous vehicle or self-driving car stands for a new way forward in mobility, which every significant automakers pay attention and develop technology for autonomous vehicle enthusiastically. In recent years, there have been a lot of studies for a collision avoidance system, which is an automobile safety system designed to reduce the severity of a collision, namely to detect objects and obstacle avoidance.

Stereo vision is widely used in applications ADAS and robot navigation [6, 8] where stereo vision is used to estimate the actual distance or range of objects of interest from the camera. A stereo camera pair must have two identical cameras rigidly mounted so that they will not move with respect to each other. The two cameras on the front of vehicle take pictures of the same object with difference view. These two images contain some encrypted information about the depth of the object[28]. This information is the third dimension of the two images. Therefore the image distance and its depth can be determined by using the stereo cameras.

However, the aim of this dissertation is to reduce the number of cameras but still be able to detect objects, which one camera is used. Thus, I proposed the method of detecting objects using only one camera.

Object detection plays a major role which results in reducing the accidents. In the actual vehicle driving situation on the road, it is desirable to be able to recognize the preceding obstacles. In order to prevent a collision of the vehicles and the obstacles - of which we do not know the exact shape, size or color - it uses various sensors to detect the obstacles, such as optical sensors, radio detection and ranging (RADAR) [63], sound navigation and ranging (SONAR) [32, 38], light detection and ranging (LIDAR) [12], camera 4, and laser sensor. After the detection is done, these systems either provide a warning to the driver such as a flashing dashboard icon, a beep, a tug from the seat belt or braking autonomously without any driver input when a collision is imminent as show in Figure 1.1. However, every detection sensor has both advantages and disadvantages [42] as show in Figure 1.2(b).

The radar sensor system uses radio waves to determine the velocity, range and angle of an object. Radars can operate under all practical driving conditions such as rain, snow or fog. However, radar sensors suitable for the detection of large objects such as vehicles, may not suit small or narrow objects such as a pedestrian. In addition, the cost of the radar sensor is higher than a camera.

The lidar that measures distance using emitted light with high accuracy, which will work in every condition. Moreover, the minimum target size of lidar is a 1" square or larger. Therefore, it can detect objects of various sizes such as vehicles,

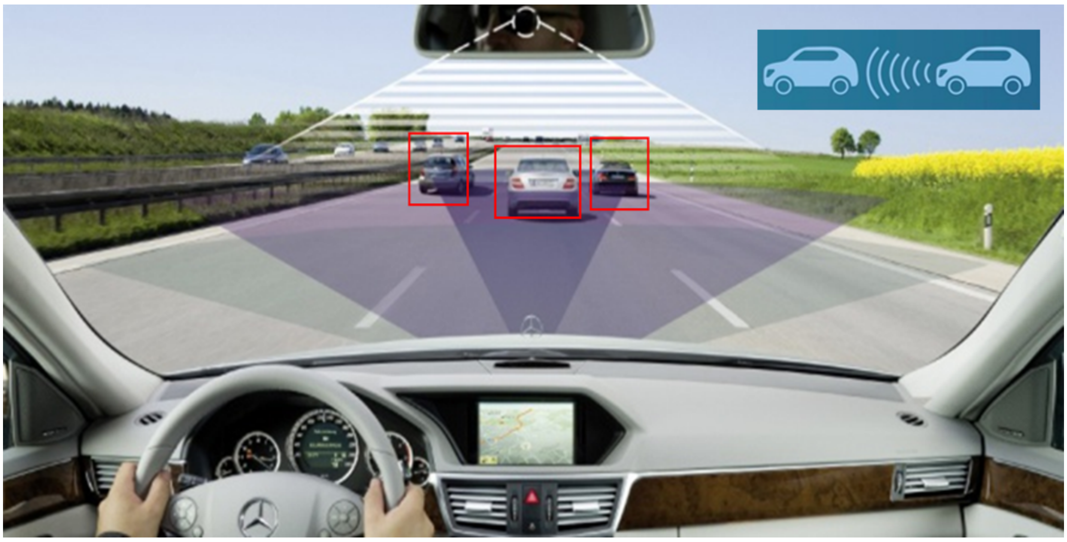


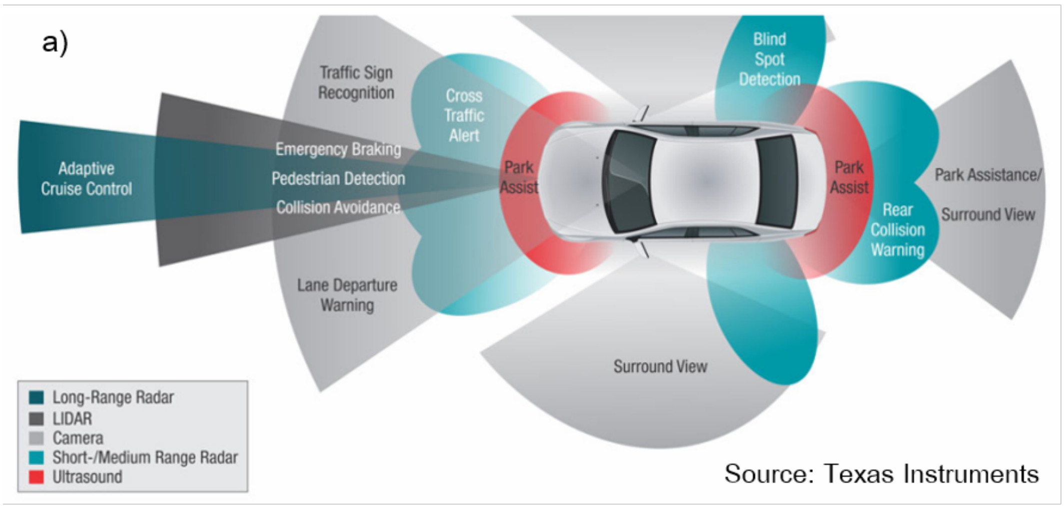
FIGURE 1.1: Preview object detection system for a collision avoidance system

motorcycles or bicycles. Nonetheless, due to lidar using light to detect objects, problems may arise when light is reflected from dark objects such as a black object, and they are still quite expensive.

The camera is a master of classification and texture interpretation, which has the cheapest and most widely available sensors of all three sensor types. Nowadays, low-price cameras with very high resolution are available. Furthermore, even a low-price camera has a resolution higher than lidar. It can detect every object, shape, size, and color. Therefore, it is able to understand things that cant be learned from lower-resolution lidar and radar.

Figure 1.2(a) demonstrates the sensing area of each sensor and the distance used to control the safety function of autonomous car, which the camera can be used to detect obstacles in a wide range. Besides, Figure 1.2(b) demonstrates sensing technology comparison-the key point and interesting is that the camera has the lowest cost.

As a result, the author would like to reduce the cost of the detector and improve the performance of the vehicle by making the vehicle has the ability to see and recognize the obstacles like human beings by computer vision system.



b) Rating: H = High, M = Medium, L = Low

Function	Camera	Radar	Lidar
Object Detection	M	H	H
Classification	H	M	L
Lane Detection	H	L	L
Traffic Sign Recognition	H	L	L
Work in Rain, Fog, Snow	L	H	M
Work in Low Light	L	H	H
Work in Bright Light	M	H	H
Size	Small	Small	Medium
Cost	$	$$	$$$$

FIGURE 1.2: Properties of object detection sensor: (a) demonstrates the sensing area of each sensor and the distance used to control the safety function of autonomous car, (b) demonstrates sensing technology comparison

1.2 Object Detection by Computer Vision with Artificial Neural Networks

Currently, computer vision, either alone or combine with other technologies whether it be sonar, radar, or lidar, is one of the key technologies of advanced driver assistance systems (ADAS). Many studies have been developed using computer vision with Artificial Neural Network (ANN) applied to obstacles recognition and classification. For example, Karthik B, et al. [29] proposed a computer vision system with neural networks to identify the traffic sign by using HOG based SVMs to extract feature of traffic sign. Then, the neural networks are used to train the traffic sign pattern and classify the 16 different traffic sign image. The result showed that 98 percent accuracy has been achieved, which it is high efficiency. Furthermore, also many work use computer vision with an ANN for the object recognition and classification such as the vehicle detection and classification in traffic [13, 20] , pedestrian detection [60, 72], vehicle license plate recognition [1, 30].

The computer vision with ANN is mathematical for a computer that can imitate the function of human brain. Hence, it can improve the performance of the vehicle has the ability to see and recognize like a human, which is an important task in an automotive safety application.

1.3 Problem Statement

For the objects that can be found on the road with several of sizes, shapes and colors. In particular, a detection of moving objects (pedestrians, cars, bicycles, etc.) as show in Figure 1.3. Such real-time obstacle detection by computer vision was crucial in that we often found fake obstacles such as a text, sign, or painting on the road as show in Figure 1.4.

For human eyes, if we look at the object from front view, we will see the real obstacle is a high object and a non-high object is a fake obstacle. Nevertheless, the camera takes pictures in front looking directly at the object from front view as show

FIGURE 1.3: The moving objects on the road

FIGURE 1.4: The fake obstacles

in Figure 1.5, even if the object is a three-dimensional object, but the camera images can be seen, it is a two-dimensional image based on the principles of an orthographic projection 45. An orthographic projection is a two-dimensional representation of a three-dimensional object, which two-dimensional drawing represents different sides of an object as in Figure 1.6.

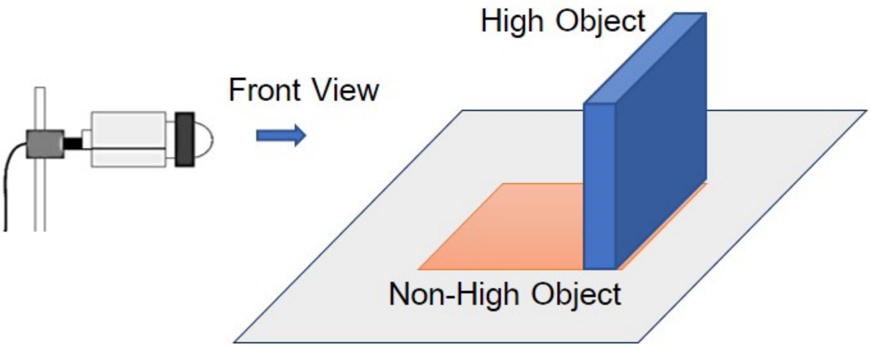


FIGURE 1.5: Preview image of the camera perspective

As mentioned above, real-time obstacle detection by computer vision may be a mistake in detecting fake obstacles becoming real obstacles. Therefore, an algorithm for obstacle detection in this dissertation need to recognize and extract the difference

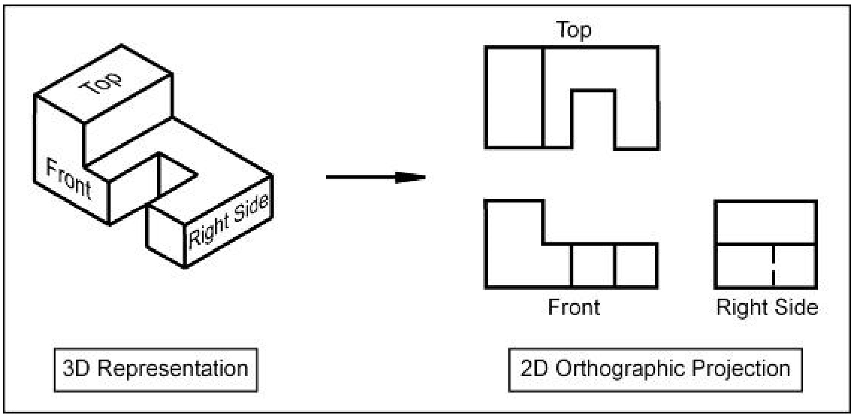


FIGURE 1.6: Preview image from the front view by an orthographic projection

of specific feature of the obstacles along with appropriate methods to learning and classify whether the object detected is a real obstacle or fake obstacle.

1.4 Objectives

The major purpose of this dissertation is to introduce a novel object detection for vehicles integrating several techniques that the author proposed in the past, such as object feature extraction, object recognition, and object classification. Such real-time obstacle detection by computer vision was crucial in that we often found fake obstacles. Thus, the aim of this dissertation is to improve general object detection method for the vehicle using computer vision with ANN, which can classify the obstacles that are real obstacle or fake obstacle. In this dissertation, the proposed technique is based on HOG to extract features of the objects and classify the obstacles by the ANN for several situations.

1.5 Outline of the dissertation

In this dissertation, the author focus on general object detection by computer vision. Note that object detection in computer vision is widely used in ADAS and many researchs have used computer vision to develop collision avoidance systems, which is one of importance task of an automobile safety system, namely to detect objects and obstacle avoidance. The object that the author is interested in this research is the generic object, which is not limited to vehicles, pedestrians, bicycles or motorcycles.

This dissertation reports the findings of a thorough study and introduces novel methods, including their evaluations. The experimental results have proven that the methods have been effective and succeeded in completing a goal of this research. The author herein proposed the algorithms to detect object in front of the vehicle in both indoor and outdoor environments.

The first, analyzed and identified the characteristics of the object in the various cases that are viewed in the camera view. The image seen from the camera's perspective is a two-dimensional image, distinct from the three-dimensional vision of the human eye. This distinctive feature of each object-type can be used to classify objects that are detectable as real or fake objects.

Second, the extraction method extracts the characteristics of each type of object to learn and remember the characteristics of the object. This approach focuses only on the objects in the image, thus reducing the role of the background and making the object stand out so that the desired result, shape and ratio are needed.

Finally, classification of objects from individual characteristics. This method using several techniques to distinguish object types based on their characteristics discovered in feature extraction to learn the difference between a height object and non-height object.

Hence, my entire studies are assembled into one main system, called object detection system for general obstacle detection, from the actual video images taken by an on-board camera. The proposed method, the author extracted feature of the obstacles to classify the obstacles that are real obstacles or fake obstacles such as a painting, sign or text on the road. The experimental results showed that this system can detect general objects, and is not restricted to vehicles, objects or pedestrians. It has provided good results along with high accuracy and reliability, which it is accurate enough to provide a warning to the driver when a collision is imminent.

1.6 Novelty of the Study

There are a lot of existing studies focusing on object detection for vehicles, as similar to this study - which most of the previous work was to detect specific types of objects such as pedestrians, cars, trucks, etc. However, this study focuses on the general object detection regardless of the size or color of the object, and is not restricted to vehicles, objects or pedestrians.

The novelty of the study is the general object detection system by computer vision with ANN, from the actual video images taken by an on-board monocular camera. The novel general object detection is based on HOG to extract features of the objects and classify the obstacles by TDNN. This system can effectively detect general objects with several sizes, shapes, and colors, which can classify the obstacles that are real obstacles or fake obstacles with high accuracy and reliability.

1.7 Scope of study

Study focuses on general object detection for vehicles by using computer vision with ANN. The key for detecting objects with the camera is the lighting conditions. As a result, places and environments directly affect the object detection. Besides, the obstacle that may be found in each case are different. Under this condition, the author decided to demonstrate my proposed method with the experiment. The author conducted three case as follows:

1. Object detection for AGV in indoor scene.
2. Object detection for senior car on the footpaths or a senior car lane.
3. Object detection for vehicle in traffic.

1.8 Structure of This Dissertation

The remainder of the dissertation is organized into 6 chapters. The rest of the dissertation is organized as follows;

The next chapter discusses previous researches related to the topics of this dissertation. From Chapter 3 to Chapter 6, the author describe methodologies of each proposed study respectively: Object-types analysis, object feature extraction, object recognition and classification, and prototype of object detection system. Chapter 7 presents experiment procedures and results. Chapter 8 is devoted to discussing new findings of each study and their limitations. Finally, the author summarize the dissertation and draw conclusions, including future works. Figure 1.7 presents an overall content structure of the dissertation

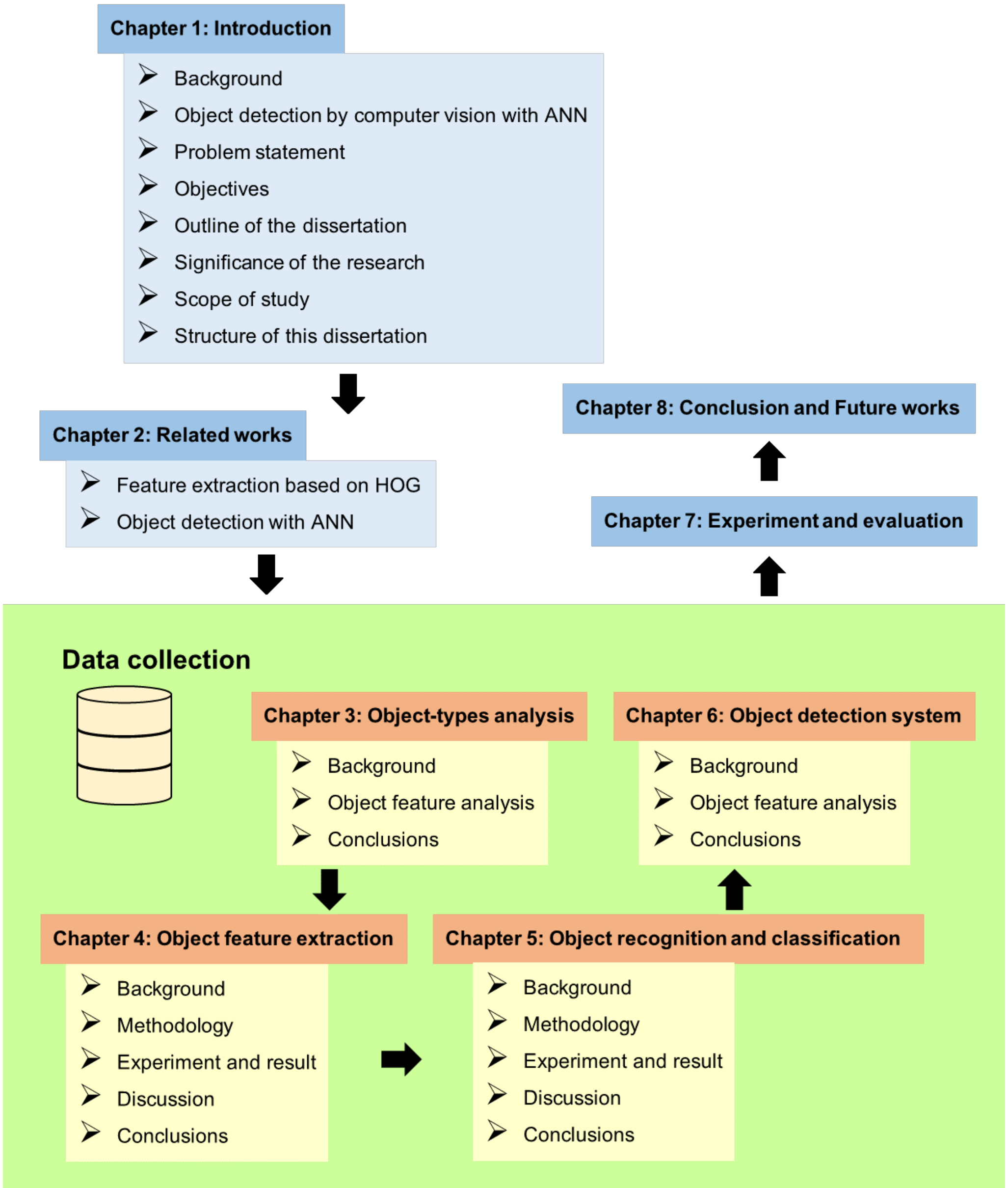


FIGURE 1.7: Content structure of this dissertation

Chapter 2

Related works

In the previous chapter, the author has explained the background of existing problems and the research objectives. This chapter surveys existing works related to the topic of this dissertation. This chapter surveys the works and theory that related to the topic of this dissertation.

When considering the actual vehicle driving situation on the road, it is desirable to be able to recognize the preceding obstacles in order to avoid accident. Thus, one of the most important obstacle avoidance is the vehicle has to be able to detect, recognize and classify the obstacles that are real obstacles or fake obstacles. So far, many works have been developed using neural networks for image analysis applied to obstacles recognition and classification, which is an important task in an automotive safety application. In the following, the author will briefly describe some recent approaches for the object recognition and classification. From the aim of this dissertation is to improve general object detection method for the vehicle using computer vision with ANN, which can classify the obstacles that are real obstacle or fake obstacle. Therefore, an algorithm for obstacle detection in this dissertation need to recognize and extract the difference of specific feature of the obstacles along with appropriate methods to recognize and classify the obstacles. This chapter is organized into two sections: feature extraction, and object recognition and classification.

2.1 Object feature extraction method

In various computer vision applications widely used is the process of finding the local features and extracted from the images refer to a pattern or distinct structure found in an image, such as a point, edge, orientation, or small image patch. These systems called feature extraction [16, 51] have received intensive attention in the literature of object detection and classification, tracking, and motion estimation. Consequently, a broad range of techniques has been proposed, such as BRISK [33], FREAK [25], SIFT [35, 47], SURF [40], HOG [14] and optical flow [24, 48, 49].

The feature extraction method depends on considering the criteria of application and the nature of data. For example, binary descriptors, such as BRISK or FREAK, which are encoded into binary vectors based on the different local intensity. These descriptors are suitable used for finding point correspondences between images, which are used for registration. In contrast, the SIFT, SURF, optical flow and HOG descriptors are suitable for recognition, classification, and detection tasks.

SIFT is an algorithm for detecting distinctive invariant image features which provide robust matching between different two images and are invariant to image scale and rotation. SIFT has four main steps consist of scale-space extrema detection, keypoint localization, orientation computation and keypoint descriptor. This method provides highly effective results in matching images in different views of an image, which is a popular method for studying to improve the efficiency of object recognition [59] or classification [22], but it is suffered with speed. For example, Ramisa [47] evaluated and also proposed several modifications to the original schema to improve the SIFT for object recognition method in a mobile robotics setting.

SURF is one of local feature detector and descriptor, which is inspired by SIFT descriptor. The characteristics of SURF are fast interest point detection, distinctive interest point description, speedup descriptor matching, and invariant to common image transformations: image rotation, scale changes, illumination change, or small change in viewpoint. The computation speed of SURF faster than SIFT [27, 52], but comparable performance.

Optical flow is descriptor that use for motion detection algorithms. The main objective of motion detection algorithms is to mark video frame regions that contain motion and also estimate the direction and magnitude of movement. Optical flow is a motion pattern of the object in the scene caused by relative motion between the scene and the eye (a camera). This method is used for specific fields, such as motion-based object detection, image segmentation to tracking and predict the motion vector of a moving object by motion estimation. For example, moving object detection and tracking based on optical flow [3], this system proposed motion object detection and tracking in image sequences from stationary camera images in traffic surveillance. It used optical flow to detecting motion vectors, the vector magnitudes threshold are used to segment objects from the background in urban videos. However, this system is done for detection and tracking in image sequences from stationary camera. Besides, Hariyono et al. [24] proposed a method for pedestrian detection from a moving vehicle by using optical flows to segment moving object regions. Then, HOG features is used to object recognition and classified using linear support vector machine (SVM). Moreover, because the calculating the optical flow is computationally expensive [61], the optical flow descriptors are much more expensive to extract than the HOG descriptors.

HOG feature descriptor counts occurrences of gradient orientation in localized portions of an image. This descriptor similar to SIFT and SURF, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy and its advantages in detecting edge and texture information of image. The HOG descriptor has more important advantages over other descriptors. Since it operates on local cells, it is invariant to geometric and photometric transformations, except for object orientation.

From the author mention above, SIFT and SURF compute the gradient histogram only for patches around specific interest points, but HOG is computed for an entire by divides the image into small cell and summing up the orientation of gradient over every pixel within each cell in an image. Although the computation is concerned they are similar, but they have different applications. SIFT and SURF suitable for identification of specific objects in object recognition (e.g. in a Bag-of-features fashion). In this dissertation, the object that has to detect is an unknown object - of which we do not know the exact shape, size or color. HOG is typically used in a sliding window fashion in object detection systems (e.g. pedestrian detection). Thus, HOG is suitable to use to extract feature of an object in this problem.

In this dissertation, the target object of detection is not an exact shape, size or color. One method that is highly effective in the feature extraction of the object in the image is the HOG method. Basically, HOG used the feature of shape regardless of the size or color of the object. It counts occurrences of gradient orientation in part of an image hence it is an appearance descriptor [14], which it is the most commonly use method to find an edge. HOG features are now widely used in object recognition and detection [58].

Many works for pedestrian detection system uses the Histograms of Oriented Gradient (HOG) method [14, 58, 70]. For example, Zhang et al. [70] proposed a method base on HOG classifier for pedestrian detection to extract the specific features of the human in infrared images by the distribution of intensity gradients or edge directions regardless of the size or color of the human. Then, linear support vector machines (SVMs) [5, 70] is used as a classifier for pedestrian detection. The system achieved is up to 99%. This is a comparison and recognition process which can be applied to separate and classify the features of pedestrians efficiently.

HOG descriptors were invoked in different applications other than human detection which it was original designed for as follows: There are many traffic signs recognition and classification used HOG descriptor to extract feature of the object in images. [21, 54]. Fleyeh et al. [21] proposed the method that used HOG descriptor approach for traffic signs classification and classified the traffic signs by a Gentle AdaBoost classifier with high accuracy was 99.42%.

Vehicles recognition, classification, and detection is an important task for automotive safety system. Many works developed for this system, which the vehicle feature extraction based on HOG-combined with several classifiers [10, 53]. For example, Rybski et al. [53] proposed a vision-based algorithms for coarse vehicle orientation classification by use HOG to extract the different orientations of vehicles in images and used orientation-specific classifiers for the classification. The result of the classification was stabilized at 88% accuracy.

Moreover, also many works use HOG descriptors as features extraction for object recognition, classification, or detection such as the hand shape classification [66], and the handwritten digit recognition [15].

2.2 Object detection with Artificial Neural Networks (ANN)

One of the most important obstacle avoidance is the vehicle has to be able to recognize and classify the obstacles. So far, many recognition and classification methods have proposed. Several works have been developed using neural networks for image analysis applied to obstacles recognition and classification [11, 29], which is an important task in an automotive safety application.

Quiles et al. [46] proposed the objects detection and recognition of mobile robot, which is implemented by using sonar to detect obstacle. In the image segmentation process is performed by a color classification method based on MLP Neural Networks [9, 43, 46]. The results of the classification is a defined color, separate from other colors of the image. For example, when the red object is required, every pixel of the image is classified as red or non-red. But this system can only be used to track objects of a specific color.

Anumula et al., [2] developed an automatic road sign recognition can detect and classify the traffic signs from complex scene by using MLF-ANN with three layers, which classified into six signs by learning from 500 images, gathers a wide range of the six signs. The MLF-ANN is used to label the signs. This method has the flexibility to specify the meaning of the traffic signs accurately in different lighting conditions, including different angle of views as well.

For the objects that can be found on the road with several of sizes, shapes and colors-in particular, a detection of moving objects (pedestrians, cars, bicycles, etc.). Detection using single frame images may be faulty. Many studies have improved the detection method by using multi-frame images instead single frame[50, 57, 69].

For example, Zhan [69] used multi-frame images instead single frame to improve moving object detection algorithms. This system used the difference of the object edge between the two images by using Canny detector to moving object. The result has a high recognition rate and a high detection speed, but the problem of this system is false detecting under more complicated background.

The Time Delay Neural Network (TDNN) has the potential of learning to overcome the limitations of a multi-layer neural network, and complete image sequences at a time instead of a single image, which it can work with complex data efficiently. Several works for object recognition use TDNN [39, 64, 65]. For example, a pedestrian classification based on the typical criss-cross motion pattern of a pedestrians legs in sequences of gray-scale stereo images taken from a moving camera pair. The recognition is stabilized by feedback loops added to the feed-forward TDNN architecture.

Chapter 3

Object-types analysis

As mentioned before, object detection, which is an important task in an automotive safety application. In the previous chapter, the author mentioned related works and their addressed problems on object classification and detection. Many state-of-art studies attempted to propose solutions to solve existing obstacles effectively.

In this dissertation, the author also introduced novel methods to detect general object that is the obstacle of the vehicle. My methodology will be described in Chapter 3 to 6 divided by my proposed systems in chapter 1. Chapter 3 presents ideas for analyzing specific features of the objects from a camera perspective and identify the ROI of detection by conditions in each case. Next, object feature extraction and identification will describe in Chapter 4. This method can be can be utilized to identify and extract specific feature of shape of both real obstacle and fake obstacle. Chapter 5 will depict object recognition and classification using several techniques to distinguish object types based on their characteristics discovered in feature extraction. Finally, the author will introduce a prototype of object detection system that integrates entire proposed systems.

3.1 Background

In the actual vehicle driving situation on the road, it is desirable to be able to recognize the preceding obstacles-of which we do not know the exact shape, size or color-in order to prevent severity of a collision.

The work that the author reviewed in the previous chapter, these systems are classified to identify and detect specific objects in the image such as pedestrian detection, traffic signs recognition, vehicle license plate recognition. Nevertheless, this dissertation proposed novel methods to detect general object, and is not restricted to vehicles, objects or pedestrians. Such real-time obstacle detection by computer vision was crucial in that we often found fake obstacles such as a text, sign, or painting on the road.

This study focuses on general object detection for vehicles by using computer vision. The key for detecting objects with the camera is the lighting conditions. As a result, places and environments directly affect the object detection. Besides, the obstacle that may be found in each case are different. Under this condition, the author decide to study in three cases. First is object detection for AGV in indoor scene. Second, electric senior vehicle in a senior car lane. Finally, object detection for vehicle in traffic.

3.1.1 The object detection for AGV in indoor scene

Currently, the AGV is a transport vehicle widely used in manufacturing industry. This system can also reduce labor costs in industry, as well as increasing the safety of its employees from working in dangerous environments as in Figure 3.1. It can move materials and equipment to various locations that work automatically according to conditions, and the directions to destination are defined, but it cannot avoid obstacles in front by itself. When considering the actual AGV driving situation on the path, it is desirable to be able to recognize the preceding obstacles. In order to prevent a collision the AGVs and the obstacles-which most objects might be found in factories are simple shape as in Figure 3.2(a). Such real-time obstacle detection was crucial in that we often found fake obstacles such as flat tiles, papers, or painting on the floor as in Figure 3.2(b).

FIGURE 3.1: Automated Guided Vehicle (AGV)

FIGURE 3.2: The obstacle images in factory environment (indoor scene): (a) real obstacle; (b) fake obstacle

3.1.2 The object detection for electric senior vehicle in a senior car lane

Nowadays, in many countries, the proportion of elderly people is steadily increasing. According to the statistical handbook of Japan by Statistics Bureau (2017), the average life expectancy in Japan climbed sharply after World War II, and in 2015, the average life expectancy at birth was 87.1 years for women and 80.8 years for men. Similarly, in 2016, the elderly population (aged 65 or over) has increased from 1950 constituting 22.4 percent of the total population. Furthermore, the elderly population also projection to 37.7 percent in 2050 as in Figure 3.3.

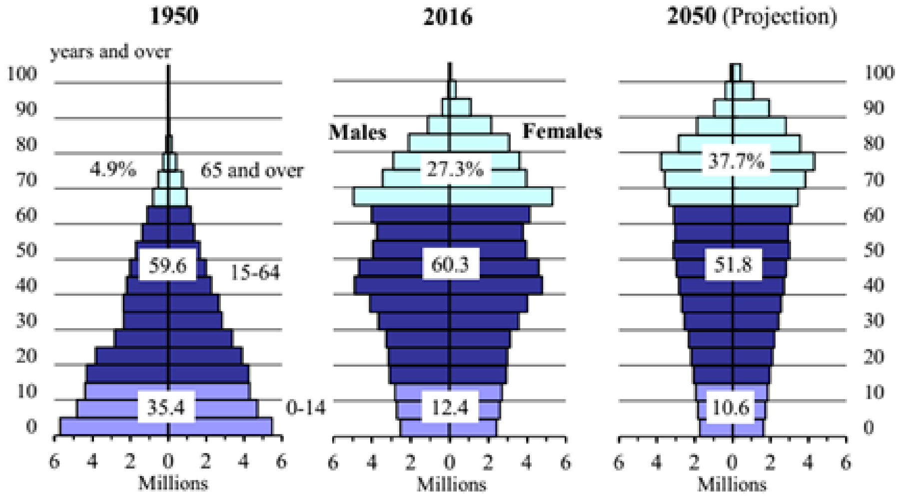


FIGURE 3.3: Changes in the Japan Population Pyramid, Source: Statistics Bureau, MIC; Ministry of Health, Labour and Welfare

The electric senior vehicle widely used to support the elderly in daily life. It is often referred to as a mobility scooter or a senior car as well. It is intended to drive pedestrians on footpaths equivalent to a wheelchair, but configured like a motor-scooter. It can help the elderly and the disabled regain their freedom, confidence and independence inside and outside the home. Moreover, it is convenient and safety for elderly to go around their neighbors such as a supermarket or a convenience store as in Figure 3.4.

Normally, the electric senior vehicle is operated in a bike lane and side walk in case to leave or enter the property. The object that is the most common objects that can be found on the sidewalk and a bike lane in Japan are not exact shape-in particular, pedestrians and bicycles as in Figure 3.5(a). Moreover, the fake obstacles that often found in real-time obstacle detection in this situation, such as painting, sign or text on the street as in Figure 3.5(b).

FIGURE 3.4: Preview of electric senior vehicle usage

a)

b)

FIGURE 3.5: The obstacle images on the sidewalk and a bike lane: (a) real obstacle; (b) fake obstacle

3.1.3 The object detection for vehicle in traffic

Currently, there are many accidents that occur on the road. The most common cause is rear-end collisions and in several instances it is driver error such as fatigue, discomfort, drowsiness, use of a phone while driving, or various reckless driving. These accidents can be reduced if these driver errors are eliminated. In recent years, there have been a lot of studies for a collision avoidance system, which is an automobile safety system designed to reduce the severity of a collision, namely to detect objects and obstacle avoidance. In the actual vehicle driving situation on the road, it is desirable to be able to recognize the preceding obstacles. In order to prevent a collision of the vehicles and the obstacles. For the real objects that can be found on the road with several of sizes, shapes and colors. In particular, a detection of moving objects (pedestrians, cars, bicycles, etc.) as show in Figure 3.6(a). Moreover, there are various symbols and messages on the road, which may lead to misunderstandings as real objects as show in Figure 3.6(b).

FIGURE 3.6: The obstacle images on traffic: (a) real obstacle; (b) fake obstacle

In this dissertation, detecting objects is a detection regardless of the object color. Hence, it is the detection based on the object characteristic analysis.

3.2 Object feature analysis

As the author mentioned above, the objects that can be found in each cases in several sizes, shapes or colors, such real-time obstacle detection by computer vision was crucial we often found fake obstacles. Identifying whether the object is a real or fake object needs to know the different characteristics of each type. Then, it uses the different characteristics of object to identify whether the object being detected is a real or fake object.

In general, the camera will take pictures in front looking directly at the object in the front view as in Figure 3.7. However, even if the object is a three-dimensional object, but the camera images can be seen, it is a two-dimensional image based on the principles of an orthographic projection. An orthographic projection is a two-dimensional representation of a three-dimensional object, which two-dimensional drawing represents different s ides of an object. We can use this principle to analyze the specific features of the obstacles that need to detect.

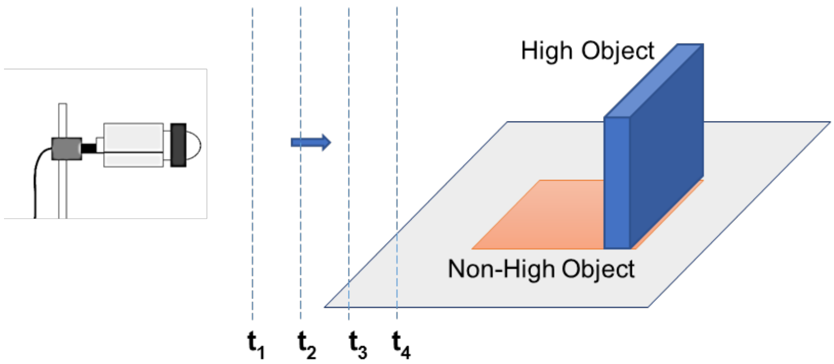


FIGURE 3.7: Preview image of the camera perspective at various distances

In case of the object as a simple shape, when the camera shoots a real object from the front view, the height of the object is vertical line based on the principles of an orthographic projection as the author mentioned in chapter 1. In contrast, the fake object image is no border in the vertical line based on construction of perspective viewing, showing the formation of a VP [31], which is the image of the point at infinity as show in Figure 3.8.

From the object projection to VP show that the angle of the object edge from the camera view is different as shown Figure 3.9. From this result, these different edge features can be used to classify objects.

For the analysis of the edges of objects from object projection to VP to use vertical line detection to identify a real or fake object, it may work well only with simple shapes. If the object is more complex shape, it will be necessary to find other features to analyze together.

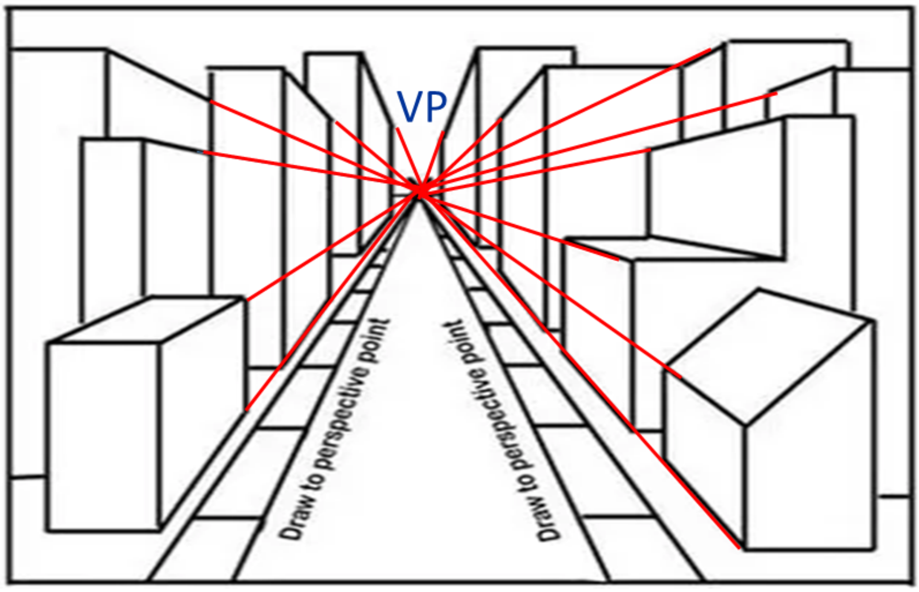


FIGURE 3.8: A 2D construction of perspective viewing, showing the formation of a vanishing point

When the vehicle moves closer to the height object, as long t1, t2, t3 as in 3.7[C.4, C.5], though the size of the object has changed, the shape of the object has not changed. However, when the vehicle moves closer to the non-height object, the size and shape of the object have changed as in 3.10. From changes in the shape of the edge of the object can be analyzed to classify the object, which can be illustrated by comparing the difference of the orientation of the edges in each frame. One method that is highly effective in the feature extraction of the object in the image is the HOG method. Basically, HOG used the features of shape regardless of the size or color of the object in the image. HOG counts occurrences of gradient orientation in part of an image hence it is an appearance descriptor. Thus, this HOG method is best suited to solve this problem.

3.3 Discussion

The aim of this dissertation is to detect the obstacles in front of a vehicle. Real time object detection by computer vision was crucial in that fake obstacles may be

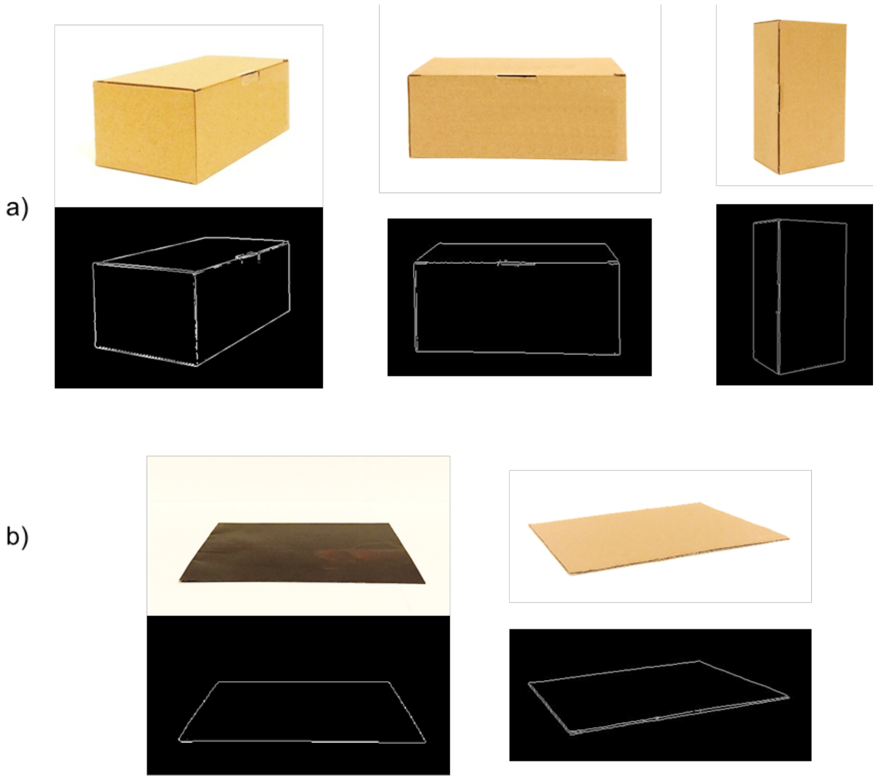


FIGURE 3.9: The examples of the edge detection: (a) the edge detection of the high objects; (b) the edge detection of the non-high objects

found, such as text or symbols on the road. Therefore, this algorithm needs to recognize the difference of specific features of the obstacles, where the real obstacle is a high object and the fake obstacle is a non-high object.

Based on the result of analysis of the characteristics of each object, the object feature that can be used to classify the obstacles that are real obstacles or fake obstacles as follows:

1. Comparison of the characteristics of the object edge base on orthographic projection (METHOD 1). From orthographic projection and construction of perspective viewing, the height of the real object is vertical line and the fake object image is no border in the vertical line.

FIGURE 3.10: A comparison of changes in the shape of an object at different distances; (a) the high object; (b) the non-high object

2. Comparison of the difference of the edge orientation in each frame (METHOD 2), which the difference of the orientation of the edges the real object is very small compared to the fake object.

3. Comparison of the shape variation ratio by calculating the ratio between the width and height of the object (METHOD 3), which the fake object has a shape variation ratio over the real object.

In conclusion, characteristics of objects used in the proposed methods presented in this dissertation-it is based on the comparison of learning outcomes of the object feature in three cases as follows:

1. Feature recognition by METHOD 1
2. Feature recognition by METHOD 2
3. Feature recognition by METHOD 3
4. Feature recognition by METHOD 2 and METHOD 3

Chapter 4

Object feature extraction

In the previous chapter, the author introduced the object-type analysis based on characteristics of objects such as orientation of edge, the difference orientation of edge in each frame, and shape variation ratio. My focus of this dissertation was to extract feature of the object with several sizes, shape, and color-to classify the object is a real object or fake object.

In this chapter, the author describe object feature extraction and identification. This method can be can be used to identify and extract feature of shape of both real obstacle and fake obstacle.

I will begin by explaining a background of the study and then move to seeing how the method works. the author conducted several experiments to evaluate results to prove a validation of the method. Findings will be discussed in a discussion section in this chapter. Finally, the author present a conclusion.

4.1 Background

In this dissertation, the target object of detection is not an exact shape, size or color. Before learning and recognizing an object process, in order to compare the differences between objects, both real objects and fake objects, it is necessary to extract the dominant features of the objects. Thus, the method used should be appropriate for the analysis and extraction of the image features, regardless of the color or size of the object in the image. From the results of the analysis of object types in the previous chapter.

From the results of the analysis of object types in the previous chapter, the object feature to be extracted to classify the obstacles that are real obstacles or fake obstacles as follows:

1. The object edge feature extraction
2. Comparison of the difference of the edge orientation in sequence of time
3. The shape variation ratio by calculating the ratio between the width and height of the object.

4.2 Methodology

As the author mentioned in chapter 2, the three feature extraction that popular in object recognition, classification, and detection such as SIFT, SURF and HOG method. Although the computation is concerned they are similar, but they have different applications. SIFT and SURF descriptor compute the gradient histogram only for patches around specific interest points, which it is highly effective in identification of specific objects. Nevertheless, the object that has to detect in this dissertation is an unknown object - of which we do not know the exact shape, size or color. HOG is computed for an entire by divides the image into small cell and summing up the orientation of gradient over every pixel within each cell in an image. Besides, HOG is typically used in a sliding window fashion in object detection systems (e.g. pedestrian detection). Thus, HOG is suitable to use to extract feature of an object in this problem.

The HOG method was first proposed by Dalal & Triggs [14], devised as a method to be used to detect humans. Basically, HOG used the features of shape regardless of the size or color of the object in the image. HOG counts occurrences of gradient orientation in part of an image hence it is an appearance descriptor. It is the most commonly used method to find an edge. It then divides the image into sub-images (block). Heres how the image is divided into two types as in the Figure 4.1(a) Regtangle-HOG type (R-HOG) and in the Figure 4.1(b) is Circle-HOG type (C-HOG).

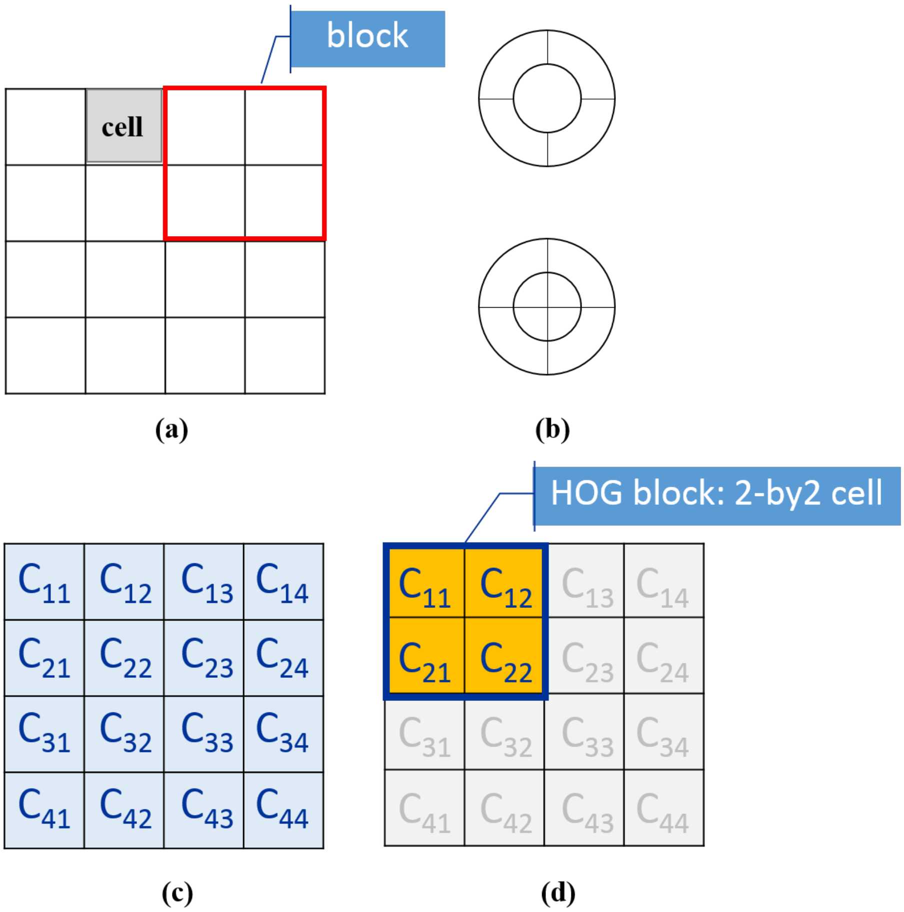


FIGURE 4.1: The structure of cell division: (a) R-HOG type; (b) C-HOG type; (c) example of an image with sixteen cells; (d) example of setting block size 2-by-2 cells

Each block is divided into small cells as in Figure 4.1(a), and each cell will contain the orientation of gradient, which is stored in the form of a histogram.

For example, Figure 4.1(c) shows an image with sixteen cells. If the block size is 2-by-2 cells, hence each block consists of four cells as show in Figure 4.1(d).

The computation of the gradient values can be calculated by using 1D - discrete derivative masks in both the horizontal and vertical directions. This method requires filtering the grayscale image with the following filter kernels by equation 4.1-4.2.

$$D_x = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \tag{4.1}$$

$$D_y = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T \tag{4.2}$$

Hence, being given an image I, we obtain the x and y derivatives using a convolution operation by equation 4.3-4.4.

$$I_x = I * D_x \tag{4.3}$$

$$I_y = I * D_y \tag{4.4}$$

Each pixel in the cell will have the magnitude and orientation similar to equation 4.5-4.6.

$$|G| = \sqrt{{I_x}^2 + {I_y}^2} \tag{4.5}$$

$$\theta(x, y) = \arctan(I_x/I_y) \tag{4.6}$$

$\theta(x, y)$ = the angle of the gradient vector coordinates (x, y).

The key parameter is bin, which is evenly spread over 0 to 180 or 0 to 360 degrees, depending on whether the gradient is unsigned or signed. The final feature vector includes all of the block in 1D matrix form.

Find the sum of each direction by equation 4.7 and calculated feature of each block by equation 4.8.

$$C_b = \sum_{i=1}^{n} \theta(x, y)_i \tag{4.7}$$

$$\upsilon_k = \sum_{i=1}^{n} (|G(x, y)_{k=i}| * C_b)|\theta(x, y)_{k=i} \tag{4.8}$$

where

n = number of position (x, y) for each block.

b = direction

C_b = a summary of each direction

The feature include all of block by equation 4.9.

$$\upsilon_k = \begin{bmatrix} \upsilon_{k=1} \\ \upsilon_{k=2} \\ \vdots \\ \upsilon_{k=k} \end{bmatrix} = [] \tag{4.9}$$

where k = the order of the block.

To optimize accuracy, the histograms have been normalized for releasing the calculation of the indicators and the intensity of overlap of the cells within the block to reduce the impact of the illumination and contrast variation by block normalization. The block normalizations are explored in four different methods for block normalization by Dalal and Triggs [14]. Let $||\upsilon||_k$ be the non-normalized vector containing all histograms in a given block, be its k-norm for k = 1, 2 and e be constant.

Then the normalization factor can be calculated by one of the following as equation 4.10-4.12). Optimized HOG feature extracting flow chart as shown in Figure 4.2.

L2-norm:

$$f = \frac{v}{\sqrt{||v||_2^2 + e^2}} \tag{4.10}$$

L2-hys: L2-norm followed by clipping (limiting the maximum values of v to 0.2) and renormalizing.

L1-norm:

$$f = \frac{v}{(||v||_1 + e)} \tag{4.11}$$

L1-sqrt:

$$f = \sqrt{\frac{v}{||v||_1 + e}} \tag{4.12}$$

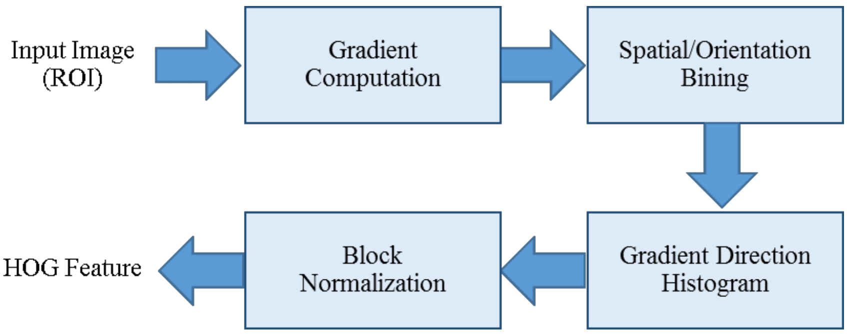


FIGURE 4.2: HOG feature extracting flow chart

I proposed a HOG method to extract feature of the obstacles, which is highly effective in the feature extraction of the object in the image. Therefore, the author used the HOG method to extract features of the obstacles which can detect objects and shapes within an image by analyzing the distribution of the intensity gradient and edge direction, and then explain the image in a histogram. The proposed method is separated into two parts, i.e., pre-processing and feature extraction.

4.2.1 Pre-processing

The pre-processing process aims to prepare the data for the next stage. The output of this stage would be ready for the next stage to perform complicated image processing tasks on the data. Prior to extracting features, training and testing a classifier. On the basic of HOG to use the features of shape regardless of size or color of the object in the image. So, a pre-processing step is image enhancement, the input images of our system are initially converted to binary image, applied to remove noise for highlight certain features of interest in the images, cropped to ROI. This provides better feature vectors for training the classifier. The most important thing is to detect and treat the edges of the object in the image, which the edge detector is Canny detector [23, 36]. The simple pre-processing flowchart is given as in Figure 4.3.

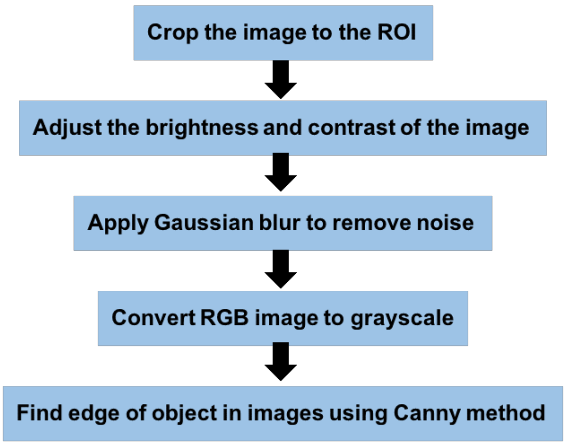


FIGURE 4.3: Pre-processing Flow Chart

4.2.2 Feature extraction

The feature extraction is separated into three parts, i.e., the object edge feature extraction, the edge orientation in each frame, and the shape variation ratio. Basically, HOG used the feature of shape regardless of the size or color of the object. It counts occurrences of gradient orientation in part of an image hence it is an appearance descriptor [14], which it is the most commonly use method to find an edge. HOG features are now widely used in object recognition and detection such as pedestrian detection system, traffic signs recognition and classification, Vehicles recognition, classification, and detection. With these properties, the author used HOG to extract the feature of the object that the author described in chapter 2 as shown in Figure 4.4.

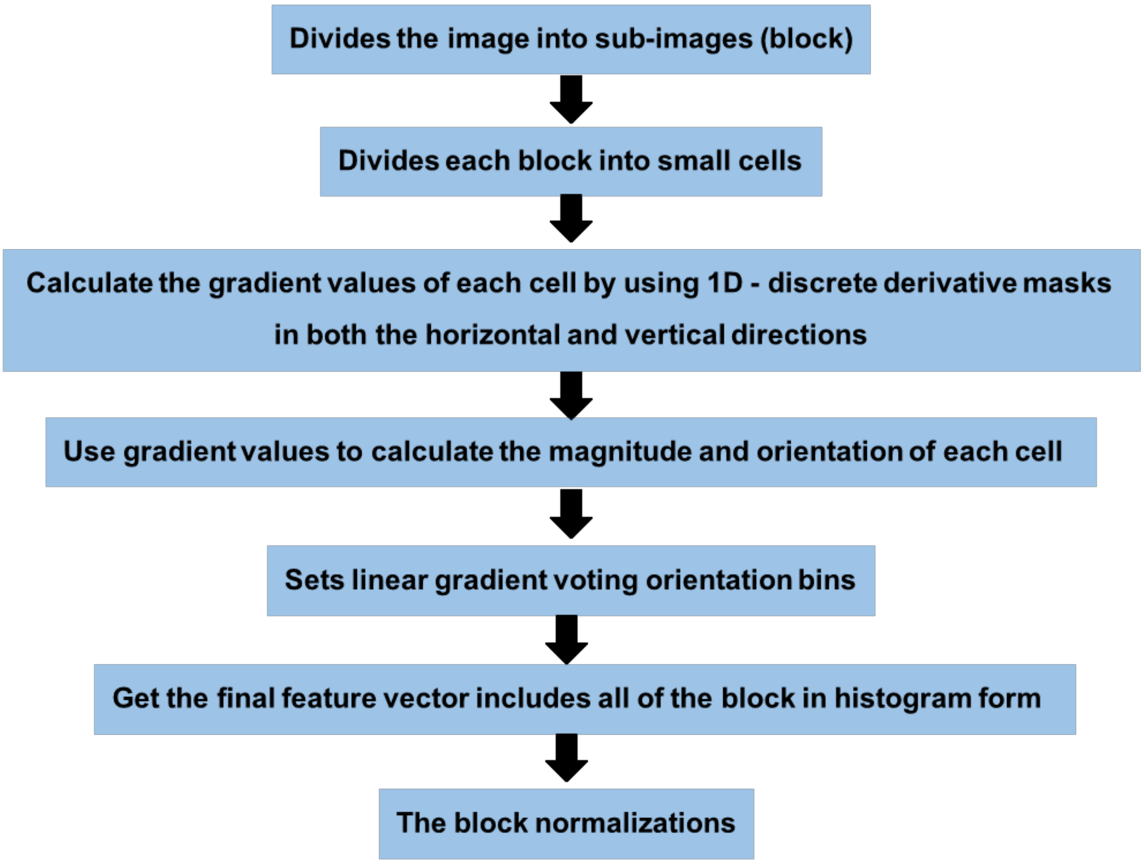


FIGURE 4.4: The HOG method flowchart

4.3 Experiment and results

I conducted an experiment to evaluate whether the proposed method can extract feature suitable for each cases. The experiment is separated into four parts, i.e., pre-processing, the object edge feature extraction, the extraction of the difference of the edge orientation in each frame, and extract shape variation of the object.

4.3.1 Pre-processing

For this experiment, the author collected the samples of various obstacle that include sample of real obstacles and sample of fake obstacles, which taken from the front view by an on-board camera and downscale to 256x256 size as shown in Figure 4.5. Then, enhance image quality follow the process in flowchart as show in Figure 4.4.

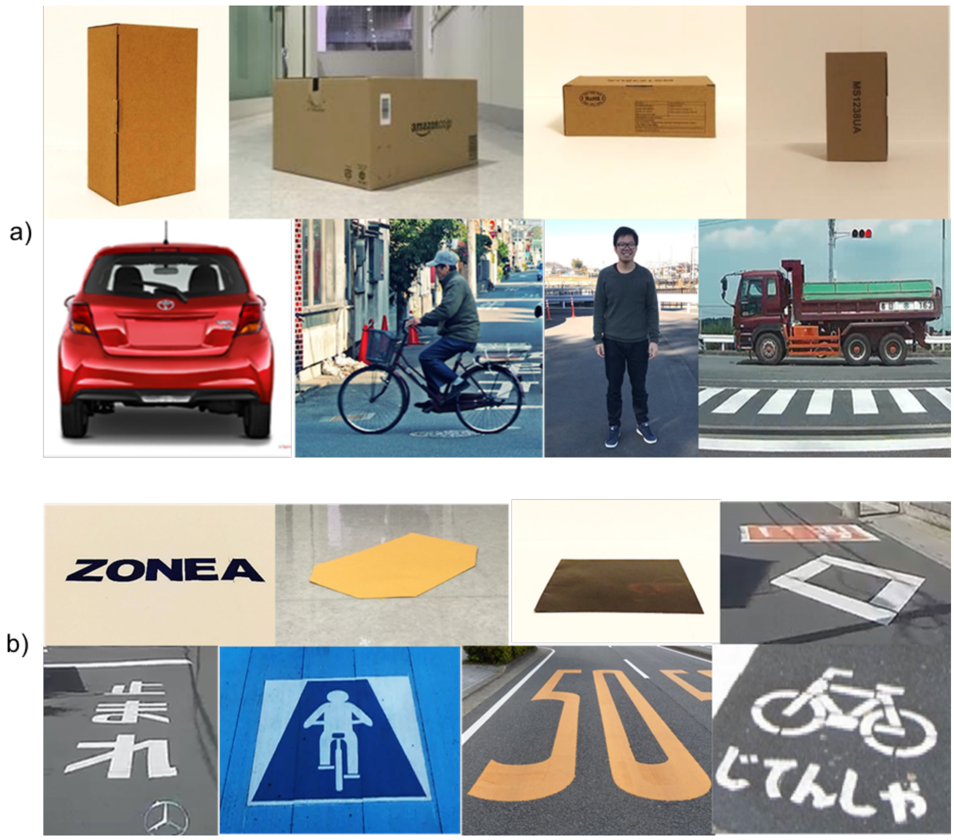


FIGURE 4.5: Example of samples of various obstacle: (a) the sample of real obstacles; (b) the sample of fake obstacles

The results of this experiment can detect the edges of the object in every shape, height and orientation of the height objects and the non-height objects as in Figure 4.6.

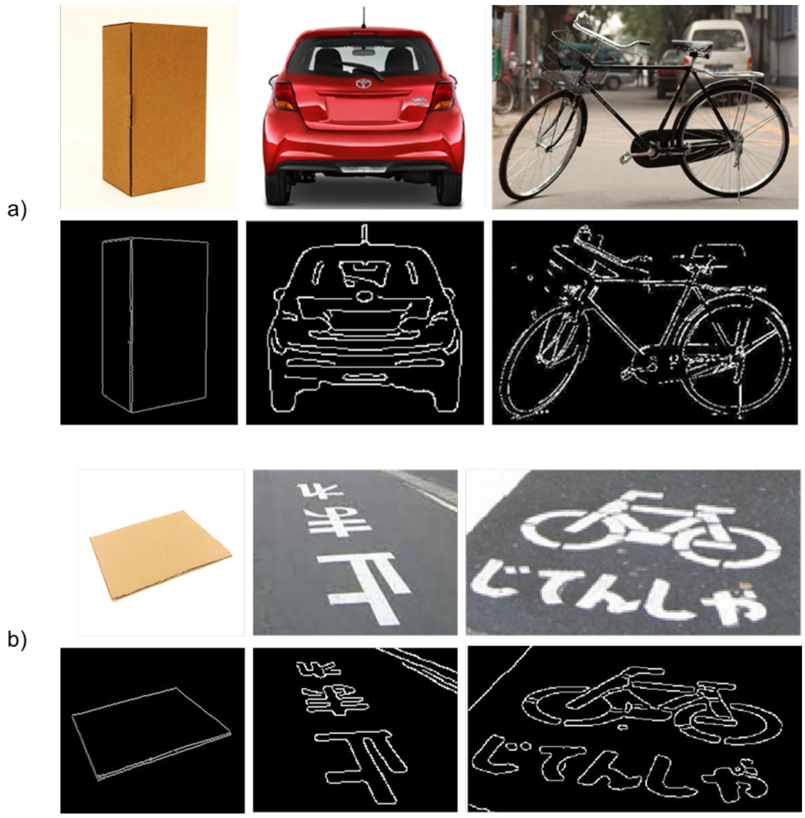


FIGURE 4.6: The edge detection of the object: (a) high object, (b) non-high object

4.3.2 The object edge feature extraction (METHOD 1)

HOG can calculate the orientation of the border in the form of a histogram. This method can analyze the distribution of the orientation in the form of a histogram. It can classify the obstacle by learning the orientation of object edges, which the real object-the height of the object is arranged in a vertical line. In contrast, the fake object image is no border in the vertical line. This experiment was conducted to investigate the feasibility of using object edge feature extraction to recognize and classify the object that are real object or fake object. The experiment is separated into two parts, i.e., object edge feature extraction for simple shape and object edge feature extraction for complex shape.

4.3.2.1 Object edge feature extraction for simple shape

As the author mentioned before, in case of the AGV, the object is a simple shape [C.1, C.2]. In case of the object as a simple shape, when the camera shoots a real object from the front view, the height of the object is vertical line based on the principles of an orthographic projection as the author mentioned in chapter 1. In contrast, the fake object image is no border in the vertical line based on construction of perspective viewing.

It is important to make sure the HOG feature vector encodes the right amount of information about the object. the author has tested the effect the cell size parameter has on the amount of shape data encoded in the feature vector. By varying the HOG cell size parameter and visualizing the result as show in Figure 4.7.

The HOG visualization plot shows that a cell size of 24-by-24 does not encode much shape information every edges, while a cell size of 16-by-16 encodes a lot of shape information but increases the dimensionality of the HOG feature vector significantly. A good compromise is a 20-by-20 cell size. This size setting encodes enough spatial information to visually identify a digit shape while limiting the number of dimensions in the HOG feature vector, which helps speed up training.

Therefore, the author created an experiment to analyze the characteristics of both types of obstacles by using the HOG feature method. This experiment sets linear gradient voting into 9 orientation bins in 0 to 180 degrees. Then it divides the image into sub-images by 2-by-2 blocks and 20-by-20 pixel cells. From the results, the feature length of each image is 4356.

The result of edge feature extraction for AGV, a common features of many images of height objects is the northward vectors, aligned as a vertical line to narrow the scope of the object as in Figure 4.8(a) and the non-height objects do not have the northward vectors, aligned as a vertical line as in Figure 4.8(b). Moreover, the

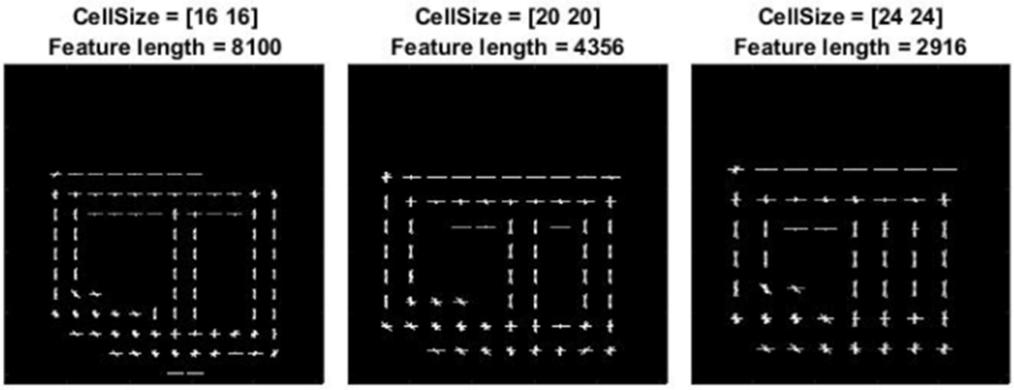


FIGURE 4.7: The effect the cell size parameter has on the amount of shape data encoded in the feature vector for simple shape

orientation of gradient values pattern of each object are different as show in Figure 4.9. In this graph, show the example the comparison of the orientation of gradient values from HOG feature between high object and non-high object consist of three sample pairs. The difference of the orientation of gradient values pattern of each object has the same tendency and the first to fifth bin, the number of histogram feature of high object much more than non-high object.

Thus, the distinguishing feature of the real obstacles that the author interested to taken into consideration is the orientation of the edges of the objects aligned as a vertical line from HOG features.

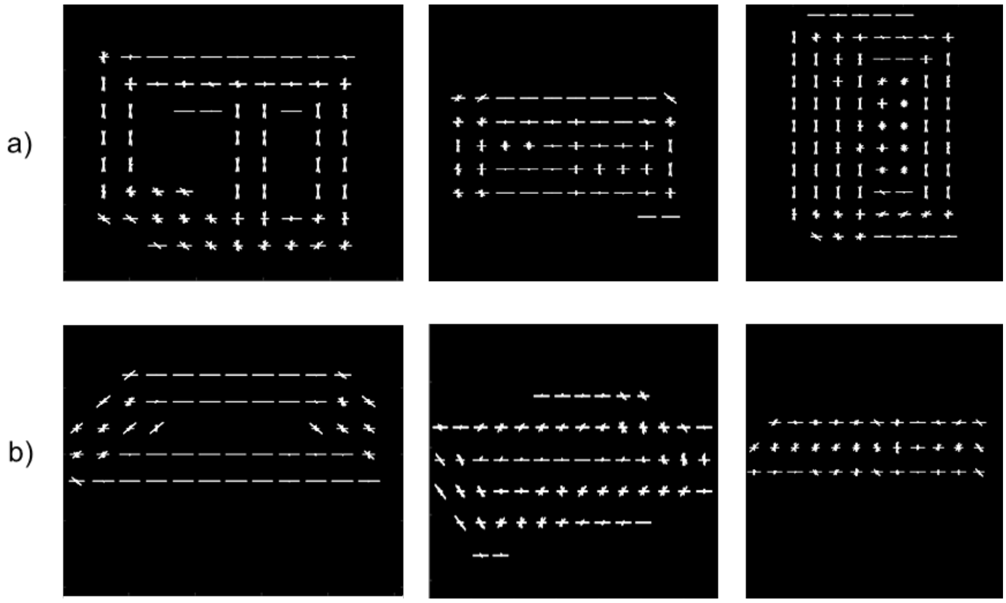


FIGURE 4.8: HOG feature visualization for simple shape: (a) high object; (b) non-high object

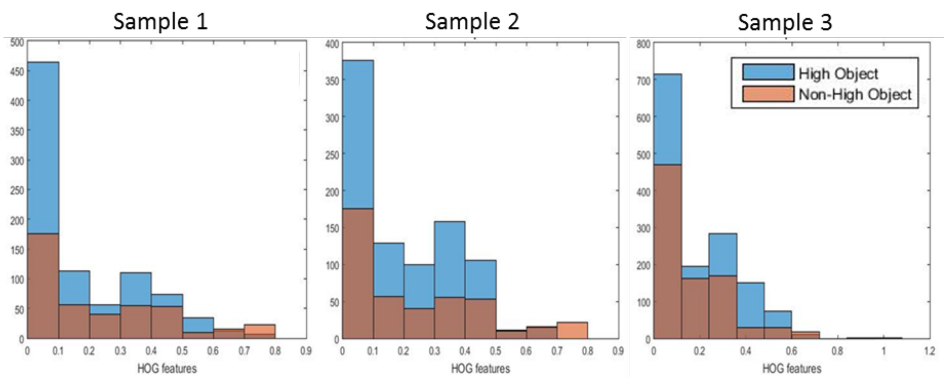


FIGURE 4.9: Illustration of a comparison of the orientation of gradient values from HOG feature between high object and non-high object for simple shape

4.3.2.2 Object edge feature extraction for complex shape

In case of senior vehicle and vehicle in traffic, the object found would be a complex shape. The effect the cell size parameter test has on the amount of shape data encoded in the feature vector. By varying the HOG cell size parameter and visualizing the result as show in Figure 4.10.

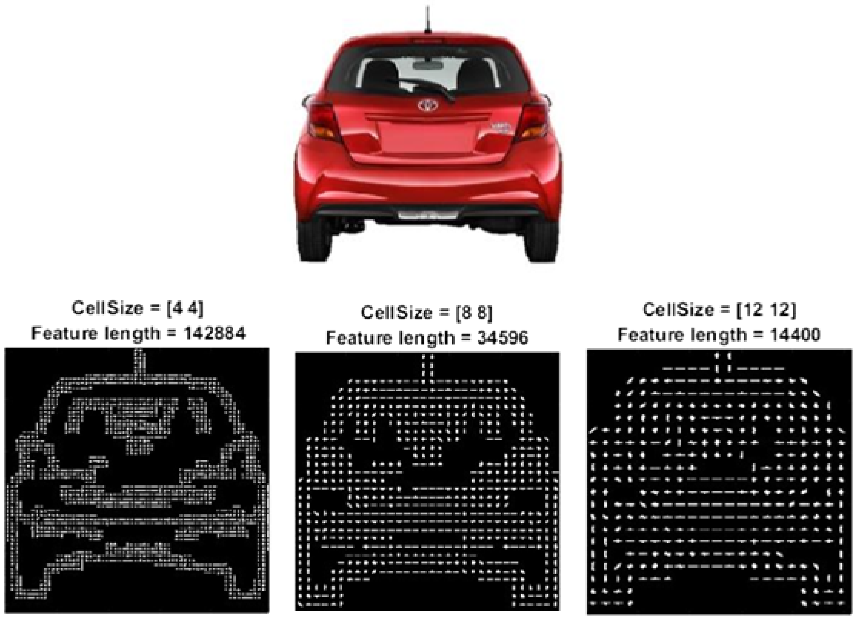


FIGURE 4.10: The effect the cell size parameter has on the amount of shape data encoded in the feature vector for complex shape

The HOG visualization plot shows that a cell size of 12-by-12 does not encode much shape information every edges, while a cell size of 4-by-4 encodes a lot of shape information but increases the dimensionality of the HOG feature vector significantly. A good compromise is a 8-by-8 cell size. This size setting encodes enough spatial information to visually identify a digit shape while limiting the number of dimensions in the HOG feature vector, which helps speed up training.

Therefore, the author created an experiment to analyze the characteristics of both types of obstacles by using the HOG feature method. This experiment sets linear gradient voting into 9 orientation bins in 0 to 180 degrees. Then it divides the image into sub-images by 2-by-2 blocks and 8-by-8 pixel cells. From the results, the feature length of each image is 34596.

The result of edge feature extraction for the complex shape as show in Figure 4.11. Figure 4.12, show the difference of the orientation of gradient values pattern of each object has the same tendency and the first to third bin, the number of histogram feature of high object less than non-high object, other bins are unstable.

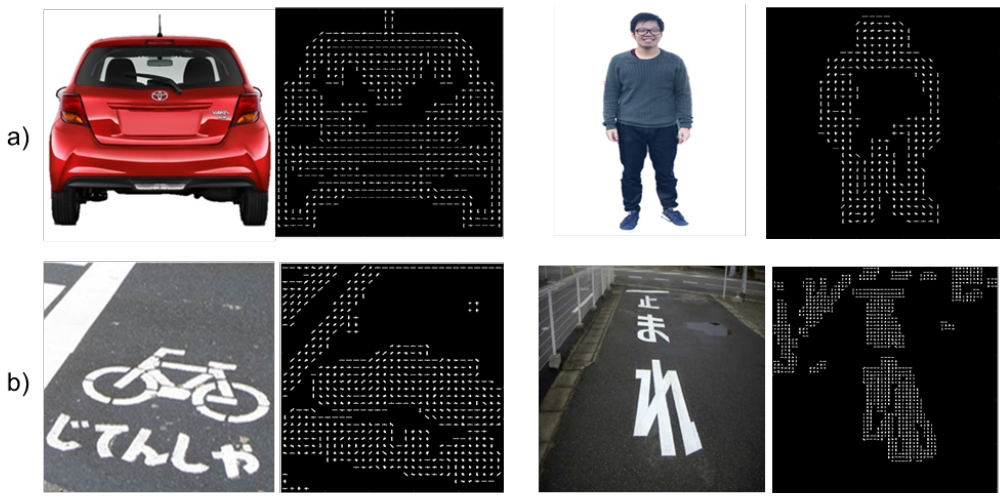


FIGURE 4.11: HOG feature visualization for complex shape: (a) high object; (b) non-high object

From the result of the object edge feature extraction for the simple shape and complex shape, the features of the edge of the object are distinctly different in case of simple shape. For the complex shape, the features of the edge are different as well, but not distinct when compared with simple shape. This method may be error under more complicated shape. If the object is more complex shape, it will be necessary to find other features to analyze together.

4.3.3 The extraction of the difference of the edge orientation in each frame (METHOD 2)

As mentioned above, when the vehicle moves closer to the high object, though the size of the object has changed, the shape of the object has not changed. However,

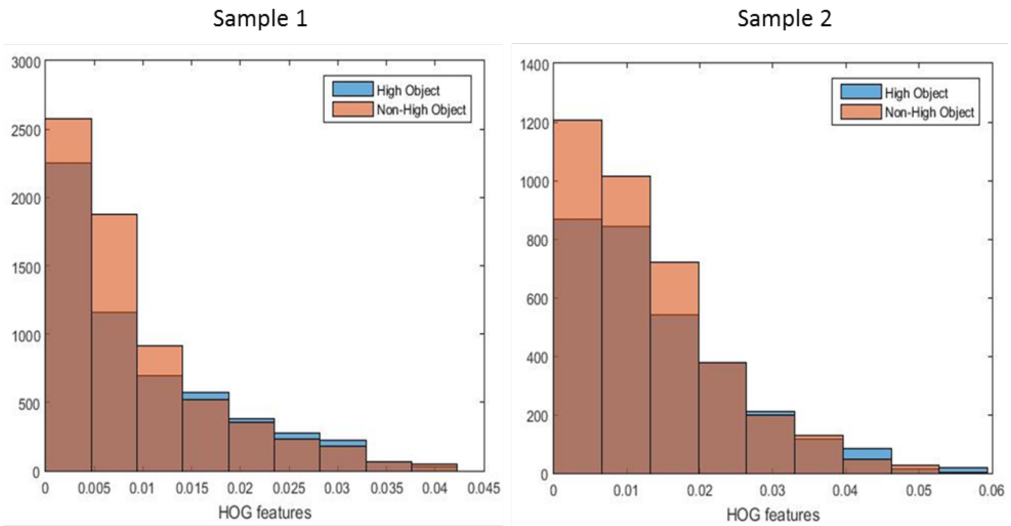


FIGURE 4.12: Illustration of a comparison of the orientation of gradient values from HOG feature between high object and non-high object for complex shape

when the vehicle moves closer to the non-high object, the size and shape of the object have changed. the author is concerned about the difference of the edge orientation in each frame of obstacle. Hence, we can also use this relationship to learn the difference between high and non-high objects.

This experiment was conducted to investigate the feasibility of using extraction of the difference of the edge orientation in each frame to recognize and classify the object that are real object or fake object. For the feature extraction used HOG descriptor, which setting parameter same as METHOD 1. Then, gauge the difference between the HOG features in different frame by computing the square error between them.

The result of the extraction of the difference of the edge orientation in each frame as show in Figure reffig:4-11. From this result show that when the vehicle moves closer to the high object, HOG features in different frame is slightly change as shown in Figure 4.13(a). However, HOG features in different frame is greatly changed when the vehicle moves closer to the non-high object as shown in Figure 4.13(b). Figure 4.13(c) show the square error to compare the difference between the HOG features in different frame of each object. The square error is smaller when images have similar edge orientation. Thus, HOG features in different time of the high-object has a little bit change when the vehicle moves closer to the object. In contrast, the non-high-object has higher difference of HOG features in different time.

This comparison of the difference of the edge orientation in each frame (METHOD 2) can use to object classification, which the difference of the orientation of the edges the real object is very small compared to the fake object.

4.3.4 The shape variation ratio (METHOD 3)

From the previous chapter, when the vehicle moves closer to the object, the real obstacle has a low shape variation ratio. In contrast, the fake obstacles has a high shape variation ratio. This experiment was conducted to investigate the feasibility of compare the ratio between the width and height of the object for identification of objects in object recognition [C.7].

Computation of shape variation ratio, the beginning segments the 2-D grayscale image into object and background regions using active contour [7] based segmentation based on Chan-Vese method. The output image bw is a binary image where the foreground is white (logical true) and the background is black (logical false). Mask is a binary image that specifies the initial state of the active contour. The boundaries of the object region(s) (white) in mask define the initial contour position used for contour evolution to segment the image.

Then, measure width and height of the object form the region of active contour result and compute the shape variation ratio calculating the ratio between the width and height of the object. The result of comparison of the shape variation ratio as show in Figure 4.14, which the object segmentation show in Figure 4.14(a). Moreover, the author can compare this variation by calculating the ratio between the width and height of the object. The real obstacle has a low shape variation ratio. In contrast, the fake obstacles has a high shape variation ratio as shown in Figure 4.14(b). As a result, the author can take pattern of the shape variation of the obstacle to recognize the obstacles.

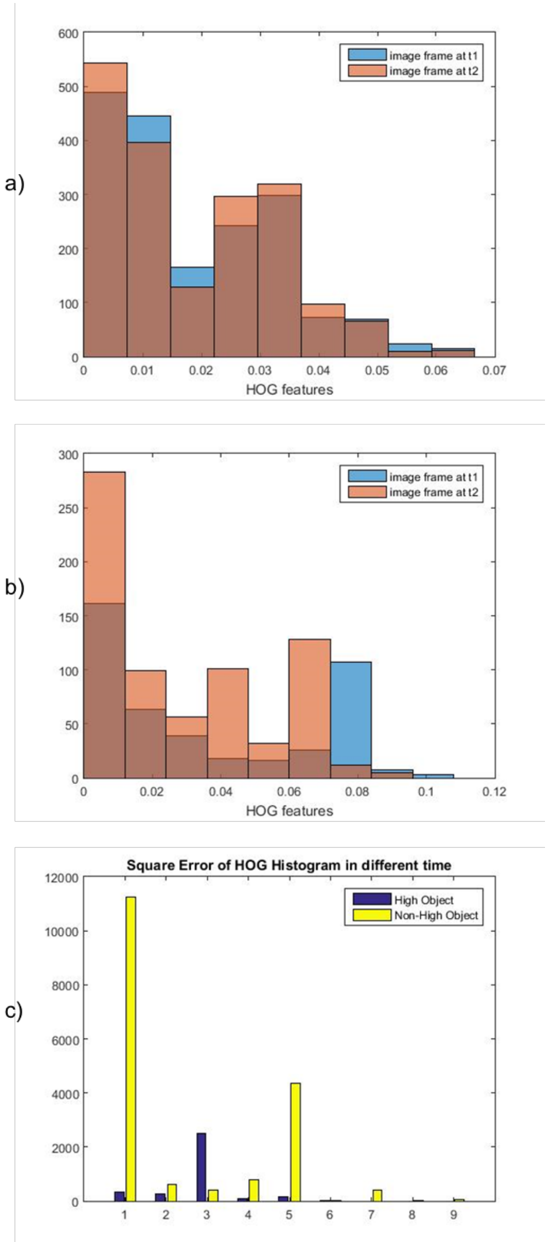


FIGURE 4.13: The result of the extraction of the difference of the edge orientation in each frame: (a) comparison of HOG feature of high object; (b) comparison of HOG feature of non-high object; (c) comparison of square error of HOG feature in different time

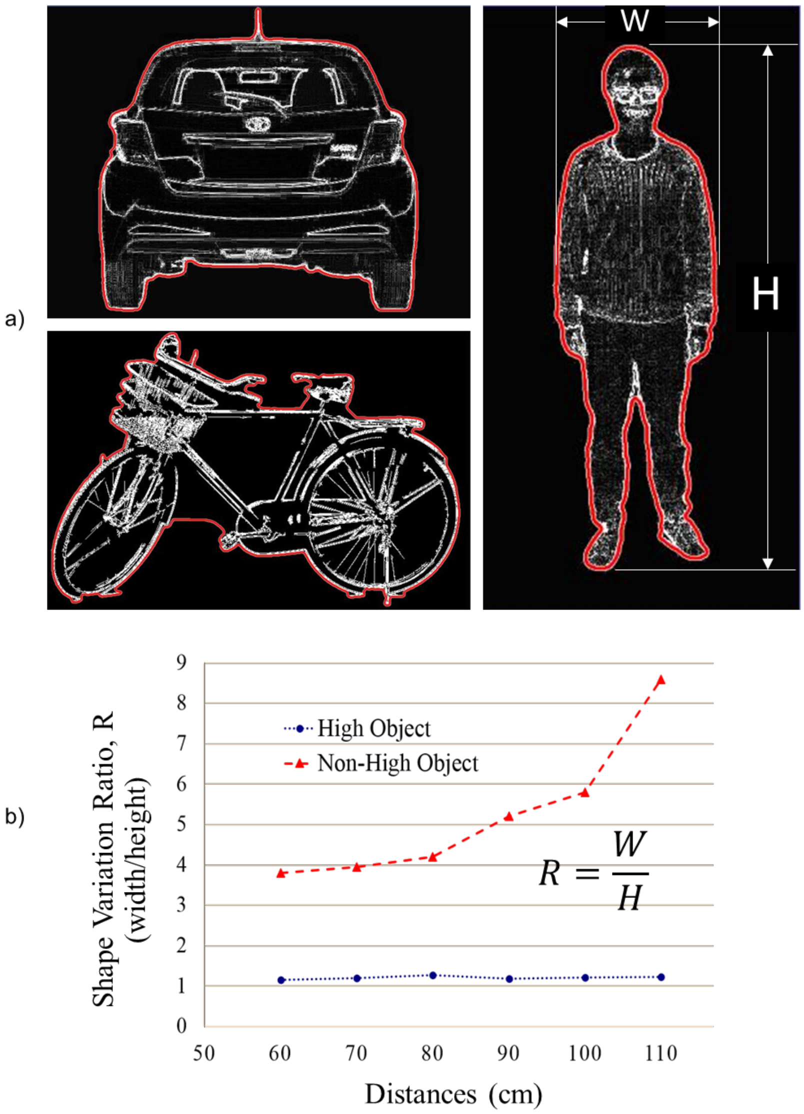


FIGURE 4.14: The result of comparison of the shape variation ratio (a) object segmentation by active contour; (b) shape variation ratio of the obstacles

4.4 Discussion

This experiment was conducted to investigate the feasibility of using various feature extraction to recognize and classify the objects that are real object or fake object. The experiment is separated into two parts based on shape, i.e., feature extraction for simple shape and complex shape. Moreover, three parts based on object feature, i.e., the object edge feature extraction, the edge orientation in each frame, and the shape variation ratio. For object segmentation using active contour based on Chan-Vese method. In addition, HOG is used to extract feature of the object to classify the objects that are real object or fake object.

From the result of the METHOD 1, the difference of object edge feature from HOG feature between high object and non-high object is quite distinctly different under simple shape object. However, the difference of object edge feature is lightly different for complex shape object. Thus, METHOD 1 may be error under more complicated shape. If the object is more complex shape, it will be necessary to find other features to analyze together.

METHOD 2, the result show that show that when the vehicle moves closer to the high object, HOG features in different frame is slightly change as shown in Figure 4.13(a). However, HOG features in different frame is greatly changed when the vehicle moves closer to the non-high object as shown in Figure 4.13(b). Figure 4.13(c) show the square error to compare the difference between the HOG features in different frame of each object. The square error is smaller when images have similar edge orientation. Thus, HOG features in different time of the high-object has a little bit change when the vehicle moves closer to the object. In contrast, the non-high-object has higher difference of HOG features in different time.

This comparison of the difference of the edge orientation in each frame (METHOD 2) can use to object classification, which the difference of the orientation of the edges the real object is very small compared to the fake object.

Finally, the experiment for METHOD 3 was conducted to investigate the feasibility of compare the ratio between the width and height of the object for identification of objects in object recognition. The result of shape variation ratio comparison, the result of this method is pattern of the shape variation of the obstacle to recognize the obstacles. We can compare this variation by calculating the ratio between the width and height of the object. The real obstacle has a low shape variation ratio. In contrast, the fake obstacles has a high shape variation ratio.

4.5 Conclusions

In this dissertation, the object that has to detect is an unknown object - of which we do not know the exact shape, size or color. From the result of feature extraction show that HOG is suitable to use to extract feature of an object in this problem as follow;

METHOD 1, a comparison of the difference of object HOG feature suit for simple shape object classification, but the problem of this method may false classifying under more complicated shape.

METHOD 2, this system used the difference of the object edge between the two images by using HOG feature, which the square error of the orientation of the edges of the real object in sequence of time is very small compared to the fake object. This method effective in identifying both simple shape and complex shapes object.

METHOD 3, this method compare shape variation by calculating the ratio between the width and height of the object. The real obstacle has a low shape variation ratio. In contrast, the fake obstacles has a high shape variation ratio when the vehicle move closer to object. As a result, the author can take pattern of the shape variation of the obstacle to recognize the obstacles.

METHOD 1 and METHOD 2 are feature extraction based on multi-frame extraction, Thus, the classifier has to be an effectively of learning to overcome the limitations of single image.

Chapter 5

Object classification

In the previous chapter, the author presented three methods to extract feature of the object. Each method must consider the criteria of application and the nature of data to select the best feature detector and descriptor. However, classifier is also important in object detection. The accuracy of the classification depends on the selection of the discriminator to suit the feature to be classified.

This chapter presents the method for classify the obstacles based on the three feature extractions described in the previous chapter as follows;

1. Object classification by learning the difference of HOG feature of the object in single image (METHOD 1).
2. Object classification by learning pattern of the difference of HOG feature of the object in sequence of images (METHOD 2).
3. Object classification by learning the shape variation ratio of the object (METHOD 3)

Herein, the author will present the background of the study and then introduce the methodology of object classification. Next, the author will describe several simulations based on each feature extraction. Finally, the author will summarize the study in a conclusion section.

5.1 Background

Image classification is a complex process that may be affected by many factors. As the author mentioned examines current problems, previous works and prospects of image classification.

In this chapter, the author focused on the analysis and the summarization of major advanced classification approaches. In addition, important issues affecting classification performance are discussed, which the designing a suitable imageprocessing procedure is a prerequisite for a successful classification. Effective use of multiple features of data and the selection of a suitable classification method are especially significant for improving classification accuracy.

As the author described in chapter 2, Stereo vision system is widely used in applications ADAS, robot navigation, and autonomous driving [41]. This system have to use two cameras on the front of vehicle take pictures of the same object with difference view. These two images contain some encrypted information about the depth of the object. This information is the third dimension of the two images. Therefore the image distance and its depth can be determined by using the stereo cameras.

However, the aim of this dissertation is to reduce the number of cameras but still be able to detect objects, which one camera is used. Thus, the author proposed the method of detecting objects using only one camera.

In recent year, many image recognition algorithms have been proposed for object classification such as K-NN [37], SVM, ANN

The k-NN classifier is a conventional non-parametric, which the principle is to calculates the distance between the feature vector of the input image and feature vector of training image dataset [71]. Then, it assigns the input image to the class among its k-NN, where k is an integer. The classification doesn't need a lot of examples per class. Thus, it is not suitable for large and complex data.

SVM and ANN are two popular strategies for supervised machine learning and classification. The main concept underlying SVMs is to find a proper hyperplane [55], defined by types of kernels, to separate data belonging to multiple categories. Both linear and RBF kernels have been frequently applied to classification problem via SVMs. Linear SVMs can be trained more quickly, but they are less accurate than non-linear approaches.

In this dissertation, the object that has to detect is an unknown object - of which we do not know the exact shape, size or color, in particular, a detection of moving objects (pedestrians, cars, bicycles, etc.). Detection using single frame images may be faulty. Many studies have improved the detection method by using multi-frame images instead single frame. As the author described in chapter 2, the ANN has been developed as biological neural networks, which in a neural networks consists of neurons are connected together to form a network which mimics a biological neural network. In addition, there is also a time-delay classification algorithm that is capable of solving this problem, which it is TDNN. The TDNN has the potential of learning to overcome the limitations of a multi-layer neural network, and complete image sequences at a time instead of a single image, which it can work with complex data efficiently.

I proposed the object recognition and classification based on HOG feature by using ANN. The features of the object are both static and dynamic data, in this chapter the author proposed two classifiers to compare the performance and selected the best classifier to my novel object detection method. Here are MLFANN and TDNN.

5.2 Methodology

I proposed the object recognition and classification based on HOG feature by using ANN. The features of the object are both static and dynamic data, in this chapter the author proposed two classifiers to compare the performance and selected the best classifier to my novel object detection method. Here are Multi-Layer Feed-forward Artificial Neural Network (MLFANN) and TDNN.

5.2.1 Multi-Layer Feed-forward Artificial Neural Network

The artificial neural networks has the ability to learn on their own like human brain, a beautiful biologically-inspired programming paradigm which enables a computer to learn from observational data. The artificial neural networks has been developed as biological neural networks, which in a neural networks consists of neurons are connected together to form a network which mimics a biological neural network. In a simple mathematical model of neural networks consists of nodes organized into three class called "Layer" include input layer, hidden layer and output layer as in Figure 5.1, the effects of the synapses are represented by connection weights that modulate the effect of associated input signals. The hidden layer is responsible for the processing of the input signal by calculated the weighted sum of input signals, with the help of the transfer function as in Figure 5.2. After that, the network will be classify by comparing the value of the weighted sum of the input signal and the threshold value, with using the activation function for converts a neuron's weighted input to its output activation.

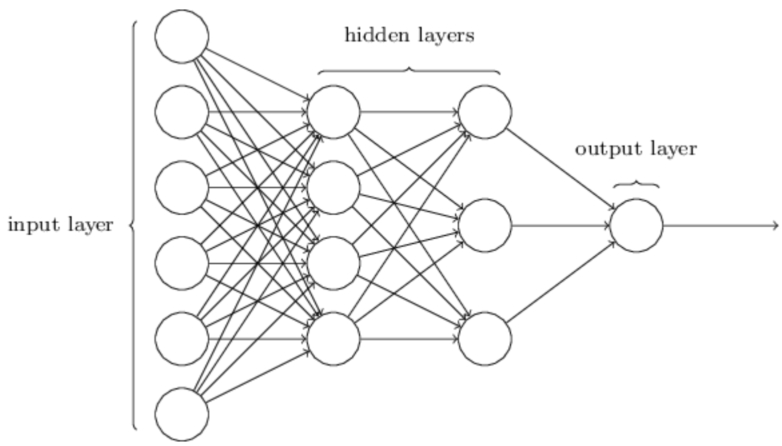


FIGURE 5.1: The example, the following four-layer neural network has two hidden layers

The artificial neural network can be trained in to two groups that are supervised and unsupervised learning. The supervise learning is the learning task of inferring a function from labeled training data. The training data consist of a set of training examples, each example is a pair consisting of an input object and a desired

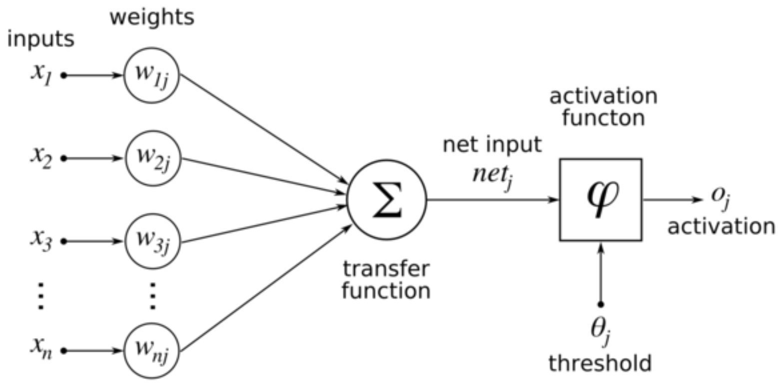


Figure 5.2: A neuron in artificial neural network

target value, whereas in unsupervised learning, learning task from unlabeled data. Unsupervised learning is difficult in practice, but useful if you lack labeled examples.

To solve more complex pattern recognition or classification problems, such as analysing the shape of the obstacles form the input images and determining what obstacle it is, can be performed by adding multi layers of neurons interconnected. The multiple layers neural network usually interconnected in a feed-forward way. Each neuron in one layer has directed connections to the neurons of the subsequent layer. In many classifications tasks apply a sigmoid function as an activation function. The most popular learning techniques that use in multi-layer neural networks being back propagation algorithm, In backward propagation, an abbreviation for Backward Propagation of Errors, is a method used for training of artificial neural network used in conjunction with an optimization method. This method uses to update the weights, in an attempt to minimize the loss function. From the actual output, the network learns from number of inputs, which is supervised learning method. It requires a dataset of desired target output from set of many inputs, making up the training set. A back propagation network consists of three layers as shown in Figure 5.2. These layers of network are connected in feed forward manner, the information processing circuits in the network will be sent in the only one way as the neurons of input layer forward to the neurons of hidden layer and the neurons of hidden layer forward to the neurons of output layer without reversing or even neurons in the same layer is not connect. Back propagation is an iterative process that starts with last layer and moves backward through the first layer. This algorithm is the weight adjustment is done through mean square error of output response to sample input. The rest of these sample patterns are repeatedly presented to the network until the error value is minimized, which weights are adjusted according to the error present in the network as shown in Figure 5.3.

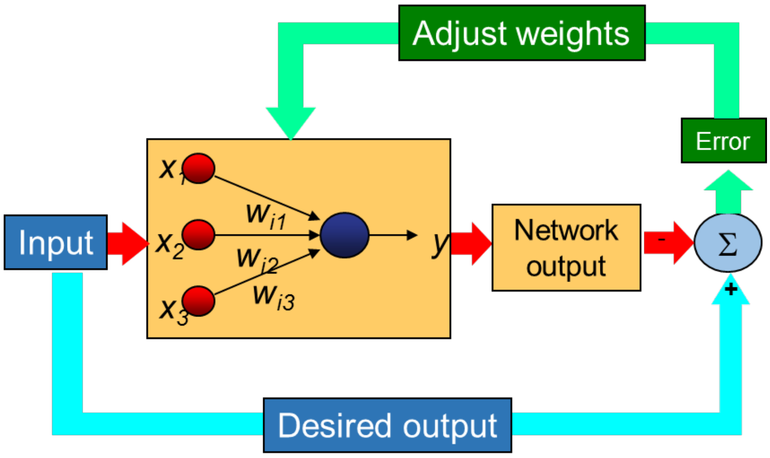


FIGURE 5.3: Supervised learning with back-propagation flow chart

5.2.2 Time Delay Neural Network

The TDNN is an artificial neural network architecture whose primary purpose is to work on sequential data. The TDNN has the ability to recognize features of time-shifting and has a larger pattern recognition system. The general TDNN concept is well known from applications in the field of speech recognition [62]. Currently, TDNN is commonly used in image-pattern shape or motion recognition tasks. The TDNN has the potential of learning to overcome the limitations of a multi-layer neural network, and complete image sequences at a time instead of a single image.

In a simple mathematical model of TDNN like other neural networks, which consists of nodes organized into three layers of clusters including input layer, output layer, and the hidden layer which handles the manipulation of the input through filters as in Figure 5.4, the effects of the synapses are represented by connection weights that modulate the effect of associated input signals. The hidden layer is responsible for the processing of the input signal by calculating the weighted sum of input signals, with the help of the transfer function as in Figure 5.5.

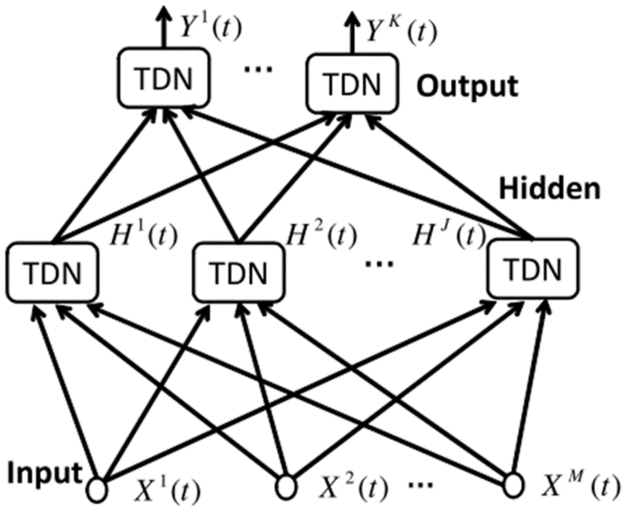


FIGURE 5.4: Overall architecture of the TDNN

After that, the network will be classified by comparing the value of the weighted sum of the input signal and the threshold value, while using the activation function to convert a neurons weighted input to its output activation. In order to achieve time-shift invariance, a set of delays are added to the input so that the data are represented at different points in time such as audio files or sequences of images. An important feature of TDNN is the ability to express relations between inputs in time, which can be used to recognize patterns between the delayed inputs. Due to their sequential nature, TDNNs are implemented as feed-forward neural networks, the flow of data in only one direction, forward from the input nodes through the hidden nodes and to the output nodes. There are no cycles or loops in the network. Supervised learning with a back propagation algorithm is generally the learning algorithm associated with TDNN.

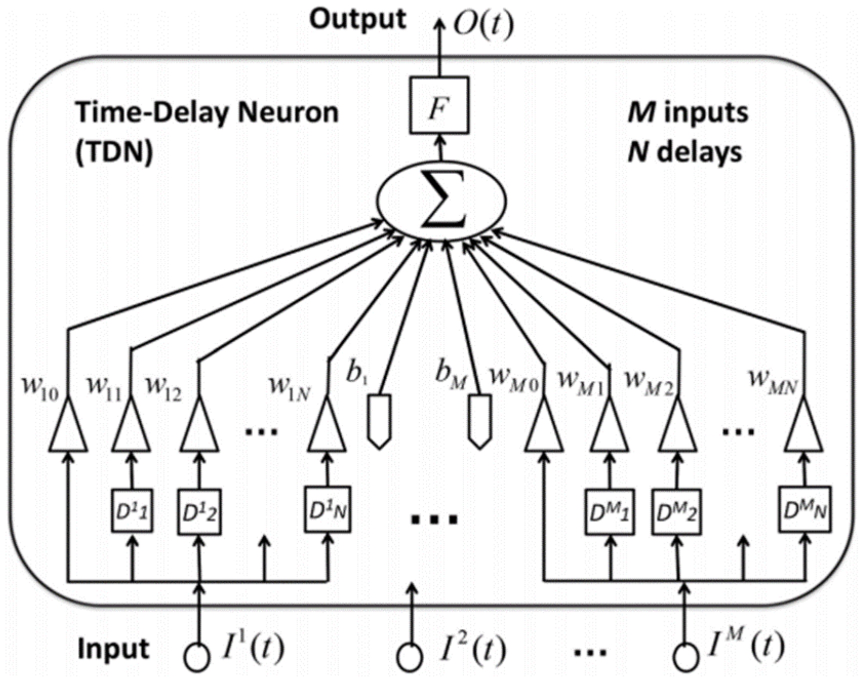


FIGURE 5.5: Single TDNN with M inputs and N delays for each input at time t. D^i_d are the registers that store the values of delayed input $I^i(t-d)$

5.2.3 Experiment and results

I conducted experiments to evaluate my method. Proposed algorithm was programmed in MATLAB and executed on a Intel Core i5-4200U, CPU 1.6GHz 2.29 GHz, 4 GB memory. The author divided them into twelve tests as presented in Table 5.1. To evaluate the object classification, the author compared results obtained from MLFANN and TDNN classifier with three case, i.e., the AGV in factory environment[C.1, C.2], the electric senior vehicles [C.8], and the vehicle in traffics [J.1, C.3, C.6].

5.2.3.1 Preparation of input for object classification

Gather sample images consisting of a real object image and a fake image (positive and negative samples) for training and test. Then, organize and partition the

TABLE 5.1: Settings of my object classification experiments

Experiment	**Case study**	**Object**	**Feature extrction**	**Classifier**
1	AGV	Simple shape	Method 1	MLFANN
2	Senior car	Complex shape	Method 1	MLFANN
3	Vehicle in traffic	Complex shape	Method 1	MLFANN
4	AGV	Simple shape	Method 2	TDNN
5	Senior car	Complex shape	Method 2	TDNN
6	Vehicle in traffic	Complex shape	Method 2	TDNN
7	AGV	Simple shape	Method 3	TDNN
8	Senior car	Complex shape	Method 3	TDNN
9	Vehicle in traffic	Complex shape	Method 3	TDNN
10	AGV	Simple shape	Method 2 & 3	TDNN
11	Senior car	Complex shape	Method 2 & 3	TDNN
12	Vehicle in traffic	Complex shape	Method 2 & 3	TDNN

images into training and test subsets as show in Figure 5.6. The input of MLFANN are set of single image. In contrast, the input of TDNN are set of delay input, here it is sequence of video images. After that, label the training images.

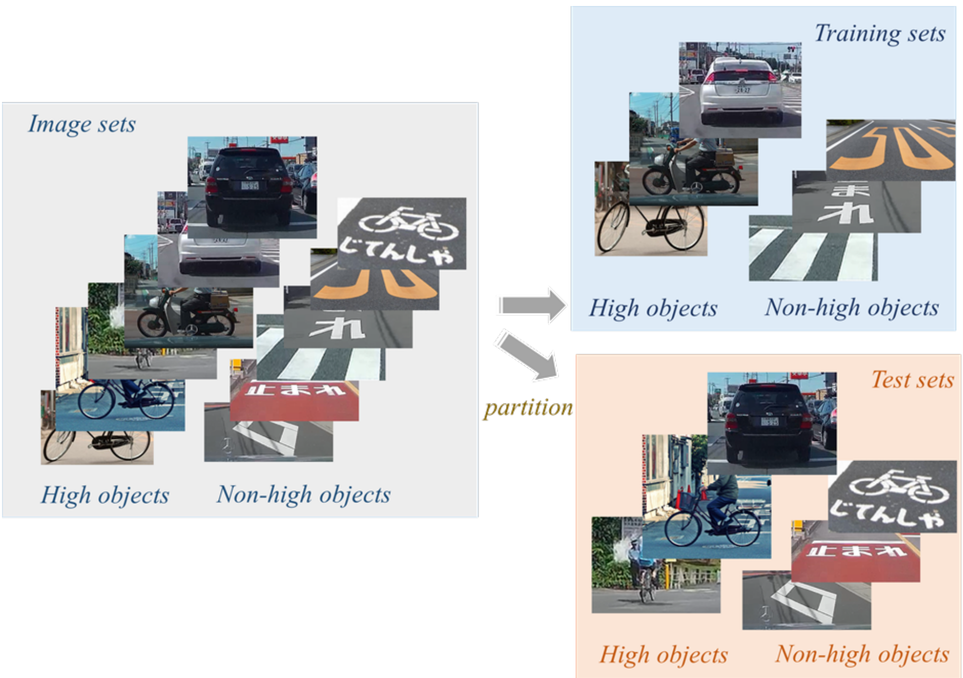


FIGURE 5.6: The sample images into training and test subsets.

5.2.3.2 The experiment configuration

As the author described in the ANN background, the hidden layer is responsible for the processing of the input signal to learning and classify the object. It is important to make sure the number of hidden neurons is appropriate for learning the feature of the object. The author has tested the effect of the number of hidden neurons on the training accuracy and training time. By varying the number of hidden neurons parameter and visualizing the result as show in Figure 5.7 for MLFANN training and Figure 5.8 for TDNN training.

The number of hidden neurons visualization plot shows that the MLFANN training by setting the number of hidden neurons is 10 neurons gives a similar accuracy to 11 and 12 neurons, but it takes less training time as show in Figure 5.7.

Similarly, the TDNN training by setting the number of hidden neurons is 20, 22 and 24 neurons gives a similar accuracy, but 20 neurons takes training time less than 22 and 24 neurons in all three cases as show in Figure 5.8.

Therefore, the optimal number of hidden neurons for MLFANN training is 10 neurons and 20 neurons for TDNN training.

Moreover, the amount of input is also important to accuracy of training and classification by ANN. The author has tested the effect of the number of input on the training accuracy and training time by varying the number of input and visualizing the result as show in Figure 5.9 for MLFANN training and Figure 5.10 for TDNN training.

The number of input visualization plot shows that the MLFANN training by using 120 samples gives a similar accuracy to 130 and 140 samples, but it takes less training time as show in Figure 5.9.

In the same way, all of three cases of the TDNN training by using 150, 160 and 170 sets of delay gives a similar accuracy, but the input 150 sets takes training time less than 160 and 170 sets as show in Figure 5.10.

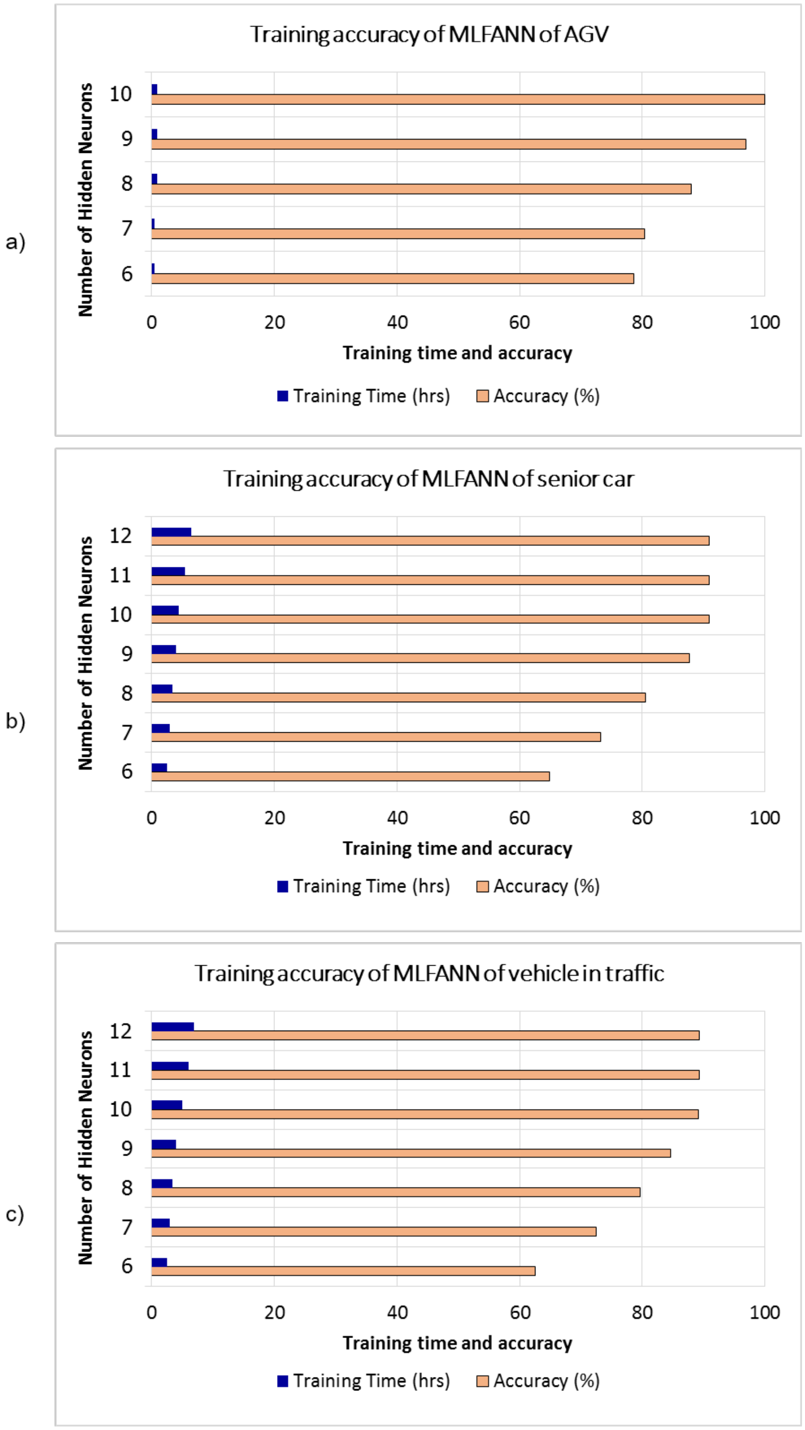


FIGURE 5.7: The number of hidden neurons visualization plot for MLFANN training; (a) the experiment for AGV; (b) the experiment for electric senior vehicle; (b) the experiment for vehicle in traffic

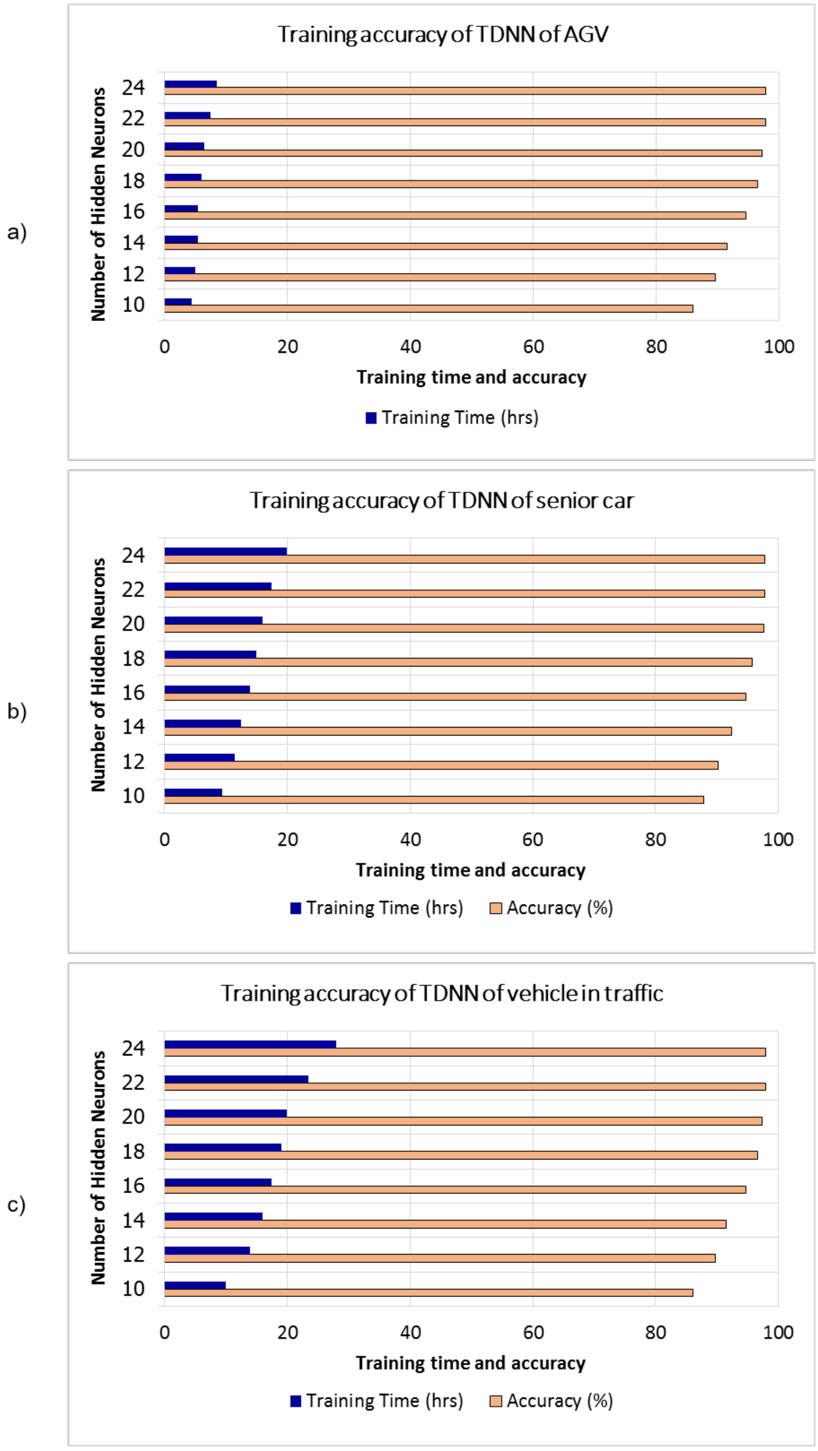


FIGURE 5.8: The number of hidden neurons visualization plot for TDNN training; (a) the experiment for AGV; (b) the experiment for electric senior vehicle; (b) the experiment for vehicle in traffic

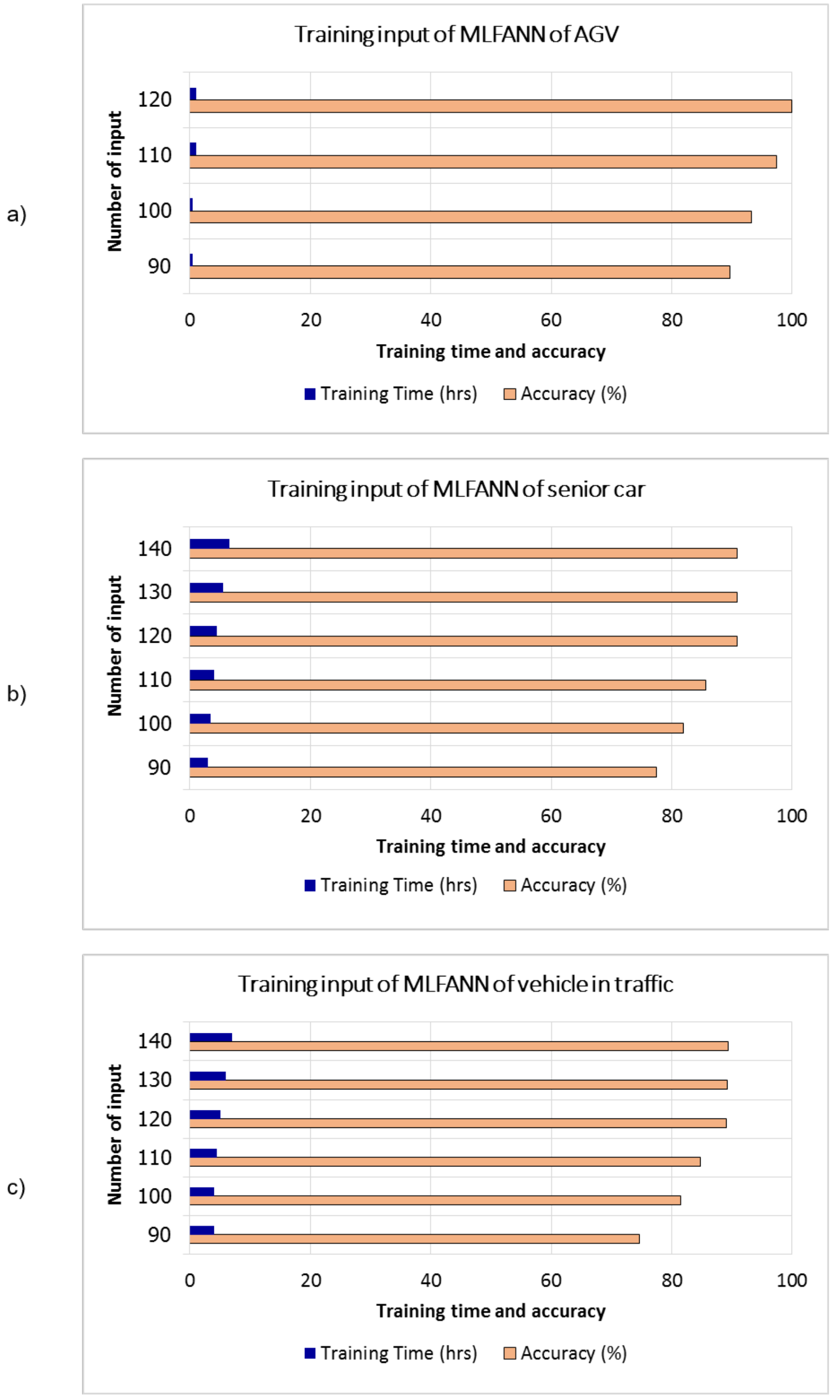


FIGURE 5.9: The number of inputs visualization plot for MLFANN training; (a) the experiment for AGV; (b) the experiment for electric senior vehicle; (b) the experiment for vehicle in traffic

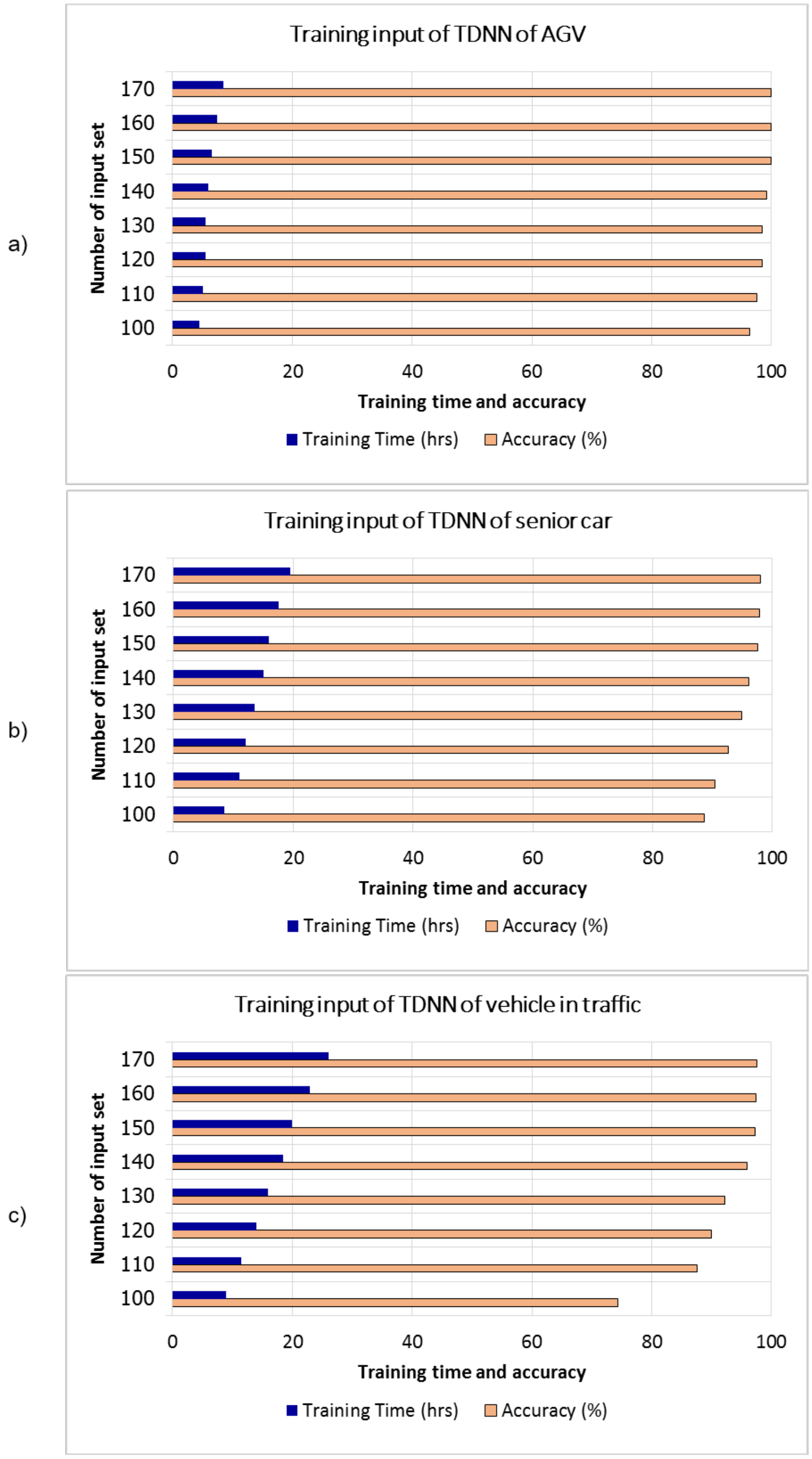


FIGURE 5.10: The number of set of inputs visualization plot for TDNN training; (a) the experiment for AGV; (b) the experiment for electric senior vehicle; (b) the experiment for vehicle in traffic

Accordingly, the MLFANN training and validation by using static images, 120 samples (70 obstacles and 50 fake obstacles). The second is the hidden layer, to recognize and classify the obstacles consisting of 10 neurons with a sigmoid activation function by learning the difference features of the obstacles from as feature. The last is the output layer, consisting of two neurons where the real obstacle and fake obstacle by the steps as show in Figure 5.11. The classification testing by using actual video images 1,500 frame (the obstacles: 900 frames, the fake obstacles: 600 frames). This set of images used in this classification test is different from the set of sample used for training process.

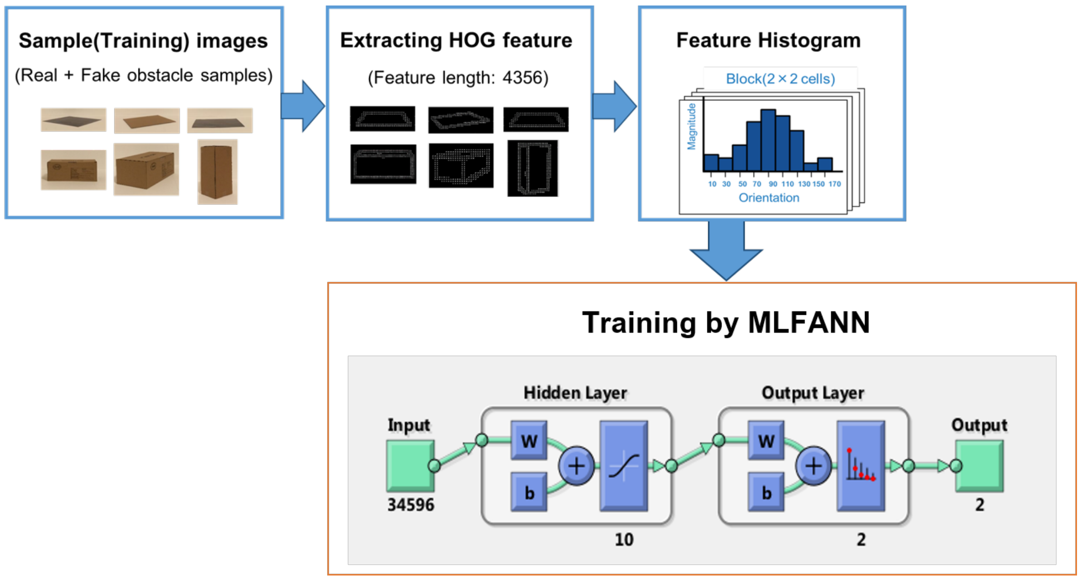


FIGURE 5.11: The MLFANN training and validation by learning HOG feature (input length for simple shape is 4356, 34596 for complex shape)

For the TDNN training process, the process is to recognize the features of both types of objects. The first is the input layer, which is 150 sets of inputs and the 5 delays are extracted sequences of video images taken by an on-board camera which are feature of HOG. The second is the hidden layer, to recognize and classify the obstacles consisting of 20 neurons with a sigmoid activation function by learning the features of the obstacles from HOG and recognizing the difference in the patterns of the obstacle shape variation ratio and the orientation of HOG feature when the vehicle is moving, where the real obstacle has a shape variation ratio lower than the fake obstacle as show in Figure 5.12. The last is the output layer, consisting of two neurons where the real obstacle and fake obstacle are as the author described in previous chapter.

Moreover, the classification test, we use 25 set of video images (the obstacles: 15 set, the fake obstacles: 10 set). This set of images used in this classification test is different from the set of sample used for training process.

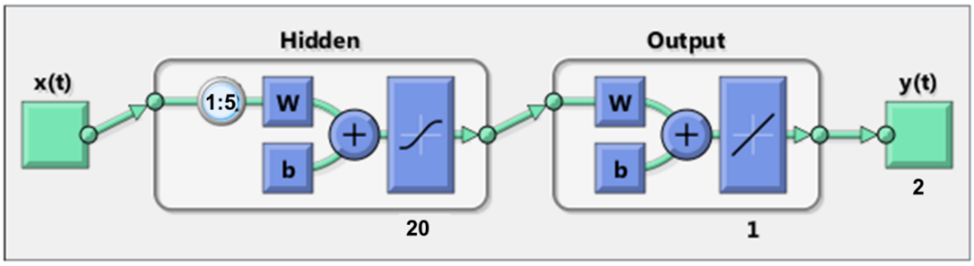


FIGURE 5.12: The TDNN training and validation

5.2.3.3 The result of the experiments

The result of the object recognition, validation of all experiments as presented in Table 5.2 and Table 5.3. The comparison of the performance and error as shown in Figure 5.13 and Figure 5.14.

As experiment 1-3, they were learning HOG feature by MLFNN, which this HOG feature is the histogram of gradient orientation of edge of the object. The accuracy as show in Figure 5.13. The verification process can be observed as having a network that at stabilized at 100%, 90.83%, 89.16%accuracy for the AGV in factory environment, the electric senior vehicles, and the vehicle in traffics respectively.

However, after examining accuracy from Experiment 4-6 which were learning HOG feature by TDNN, which this HOG feature is the pattern of the difference of HOG feature of the object in sequence of images and the input are set of delay from video images. This extraction method could efficiently training better than learning HOG feature in single frame method. Moreover, the accuracy of the experiment of the electric senior vehicles and the vehicle in traffics were increased to 93.44%

TABLE 5.2: The number of epoch and training time of object classification experiments

Experiment	Case study	Feature extrction	Classifier	Epoch	Training time
1	AGV	Method 1	MLFANN	358	1 hr.
2	Senior car	Method 1	MLFANN	1011	4.5 hrs.
3	Vehicle in traffic	Method 1	MLFANN	1500	5 hrs.
4	AGV	Method 2	TDNN	912	6 hrs.
5	Senior car	Method 2	TDNN	1896	12.5 hrs.
6	Vehicle in traffic	Method 2	TDNN	1994	12.5 hrs.
7	AGV	Method 3	TDNN	897	6 hrs.
8	Senior car	Method 3	TDNN	1952	12 hrs.
9	Vehicle in traffic	Method 3	TDNN	2011	12 hrs.
10	AGV	Method 2 & 3	TDNN	1125	6.5 hrs.
11	Senior car	Method 2 & 3	TDNN	2885	16 hrs.
12	Vehicle in traffic	Method 2 & 3	TDNN	3060	20 hrs.

TABLE 5.3: The result of the object classification experiments

Experiment	Accuracy(%)	False Positive (%)	False Negative (%)
1	100	0	0
2	90.83	5.83	3.33
3	89.16	6.67	4.17
4	100	0	0
5	93.44	4.23	2.33
6	93.12	5.61	1.27
7	100	0	0
8	92.91	3.62	3.47
9	93.47	4.81	3.92
10	100	0	0
11	97.6	2.4	0
12	97.33	2.67	0

and 99.12% respectively, whereas the performance rates of Experiment 1 were only 90.83% and 89.16%.

Experiment 7-9 were learning the pattern of shape variation ratio of the object by TDNN. The input sequence of images and the input are set of delay as in Experiment 4-6. The accuracy of the electric senior vehicles was slightly reduced to 92.91%, but still more than the first method. Moreover, the vehicle in traffics experiment Slightly increased to 93.47. Based on this results, Method 2 and 3 are similarly effective.

Finally, Experiment 10-12 were learning both of the pattern of the difference of HOG feature and the pattern of shape variation ratio, all experiments were dramatically improved comparing to the first experiment. The accuracy of the electric senior vehicles was up to 97.60%, and the vehicle in traffics reached to 97.33%.

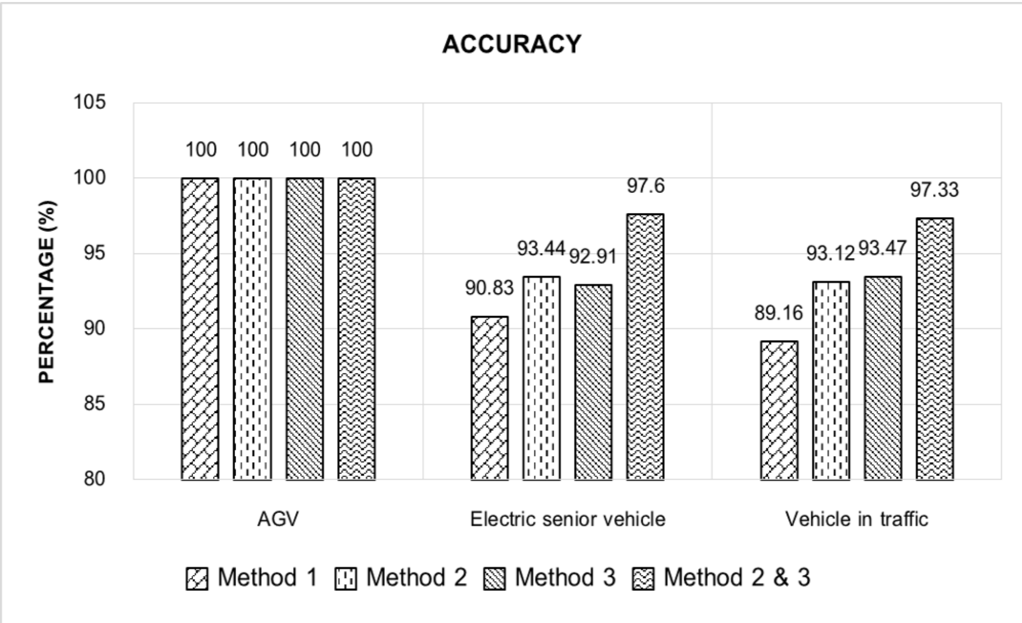


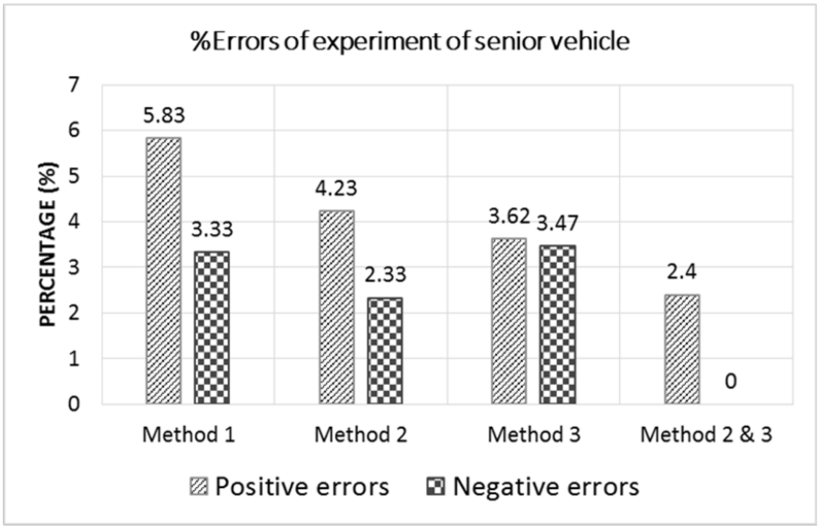
FIGURE 5.13: The accuracy of object classification of all experiment

From the results of the training of electric senior vehicle, METHOD 1 is the highest error both of false positive and false negative. The METHOD 2 can reduce the error to 4.23% and 2.33% respectively, whereas the error of Experiment 1 were 5.88% and 3.33%.

Method 3 can reduce the false positive error to 3.62%, in contrast, the false negative increase to 3.47. However, METHOD 4 reduce the false positive to 2.4% and 0% for false negative.

For vehicle in traffic have the errors in the same trend as electric senior vehicle as shown in Figure 5.14.

a)



b)

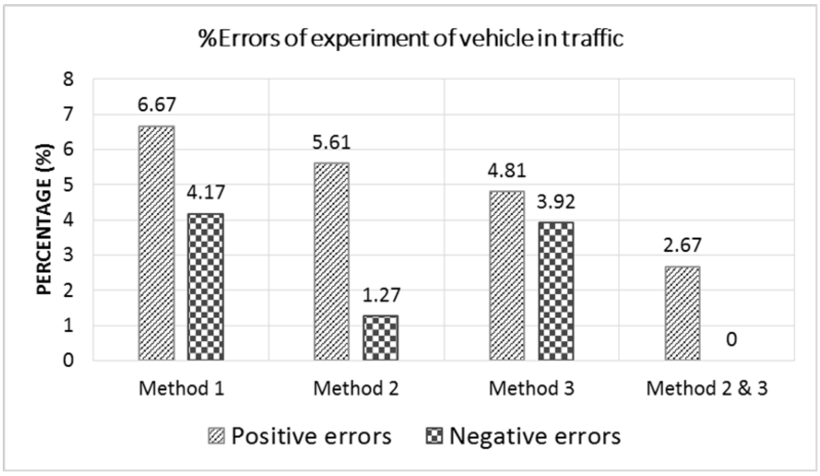


Figure 5.14: Illustration of %errors of all experiment; (a) the experiment for electric senior vehicle; (b) the experiment for vehicle in traffic

5.3 Discussion

I proposed an object recognition and classification. The author improved my idea by using HOG feature to be input of the network.

I conducted twelve tests to evaluate the performance of my method as presented in Table 5.1. The author calculated several performance rates, i.e., accuracy, false positive, and false negative. The experiment representing my method were

Experiment 9-12 is to combined the pattern of the difference of HOG feature and the pattern of shape variation ratio as input into the object recognition.

As presented in Figure 5.3 and 5.14, METHOD 1 provided the lowest accuracy, including the highest error; thus, the learning the difference of HOG feature of the object in single image is inappropriate to solve this study problems under complex shape, but it can used to learning the difference of HOG feature of the object under simple shape object as in AGV case.

For METHOD 2, the author observed that the errors reduced from the first method. Moreover, the accuracy increased. This positive situation happened because the author changed the feature that use be input from the difference of HOG feature in single image to the difference of HOG feature in video images.

However, METHOD 3 gives a similar accuracy to METHOD 2, but there are more false negative, in contrast, false positive decreased.

METHOD 4 provided high accuracy and lowest errors. To compare between METHOD 1, 2 and 3, the results showed that the feature selection and extraction affects the performance of the system, which it is a combined the pattern of the difference of HOG feature and the pattern of shape variation ratio as input into the object recognition by learning in sequence of images, not only learning in single image.

5.4 Conclusions

In this study, the author proposed the method of learning feature to classify the object. The objectives of this study were to learning the difference feature of the object to recognize the object is a real object or fake object.

I conducted twelve tests to evaluate the performance of my method. The author calculated several performance rates, i.e., accuracy, false positive, and false negative.

From the experiments that have been carried out, it is possible to conclude that my methods achieve the objectives of this study. As the experimental results, METHOD 4 presented the highest performance rates greater than other methods, and the error was small. My method is a combined the pattern of the difference of HOG feature and the pattern of shape variation ratio as input into the object recognition by TDNN. This method has only false positive without false negative.

Based on the findings, it is more efficient to learn sequential images than to learn single image. Moreover, the recognition by learning the combination of features has more effective than learning one feature. The TDNN has the potential of learning to overcome the limitations of a MLFANN, and complete image sequences at a time instead of a single image, which it can work with complex data efficiently

Chapter 6

Object detection system

In the previous chapter, the author proposed a method of classifying feature of the object, including selecting the best feature extraction that suits for my problem.

The main focuses of this chapter is proposed object detection using computer vision by ANN. Background of the study will be presented at the first section of this chapter. The next section is the methodology which describes overall system design and implementation. Regard experiment and evaluation parts, the author will explain in the next chapter.

6.1 Background

As the author describe in chapter 1, the aim of this dissertation is to reduce the number of cameras but still be able to detect objects, which one camera is used. Thus, the author proposed the method of detecting objects using only one camera.

Object detection is the process of finding instances of real-world objects such as pedestrian [17], traffic sign [56], faces [19], and car [26] in images or videos. Object detection algorithms typically use extracted features and learning algorithms to recognize instances of an object category. It is widely used in applications such as image retrieval, security, surveillance, and ADAS.

When considering the actual vehicle driving situation on the road, it is desirable to be able to recognize the preceding obstacles in order to avoid accident. Thus, one of the most important obstacle avoidance is the vehicle has to be able to detect, recognize and classify the obstacles that are real obstacles or fake obstacles. So far, many works have been developed using neural networks for image analysis applied to obstacles recognition and classification, which is an important task in an automotive safety application.

In this dissertation, the author focused on general object detection for vehicles by using computer vision. In the process of the object detection include object analysis as the author described in chapter 3, analyzing specific features of the objects from a camera perspective and identify the ROI of detection by conditions in each case. In chapter 4 presented the selecting feature detection and descriptor, this method can be can be utilized to identify and extract specific feature of shape of both real object and fake object.

In previous chapter proposed object recognition and classification and the author conducted twelve tests to evaluate the performance of classifier. Finally, the author will introduce a prototype of object detection system that integrates entire proposed systems in this chapter.

6.2 Methodology

I proposed the object detection based on HOG feature by using ANN. The features of the object are both static and dynamic data, in this chapter the author proposed two detector to compare the performance and selected the best detector to my novel object detection method. Here are Multi-Layer Feed-forward Artificial Neural Network (MLFANN) and TDNN.

This section is separated into four parts, i.e., pre-processing, feature extraction, object recognition and classification, and object detection.

6.2.1 Pre-processing

Gather sample images consisting of a real object image and a fake image (positive and negative samples) for training and test. Then, organize and partition the images into training and test subsets as the author described in section 5.3.1. A pre-processing step is image enhancement, the input images of our system are initially converted to binary image, applied to remove noise for highlight certain features of interest in the images, cropped to region of interest (ROI) as the author described in section 4.2.1.

The input of MLFANN are set of single image. In contrast, the input of TDNN are set of delay input, here it is sequence of video images. After that, label the training images as show in Figure 6.1.

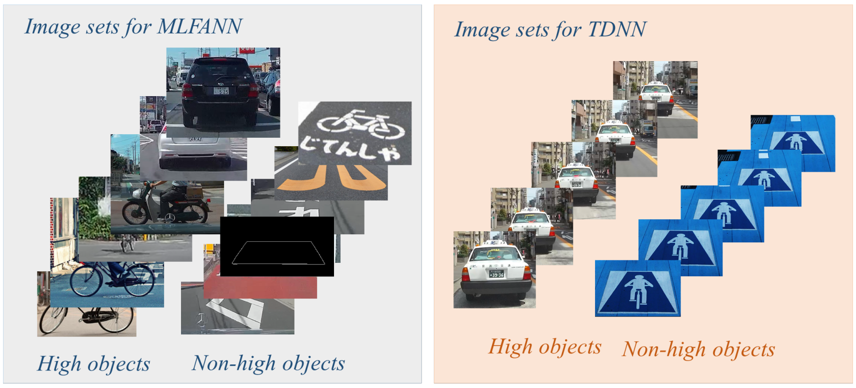


FIGURE 6.1: Illustration of Image set for object recognition and classification

6.2.2 Feature extraction

I proposed a HOG method to extract feature of the obstacles, which is highly effective in the feature extraction of the object in the image. Therefore, the author used the HOG method to extract features of the obstacles which can detect objects and shapes within an image by analyzing the distribution of the intensity gradient and edge direction, and then explain the image in a histogram. From the result of feature extraction show that HOG is suitable to use to extract feature of an object in this problem. The features extraction that suit for classify in this problem are four features as follow:

1. The difference of HOG feature of the object in single image (METHOD 1), which it is comparing of the different of gradient orientation of the object edge in single image.
2. The pattern of the difference of HOG feature of the object in sequence of images (METHOD 2), which it is comparing the difference of the orientation of the edges between two image. The real object is very small different changes when compared to the fake object.
3. The pattern of shape variation ratio of the object (METHOD 3), which the fake object has a shape variation ratio over the real object.
4. The combination of METHOD 2 and METHOD 3

The step of HOG to extract the feature of the object follow as Figure 6.2 and the principles of HOG is explained in detail in Chapter 4. The final feature vector includes all of the block in 1D matrix form containing all histograms as HOG feature. Set the following parameters:

1. Down scale images size to 256x256
2. Linear gradient voting into 9 orientation bins in 0 to 180 degrees
3. Sub-images by 2-by-2 blocks
4. Cell size 20-by-20 pixel for simple shape, 8-by-8 pixel for complex shape

6.2.2.1 Object recognition and classification

I proposed the object recognition and classification based on HOG feature by using ANN. The features of the object are both static and dynamic data, in

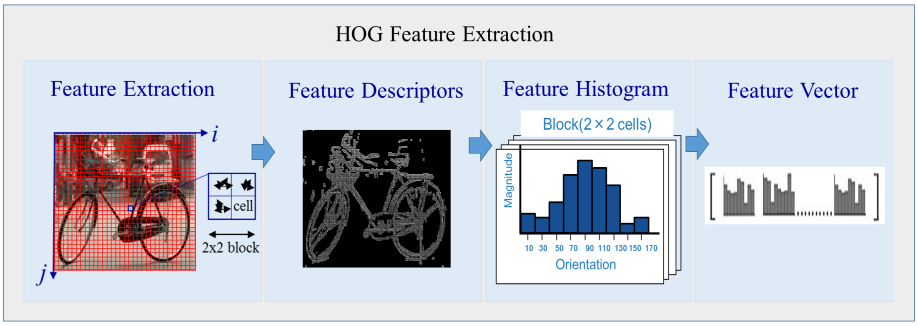


FIGURE 6.2: The process of HOG method

this chapter the author proposed two classifiers to compare the performance and selected the best classifier to my novel object detection method. Here are MLFANN and TDNN.

I designed the object classification into four method as follows:

1. Object classification by learning the difference of HOG feature of the object in single image with MLFANN
2. Object classification by learning pattern of the difference of HOG feature of the object in sequence of images with TDNN.
3. Object classification by learning the shape variation ratio of the object with TDNN.
4. Object classification by learning the combination of the pattern of the difference of HOG feature and the pattern of shape variation ratio as input into the object recognition by learning in sequence of images, not only learning in single image with TDNN.

The MLAFNN classification process is explained in detail in chapter 5.

For the training process in TDNN, where the training process is supervised learning, and the network learns by labeled examples. The process is to recognize the features of both types of objects. The first is the input layer, which is 150 sets of inputs and the 5 delays are extracted sequences of video images taken by an on-board camera which are feature of HOG. The second is the hidden layer, to recognize and classify the obstacles consisting of 20 layers with a sigmoid activation function by learning the features of the obstacles from HOG and recognizing the difference in the patterns of the obstacle shape variation ratio and the orientation of HOG feature when the vehicle is moving, where the real obstacle has a shape variation ratio lower than the fake obstacle as show in Figure 6.3. The last is the output layer, consisting of two neurons where the real obstacle and fake obstacle are as the author described in previous chapter.

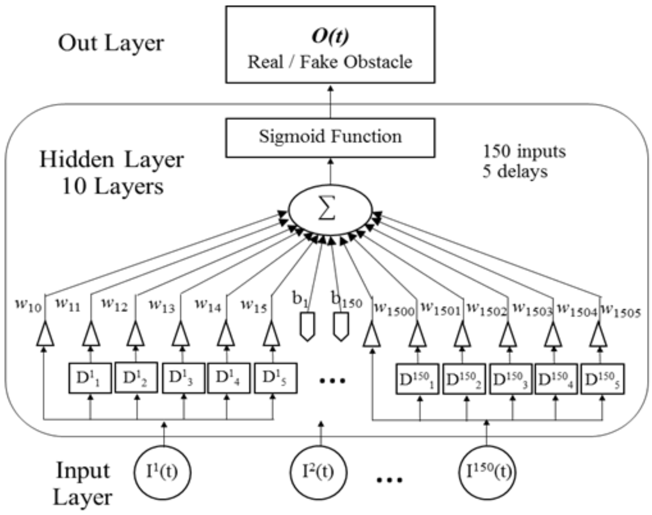


FIGURE 6.3: The design of the object recognition and classification by TDNN

6.2.2.2 Object detection

I proposed the object detection based on HOG feature by using MLFANN for learning a single data and TDNN for learning the video images. The author designed to use the detector same algorithm with the classifier as follow:

1. MLFANN detector used to detect object in case of the learning feature by MLFANN classifier. The process of MLAFNN detector as show in Figure 6.4.
2. TDNN detector used to detect object in case of the learning feature by TDNN classifier. The process of TDNN detector as show in Figure 6.5.

Moreover, the object detection test, the author use a set of video images (the obstacles: 15 set, the fake obstacles: 10 set). This set of images used in this detection test is different from the set of sample used for training process.

In each situation, the area that needs to be aware of the objects is different. Thus, to reduce the errors that may occur from analyzing unnecessary areas, reduce processing time, and simplify analysis. In this chapter, the author presented the region limitations for detecting objects depending on the environment by setting the region of interest (ROI) in image. This method applicable for both lane and non-lane based traffic scenarios focused mainly on the ROI in front of the vehicle, which limits the processing area to the ground locations as shown in Figure 6.6.

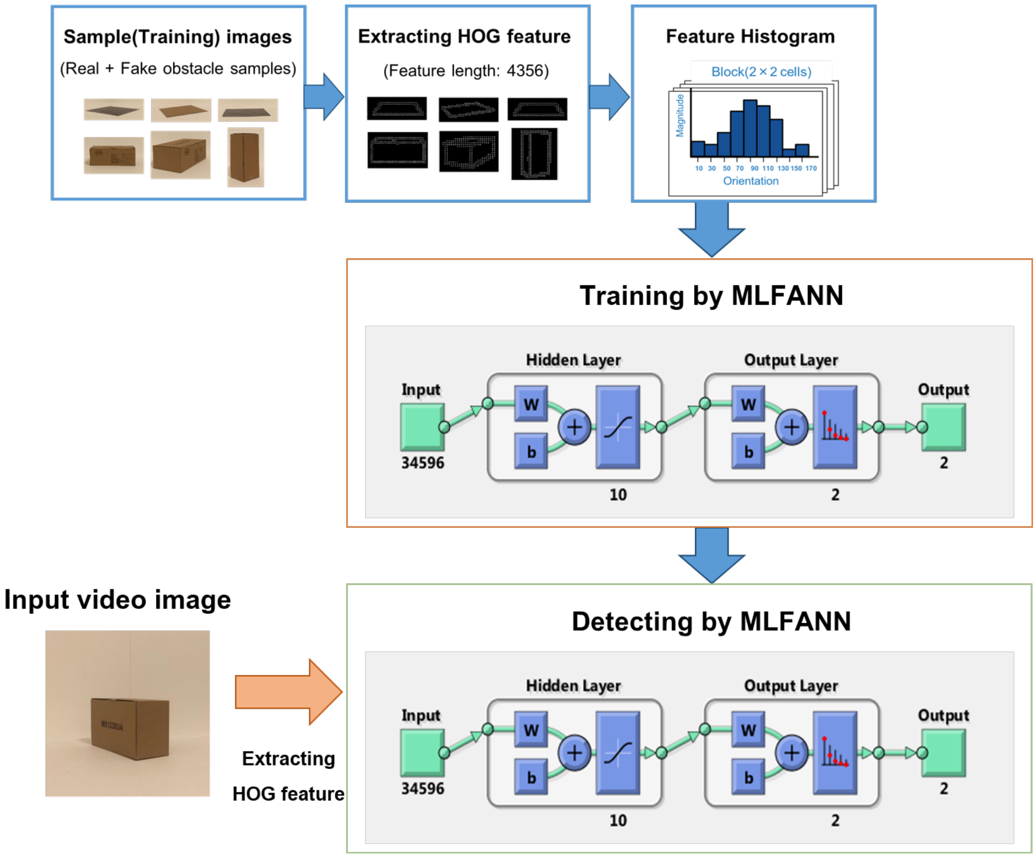


FIGURE 6.4: The process of object detection by MLFANN

The experiment and result of object detection is discussed in the next chapter.

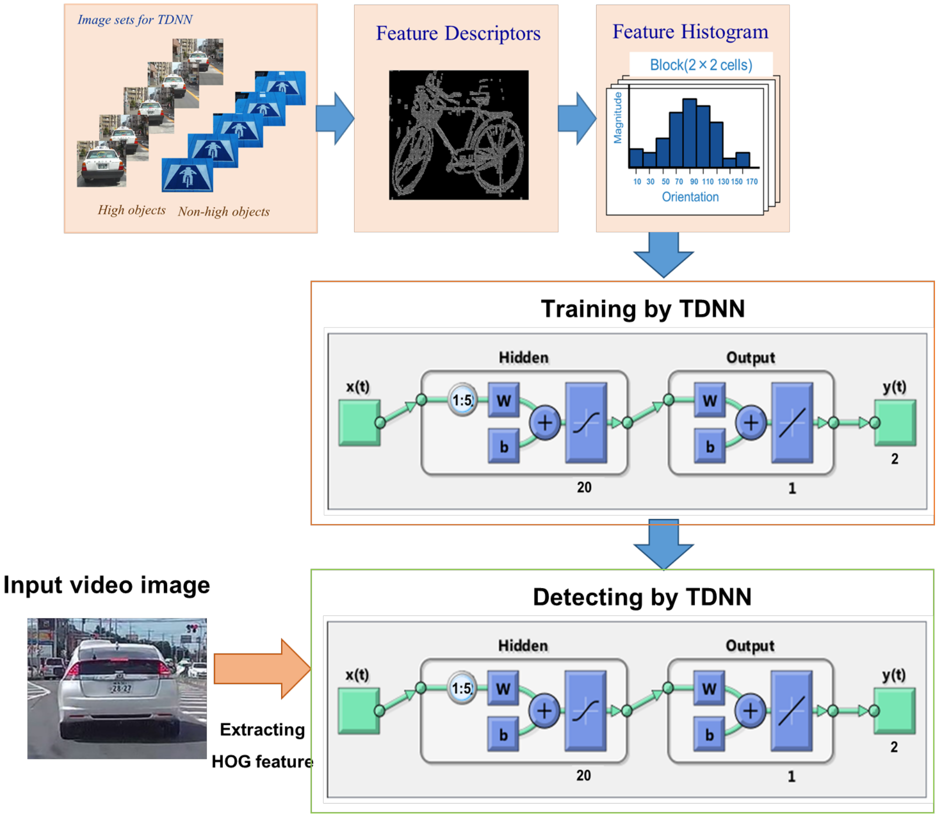


FIGURE 6.5: The process of object detection by TDNN

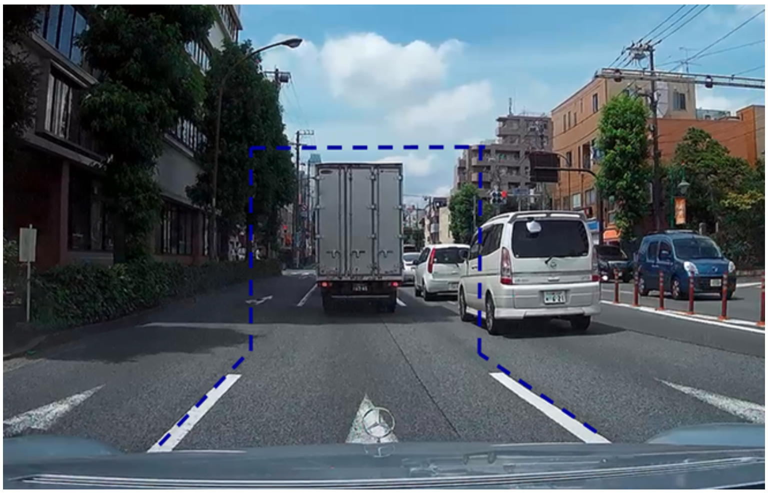


FIGURE 6.6: The ROI to detect the obstacles

Chapter 7

Experiment and Evaluation

In the previous chapter, the author presented a methodology of object detection based on HOG feature with ANN by used only one camera. This method applicable for both lane and non-lane based traffic scenarios focused mainly on the ROI in front of the vehicle. The author designed and constructed my own algorithm to extract feature of the object to classify the object that is real object or fake object. As the author mentioned in the prior chapter, the author focused on extracting explicit and implicit information (e.g., orientation of edges, the difference of edge orientation between two images, and shape variation ratio) from the image taken by an on-board camera.

In this chapter, the experiments and results will be shown; moreover, a detail about experiment procedure will be included.

Initially, the author will describe about experiment configuration describing settings used to accomplish the experiments. Furthermore, experiment procedure will be described. It briefly described total steps of this dissertations experiments. Finally, the author will show the experimental results and the system performance.

7.1 Experimental design for object detection

This study focuses on general object detection for vehicles by using computer vision. The key for detecting objects with the camera is the lighting conditions. As a result, places and environments directly affect the object detection. Besides, the obstacle that may be found in each case are different. Under this condition, the author will experiment with three scenarios as follows:

1. Object detection for AGV in indoor scene.
2. Object detection electric senior vehicle in a senior car lane.
3. Object detection for vehicle in traffic.

As the author mentioned in chapter 3, the aim of this dissertation is to detect the obstacles in front of a vehicle. Real time object detection by computer vision was crucial in that fake obstacles may be found, such as text or symbols on the road. Therefore, this algorithm needs to recognize the difference of specific features of the obstacles, where the real obstacle is a high object and the fake obstacle is a non-high object.

Based on the result of analysis of the characteristics of each object, the object feature that can be used to classify the obstacles that are real obstacles or fake obstacles as follows:

1. The difference of HOG feature of the object in single image (METHOD 1), which it is comparing of the different of gradient orientation of the object edge in single image.
2. Object classification by learning pattern of the difference The pattern of the difference of HOG feature of the object in sequence of images (METHOD 2), which it is comparing the difference of the orientation of the edges between two image. The real object is very small different changes when compared to the fake object. HOG feature of the object in sequence of images (METHOD 2).

3. The pattern of shape variation ratio of the object (METHOD 3), which the fake object has a shape variation ratio over the real object.

4. The combination of METHOD 2 and METHOD 3.

For the object classification and detection, the author proposed the object detection based on HOG feature by using MLFANN for learning a single data and TDNN for learning the video images. the author designed to use the detector same algorithm with the classifier as follow:

1. MLFANN detector used to detect object in case of the learning feature by MLFANN classifier.

2. TDNN detector used to detect object in case of the learning feature by TDNN classifier.

I conducted experiments to evaluate my object detection method. The author divided them into thirteen tests as presented in Table 7.1.

TABLE 7.1: Settings of my object detection experiments

Exp.	**Case study**	**Object**	**Feature extrction**	**Detector**
1	AGV	Simple shape	Method 1	MLFANN
2	Senior car	Complex shape	Method 1	MLFANN
3	Vehicle in traffic	Complex shape	Method 1	MLFANN
4	AGV	Simple shape	Method 2	TDNN
5	Senior car	Complex shape	Method 2	TDNN
6	Vehicle in traffic	Complex shape	Method 2	TDNN
7	AGV	Simple shape	Method 3	TDNN
8	Senior car	Complex shape	Method 3	TDNN
9	Vehicle in traffic	Complex shape	Method 3	TDNN
10	AGV	Simple shape	Method 2 & 3	TDNN
11	Senior car	Complex shape	Method 2 & 3	TDNN
12	Vehicle in traffic (daytime)	Complex shape	Method 2 & 3	TDNN
13	Vehicle in traffic (nighttime)	Complex shape	Method 2 & 3	TDNN

7.2 Experimental procedure

In this section, we will demonstrate our proposed method with the experiment we conducted. The aim of this experiment is to improve function, detecting objects to recognize and classify the obstacles that are real obstacles or fake obstacles such as a painting, sign or text on the road.

In this dissertation, the author proposed a combined on-board computer vision system based on HOG features and ANN. The author created an experiment to extract features of the obstacles in the actual video images by using the HOG method. For obstacle recognition and classification, and detection in real-time the author used MLFANN and TDNN to compare the performance of each method. These images were processed by using the sequences of video images taken by an on-board camera that was fixed on board the front of a vehicle. The experiment was done in a real environment and speed limit is 6 km/h for electric senior vehicle, 60 km/h for vehicle in traffic. The experiment consisted of four main parts; pre-processing, feature extraction, object classification, and object detection. The overview process of the experiment is shown in Figure 7.1.

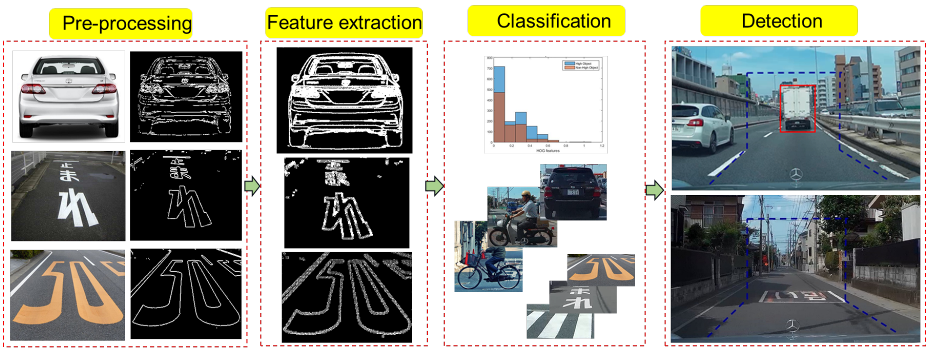


FIGURE 7.1: An overview of the experiment procedure

Initially, gather sample images consisting of a real object image and a fake image (positive and negative samples) for training and test. Then, organize and partition the images into training and test subsets. The pre-processing process aims to prepare the data for the next stage. The output of this stage would be ready for the next stage to perform complicated image processing tasks on the data. Prior to extracting features, training and testing a classifier, a pre-processing step is image enhancement applied to remove noise to highlight certain features of interest in the images, cropped to region of interest as described in chapter 3.

Next, used the HOG method to extract features of the obstacles which can detect objects and shapes within an image by analyzing the distribution of the intensity gradient and edge direction, and then explain the image in a histogram. The feature extraction configuration as described the details in chapter 4.

Then, recognize and classify the object by using ANN. The features of the object are both static and dynamic data, in this chapter the author proposed two classifiers to compare the performance and selected the best classifier to my novel object detection method. Here are MLFANN and TDNN as described the details in chapter 5.

Finally, the author tested object detection by used the algorithm same as classification process as described the details in chapter 6.

Moreover, the object detection test, the author use a set of video images. This set of images used in this detection test is different from the set of sample used for training process.

7.2.1 Experimental results and evaluation

All evaluations were done on a Intel Core i5-4200U, CPU 1.6GHz 2.29 GHz, 4 GB memory. The result of the object classification of all experiments as presented in Table 7.2 and Table 7.3. The comparison of the performance and error as shown in Figure 7.2 and Figure 7.3.

As Experiment 1-3, they were learning HOG feature by MLFNN, which this HOG feature is the histogram of gradient orientation of edge of the object. The accuracy as show in Figure 7.2. The verification process can be observed as having a network that at stabilized at 99.73%, 87.65%, 85.12%accuracy for the AGV in factory environment, the electric senior vehicles, and the vehicle in traffics respectively.

Table 7.2: The result of the object detection experiments

Experiment	Accuracy(%)	False Positive (%)	False Negative (%)
1	99.73	0.07	0.20
2	87.65	7.14	5.21
3	85.12	8.56	6.32
4	99.73	0.27	0
5	91.27	5.18	.55
6	91.60	6.30	2.10
7	99.67	0.33	0
8	91.86	3.89	4.25
9	90.19	5.25	4.56
10	99.80	0	0.20
11	93.20	4.40	2.4
12	96.67	3.3	0
13	69.23	11.16	19.61

Table 7.3: Average overall computational time (Object detection for vehicles in traffic case)

Feature	Detector	Accuracy (%)	Time (ms)
Method 1	MLFANN	85.12	445.3
Method 2	TDNN	91.6	653.1
Method 3	TDNN	90.19	608.7
Method 2 & 3	TDNN	96.67	720.4

However, after examining accuracy from Experiment 4-6 which were learning HOG feature by TDNN, which this HOG feature is the pattern of the difference of HOG feature of the object in sequence of images and the input are set of delay from video images. This extraction method could efficiently training better than learning HOG feature in single frame method. Moreover, the accuracy of the experiment of the electric senior vehicles and the vehicle in traffics were increased to 91.27% and 97.6% respectively, but the AGV testing is as comparable performance with all method.

Experiment 7-9 were learning the pattern of shape variation ratio of the object by TDNN. The input sequence of images and the input are set of delay as in Experiment 4-6. The accuracy of the electric senior vehicles was slightly increased to 91.86%. However, the vehicle in traffics experiment slightly reduced to 90.19%,

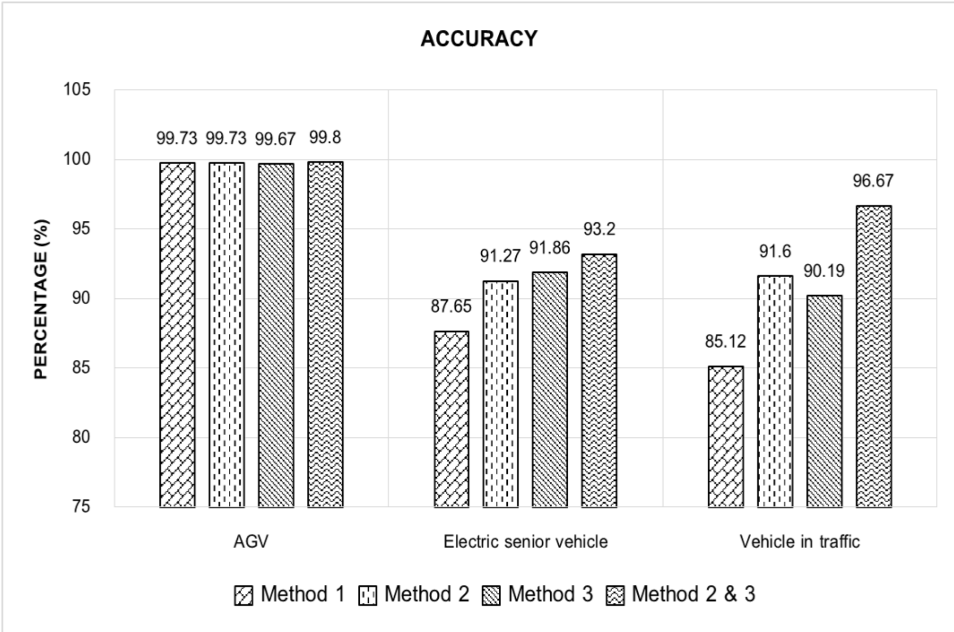


FIGURE 7.2: The accuracy of object detection of all experiment

but still more than the first method. Based on this results, Method 2 and 3 are similarly effective.

Finally, Experiment 10-12 were learning both of the pattern of the difference of HOG feature and the pattern of shape variation ratio, all experiments were dramatically improved comparing to the first experiment. The accuracy of the electric senior vehicles was up to 93.20%, and the vehicle in traffics reached to 96.67%, whereas the performance rates of Experiment 1 were only 87.65% and 85.12%.

From the results of the detecting of electric senior vehicle, METHOD 1 is the highest error both of false positive and false negative. The METHOD 2 can reduce the error to 5.18% and 3.55% respectively, whereas the error of Experiment 1 were 7.14% and 5.21%.

Method 3 can reduce the false positive error to 3.89%, in contrast, the false negative increase to 4.25%.

METHOD 4 reduce the false increase to 4.4%. However, the false negative reduce to 2.40%.

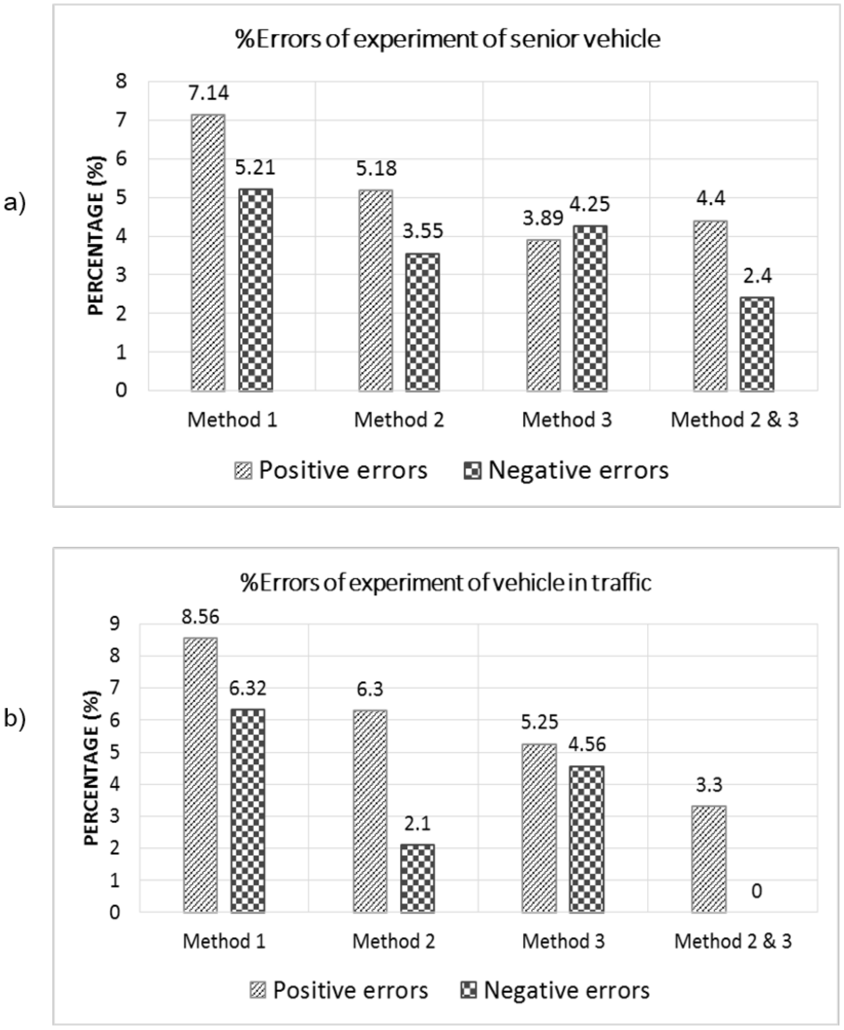


FIGURE 7.3: Illustration of %errors of all object detection experiment; (a) the experiment for electric senior vehicle; (b) the experiment for vehicle in traffic

For vehicle in traffic have the errors in the same trend as electric senior vehicle as shown in Figure 7.3.

Vehicle speed is one of the important factors that affect accuracy in real-time detecting by TDNN. In vehicle in traffic experiment, the vehicle speed range 45-60 km/h is relatively high accuracy in the range of 95.78 - 96.67%. However, the detection accuracy decreases when vehicle speed decreases as show in Figure 7.4.

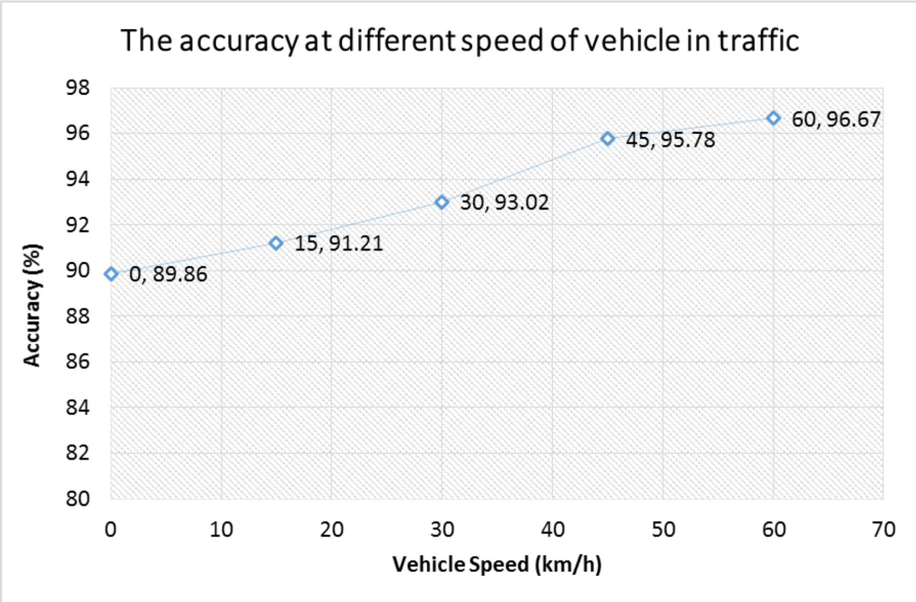


FIGURE 7.4: The accuracy at different speed of vehicle in traffic

7.2.2 Discussion

I proposed a classification and detection by improved my idea by using HOG feature to be input of the network.

I conducted twelve tests to evaluate the performance of my method (Table 7.1). The author calculated several performance rates, i.e., accuracy, false positive, and false negative. The experiment representing my method were Experiment 9-12 is to combined the pattern of the difference of HOG feature and the pattern of shape variation ratio as input into the object classification and detection.

From the result as presented in Figure 7.2 and 7.3 show that: although METHOD 1 provided the lowest accuracy, including the highest error, but highly accurate in case of AGV, because of the object is a simple shape. In case of the object as a simple shape, when the camera shoots a real object from the front view, the height of the object is vertical line based on the principles of an orthographic projection as the author mentioned in chapter 1. In contrast, the fake object image is no border in the vertical line based on construction of perspective viewing. As this result, the histogram of the gradient orientation of both object is distinctly different, which making the detection highly effective. This summary is based on the results of the AGV testing is as comparable performance with all methods, although METHOD 1 given the lowest performance in other cases; thus, the learning the difference of HOG feature of the object in single image is inappropriate to solve this study problems under complex shape, but it can used to learning the difference of HOG feature of the object under simple shape object as in AGV case.

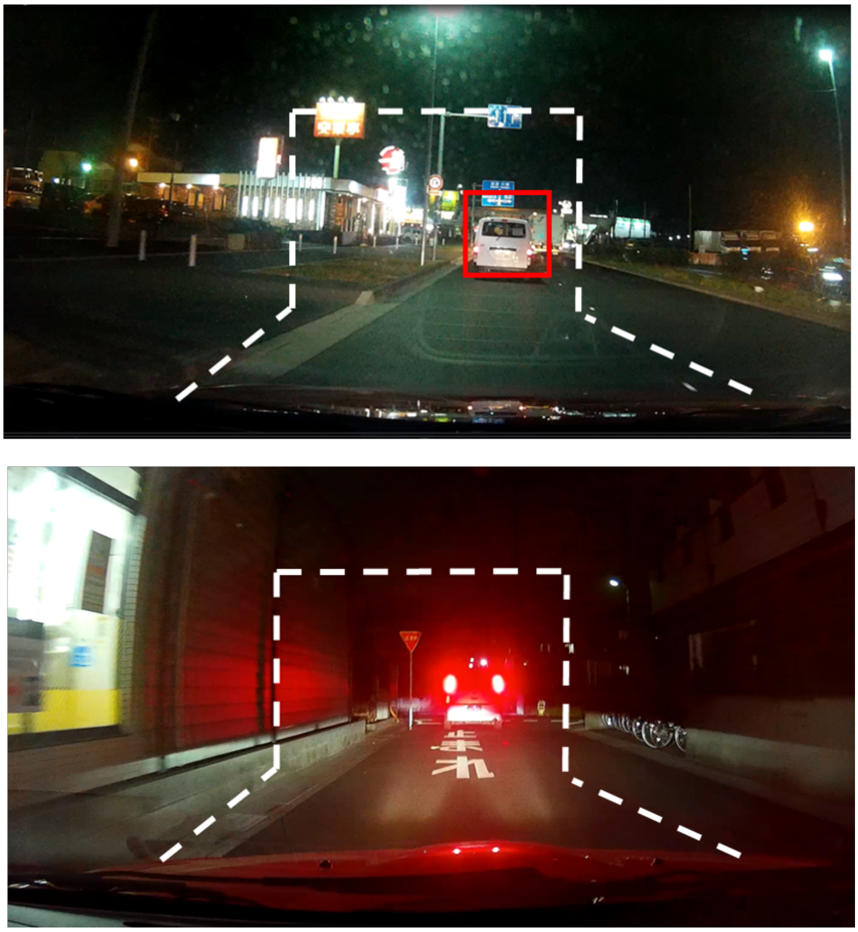
For METHOD 2, the author observed that the errors reduced from the first method. Moreover, the accuracy increased. This positive situation happened because the author changed the feature that use be input from the difference of HOG feature in single image to the difference of HOG feature in video images.

However, METHOD 3 gives a similar accuracy to METHOD 2, but there are more false negative, in contrast, false positive decreased.

METHOD 4 provided highest accuracy and lowest errors. To compare between METHOD 1, 2 and 3, the results showed that the feature selection and extraction affects the performance of the system, which it is a combined the pattern of the difference of HOG feature and the pattern of shape variation ratio as input into the object recognition by learning in sequence of images, not only learning in single image.

In Table 7.3 show the overall computational cost, the fastest algorithm is learning HOG feature (METHOD 1) by MLFANN (445.3 ms), whereas lowest accuracy. METHOD 2 and METHOD 3 that learning by TDNN are similarly effective, which more accuracy and the computational cost goes up as well. The most effective algorithm is learning the combination of METHOD 2 and METHOD 3 by TDNN. The implementation of this algorithm has a highest computational cost (720.4 ms). The results of this test indicate that improved detection accuracy has been accompanied by increased computational costs.

Usually, the method in this dissertation focuses on detecting objects in daytime. However, this method can be used at nighttime as show in Figure reffig:7-5(a), but the detection accuracy is reduced to 69.23%, with a false negative at 19.61% in case of vehicle in traffic as presented in Table reftab:7-2. For example; some cases have cars in front of vehicles, but cannot detect as show in Figure reffig:7-5(b).



a)

b)

FIGURE 7.5: The result of object detection testing in nighttime; (a) the obstacle detected; (b) The detection miss an obstacle

All method have false positive, which it can detect the fake object as the real object. Although it will not cause damage, this system is not suitable for use with an automatic braking system because it can cause an accident with a vehicle that follows behind it. Therefore, this system provides a warning to the driver when there is an imminent collision in order to prevent an accident and reduce the severity of a collision. Those actions may start with warning the driver, such as through a flashing dashboard icon, a beep, or a tug from the seat belt.

The results of this experiment show that it can detect general objects, and is not restricted to vehicles, objects or pedestrians, , which do not know the exact shape, size and color. as shown in Figure 7.6.

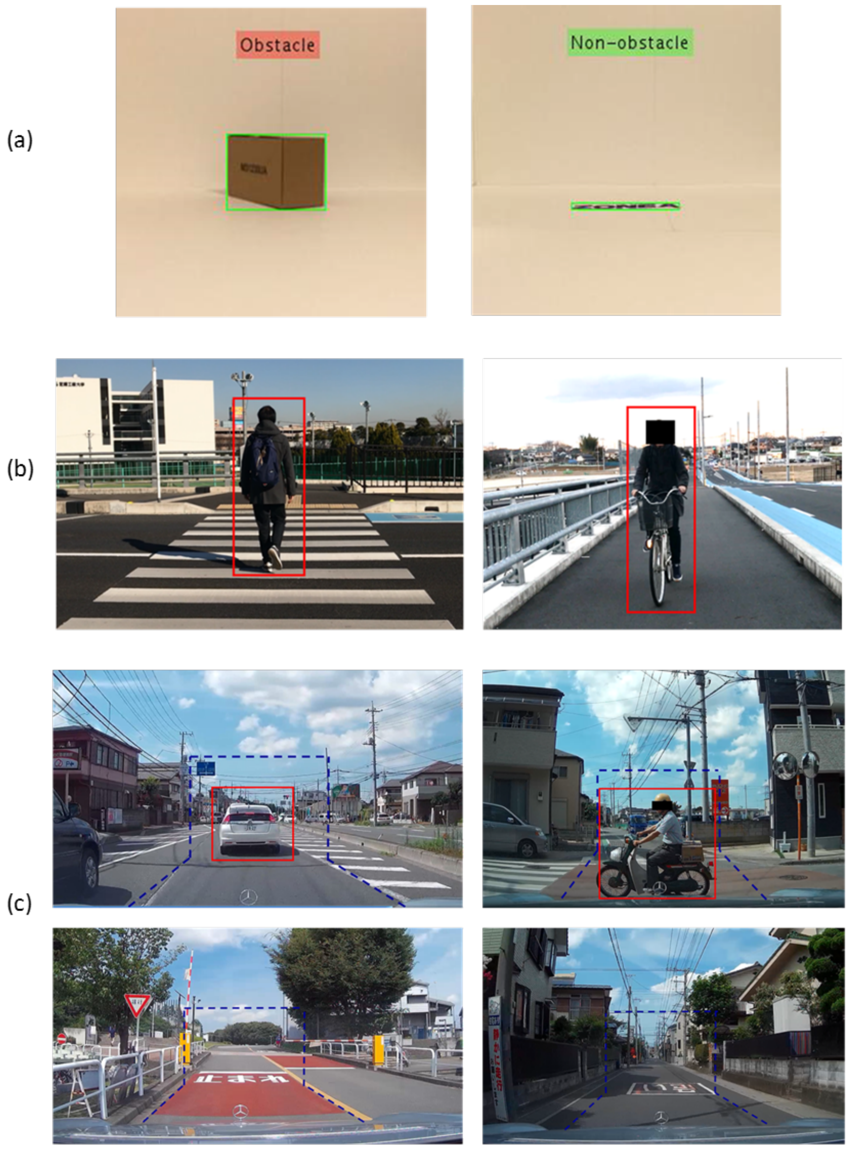


FIGURE 7.6: The result of object detection testing by using actual video images; (a) AGV in indoor scene; (b) electric senior vehicle; (c)vehicle in traffic

Chapter 8

Conclusions and Future Works

In this final chapter, the author summarize my dissertation. Moreover, the author will imagine the possible future work.

8.1 Conclusions

This dissertation presented study on object detection using computer vision by ANN, which can classify the general objects that are real objects or fake objects such as a painting, sign or text on the road, including proposing feature analysis, several methods for feature extraction, classification, and detection.

My main objectives of this dissertation were presented as follows:

1. To select the appropriate features for the problem.
2. To distinguish the object types.
3. To suggest a new solution of object classification and detection.
4. To design object detection method.
5. To extract feature of the object in image.

6. To create a prototype of object detection system.
7. To evaluate several methods for object detection as proposed in this dissertation.

I principally addressed the problem of object detection in real-time by computer vision. The author conducted several experiments and evaluated the obtained results.

It clearly showed that the system can identify and extract feature of the object in image. Moreover, the system was included object classification and detection. As the results, my system can detect general object, which can classify the objects that are real object or fake object.

To achieve the objective of the system presented in Chapter 3, the author introduced analyzing of objects to find the appropriate attributes to classify objects. Based on the result of analysis of the characteristics of each object, the object feature that can be used to classify the obstacles that are real obstacles or fake obstacles as follows:

1. Comparison of the characteristics of the object edge base on orthographic projection (METHOD 1). From orthographic projection and construction of perspective viewing, the height of the real object is vertical line and the fake object image is no border in the vertical line.
2. Comparison of the difference of the edge orientation in each frame (METHOD 2), which the difference of the orientation of the edges the real object is very small compared to the fake object.
3. Comparison of the shape variation ratio by calculating the ratio between the width and height of the object (METHOD 3), which the fake object has a shape variation ratio over the real object.

For the feature extraction presented in Chapter 4, the author proposed HOG method to extract the feature of the object from the analysis in Chapter 3-all of the three methods.

To overcome the goal of the system presented in Chapter 5, the author designed an object recognition and classification by MLFANN and TDNN. The author conducted experiments to evaluate my method. The author divided them into twelve tests three case, i.e., the AGV in factory environment, the electric senior vehicles, and the vehicle in traffics. The author designed the method to learn the object feature in four methods:

1. To learning the difference of HOG feature of the object in single image (METHOD 1), which it is comparing of the different of gradient orientation of the object edge in single image by MLFANN.
2. To learning the pattern of the difference of HOG feature of the object in sequence of images (METHOD 2), which it is comparing the difference of the orientation of the edges between two image. The real object is very small different changes when compared to the fake object by TDNN.
3. To learning the pattern of shape variation ratio of the object (METHOD 3), which the fake object has a shape variation ratio over the real object by TDNN.
4. To learning the combination of METHOD 2 and METHOD 3 by TDNN.

To fulfill my target of the research in Chapter 6, the author proposed the object detection based on HOG feature by using MLFANN for learning a single data and TDNN for learning the video images. The author designed to use the detector same algorithm with the classifier as follow:

1. MLFANN detector used to detect object in case of the learning feature by MLFANN classifier.
2. TDNN detector used to detect object in case of the learning feature by TDNN classifier.

Finally, the author integrated all implemented systems into one main system to detect general object in front of a vehicle.

In conclusion, the author proposed the systems for vehicles to detect general objects, which can classify objects that are real objects or fake objects. This system applicable for both lane and non-lane based traffic scenarios focused mainly on the ROI in front of the vehicle. The most accurate method in this dissertation is the detection of general objects by using HOG extractor to extract feature of the object, which it is a combination of two feature between the pattern of the difference of HOG feature and the pattern of shape variation ratio as input into the object recognition by learning in sequence of images, not only learning in single image as input into the object recognition and detection by TDNN.

Based on the findings, it is more efficient to learn sequential images than to learn single image. Moreover, the recognition by learning the combination of features has more effective than learning only one feature. The TDNN has the potential of learning to overcome the limitations of an MLFANN, and complete image sequences at a time instead of a single image, which it can work with complex data efficiently.

The results of this experiment show that we can detect obstacles of various sizes, shapes and colors, which is not restricted to the vehicles, objects or pedestrians. The distance from the real vehicle to the object that is used to classify the obstacles are up to 50 m, which the vehicle can brake without a collision. Therefore, this method can be used to improve object function classification accurately and efficiently for vehicles by using TDNN in the sequence of video images.

This method has false positive, which it can detect the fake object as the real object. Although it will not cause damage, this system is not suitable for use with an automatic braking system because it can cause an accident with a vehicle that follows behind it. Therefore, this system can be applied to provide a warning to the driver when there is an imminent collision in order to prevent an accident and reduce the severity of a collision. Those actions may start with warning the driver, such as through a flashing dashboard icon, a beep, or a tug from the seatbelt.

8.2 Future works

Further study of the issue would be of interest. To outline the directions of future work, in the meantime, deep learning has been proving useful for a variety of scientific tasks. Currently, the application of deep learning in a variety of fields such as personal image search, vehicle detection for forward collision warning, traffic sign detection, etc. Several works have been developed using deep learning neural networks for image analysis applied to obstacles detection and recognition. Deep learning is one of a machine learning methods, which try to learn how to represent data efficiently. The deep learning method that has the potential to deal with features for learning for unsupervised or semi-supervised feature learning and hierarchical feature extraction. Deep learning has the property that if you feed it more data, it gets better and better. (William Stafford Noble). The deep learning has potential to overcome limitations of shallow learning, e.g., MLP-neural network.

Deep learning is one of a machine learning methods, which try to learn how to represent data efficiently. The deep learning method that has the potential to deal with features for learning for unsupervised or semi-supervised feature learning and hierarchical feature extraction.

In 2012, Google with Andrew Ng researchers co-founder project "Google Brain", and propose the research about deep learning by extracted some 10 million still images from YouTube videos and fed them into Google Brain, a network of 1,000 computers programmed to soak up the world much as a human toddler does. After three days looking for recurring patterns, Google Brain decided, without guidance from the supervised, there were certain repeating categories it could identify these are human faces, human bodies or cats. This network had never been told what a human face was, what a human body was, what a cat was, nor was it given even a single image labeled as a human face, a human body, or a cat. That's what we mean by self-taught learning.

The localize objects in an image, which predicts multiple bounding boxes at a time complicated and very challenging. Erhan [18] proposeda deep CNN and using feature extraction and learning model to localizing objects in an image, it is able to capture multiple instances of objects of the same class, this method results in accurate and highly effective.

Recurrent convolutional neural network used for object recognition was proposed by Ming Liang [34] , which basic idea was to add recurrent connections within every convolutional layer of the feed-forward CNN for enhanced the capability to capture statistical regularities in the context of the object. This method can increase efficiency and accuracy in recognizing objects than CNN by incorporating recurrent connections into each convolutional layer. Ouyang et al., [44] proposed deformable deep convolutional neural networks for generic object detection, which can increase the mean averaged precision obtained by RCNN by restructure mechanism training. The increase and decrease some of the major components of the model series with a larger variety would have been significantly enhances the performance of the model. And also many works using deep learning such as the traffic sign recognition using extreme learning classifier with deep convolutional features [68], and multi-stage contextual deep learning for pedestrian detection [67]

Thus, the method used in this paper can be further extended to detect the obstacles in bad weather such as fog or rain by deep learning. Along with the principle of deep learning neural networks to be design the algorithm to improve the detection of objects. In particular, a detection of moving objects is available to realize safer path planning, a form of informative support for the driver. We expect this research will lead to more control over vehicles to avoid oncoming obstacles automatically and efficiently.

Appendix A

List of Publications

A.1 International Journal Papers

[J.1] Varagul, J. & Ito, T. (2017). General Object Detection Method by On-Board Computer Vision with Artificial Neural Networks. International Journal of Automotive Engineering, 8(4), 149-156.

A.2 International Conference Papers (Peer-reviewed)

[C.1] Varagul, J., & Ito, T. (2016). Improving function detecting object for AGV. In Doctoral Consortium-11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2016. SciTePress.

[C.2] Varagul, J., & Ito, T. (2016). Simulation of Navigation Control and Obstacle Avoidance for AGV by Computer Vision with Artificial Neural Networks. In Proceedings of the 10th South East Asian Technical University Consortium Symposium, SEATUC 2016

[C.3] Varagul, J., & Ito, T. (2016). Simulation of Detecting Function object for AGV using Computer Vision with Neural Network. In Proceedings of the 20th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, KES 2016.

[C.4] Varagul, J., & Ito, T. (2016). Object Detection Method for AGV by Computer Vision with Artificial Neural Networks. In Proceedings of the 23rd World Congress on Intelligent Transport Systems, ITS 2016.

[C.5] Varagula, J., & ITOb, T. (2017). Object Detection Method in Traffic by On-Board Computer Vision with Time Delay Neural Network. Procedia Computer Science, 112, 127-136.

[C.6] Varagul, J. & Ito, T. (2017, Oct). Object Detection Method by On-Board Computer Vision with Time Delay Neural Network. 2017 JSAE Congress (Autumn).

[C.7] Varagul, J., Kulprom, P., & Ito, T. (2017, Oct). Improving Function Detecting General Object for On-Board Computer Vision by Artificial Neural Network. In Proceedings of the 24th World Congress on Intelligent Transport Systems, ITS 2017.

[C.8] Varagul, J., & Ito, T. (2018). Object Detection Method for electric senior vehicle by Computer Vision with Artificial Neural Networks. In Proceedings of the 12th South East Asian Technical University Consortium Symposium, SEATUC 2018

Bibliography

[1] Anagnostopoulos, C. N. E., Anagnostopoulos, I. E., Loumos, V., & Kayafas, E. (2006). A license plate-recognition algorithm for intelligent transportation system applications. *IEEE Transactions on Intelligent transportation systems*, *7*(3), 377–392.

[2] Anumula, N. K., Katta, P., Basa, M., & Prasad, D. (2015). Open cv implementation of object recognition using artificial neural networks. *International Journal of Computational Engineering Research (IJCER)*, *5*(02).

[3] Aslani, S. & Mahdavi-Nasab, H. (2013). Optical flow based moving object detection and tracking for traffic surveillance. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, *7*(9), 1252–1256.

[4] Basharat, A., Gritai, A., & Shah, M. (2008). Learning object motion patterns for anomaly detection and improved object detection. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, (pp. 1–8). IEEE.

[5] Bertozzi, M., Broggi, A., Del Rose, M., Felisa, M., Rakotomamonjy, A., & Suard, F. (2007). A pedestrian detector using histograms of oriented gradients and a support vector machine classifier. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, (pp. 143–148). IEEE.

[6] Bertozzi, M., Broggi, A., Fascioli, A., & Nichele, S. (2000). Stereo vision-based vehicle detection. In *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, (pp. 39–44). IEEE.

[7] Blake, A. & Isard, M. (2012). *Active contours: the application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion*. Springer Science & Business Media.

[8] Broggi, A., Bertozzi, M., & Fascioli, A. (2001). Self-calibration of a stereo vision system for automotive applications. In *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, volume 4, (pp. 3698–3703). IEEE.

[9] Burrascano, P., Fiori, S., Frattale-Mascioli, F., Martinelli, G., Panella, M., & Rizzi, A. (2002). Visual path following and obstacle avoidance by artificial neural networks. *Enabling Technologies for the PRASSI Autonomous Robot"(S. Taraglio and V. Nanni, Ed. s), ENEA Research Institute*, 30–39.

[10] Cao, X., Wu, C., Yan, P., & Li, X. (2011). Linear svm classification using boosting hog features for vehicle detection in low-altitude airborne videos. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, (pp. 2421–2424). IEEE.

[11] Cireşan, D., Meier, U., Masci, J., & Schmidhuber, J. (2011). A committee of neural networks for traffic sign classification. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, (pp. 1918–1921). IEEE.

[12] Clode, S., Rottensteiner, F., Kootsookos, P., & Zelniker, E. (2007). Detection and vectorization of roads from lidar data. *Photogrammetric Engineering & Remote Sensing*, *73*(5), 517–535.

[13] Daigavane, P., Bajaj, P. R., & Daigavane, M. (2011). Vehicle detection and neural network application for vehicle classification. In *Computational Intelligence and Communication Networks (CICN), 2011 International Conference on*, (pp. 758–762). IEEE.

[14] Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, (pp. 886–893). IEEE.

[15] Ebrahimzadeh, R. & Jampour, M. (2014). Efficient handwritten digit recognition based on histogram of oriented gradients and svm. *International Journal of Computer Applications*, *104*(9).

[16] El-Gayar, M., Soliman, H., et al. (2013). A comparative study of image low level feature extraction algorithms. *Egyptian Informatics Journal*, *14*(2), 175–181.

[17] Enzweiler, M. & Gavrila, D. M. (2009). Monocular pedestrian detection: Survey and experiments. *IEEE transactions on pattern analysis and machine intelligence*, *31*(12), 2179–2195.

[18] Erhan, D., Szegedy, C., Toshev, A., & Anguelov, D. (2014). Scalable object detection using deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 2147–2154).

[19] Farfade, S. S., Saberian, M. J., & Li, L.-J. (2015). Multi-view face detection using deep convolutional neural networks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, (pp. 643–650). ACM.

[20] Fazli, S., Mohammadi, S., & Rahmani, M. (2012). Neural network based vehicle classification for intelligent traffic control. *International Journal of Software Engineering & Applications*, *3*(3), 17.

[21] Fleyeh, H. & Roch, J. (2013). *Benchmark evaluation of HOG descriptors as features for classification of traffic signs*. Högskolan Dalarna.

[22] Gehler, P. & Nowozin, S. (2009). On feature combination for multiclass object classification. In *Computer Vision, 2009 IEEE 12th International Conference on*, (pp. 221–228). IEEE.

[23] Green, B. (2002). Canny edge detection tutorial. *Retrieved: March*, *6*, 2005.

[24] Hariyono, J., Hoang, V.-D., & Jo, K.-H. (2014). Moving object localization using optical flow for pedestrian detection from a moving vehicle. *The Scientific World Journal*, *2014*.

[25] Hong, Z., Lin, F., & Xiao, B. (2017). A novel gravity-freak feature extraction and gravity-klt tracking registration algorithm based on iphone mems mobile sensor in mobile environment. *PloS one*, *12*(10), e0186176.

[26] Jin, X. & Davis, C. H. (2007). Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks. *Image and Vision Computing*, *25*(9), 1422–1431.

[27] Juan, L. & Gwun, O. (2009). A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, *3*(4), 143–152.

[28] Kaempchen, N., Franke, U., & Ott, R. (2002). Stereo vision based pose estimation of parking lots using 3d vehicle models. In *Intelligent Vehicle Symposium, 2002. IEEE*, volume 2, (pp. 459–464). IEEE.

[29] Karthik B., Hari Krishna Murthy, M. M. (2013). Identification classification and monitoring of traffic sign using hog and neural networks. *International Journal of Science and Research (IJSR)*, *4*(8), 782–786.

[30] Kocer, H. E. & Cevik, K. K. (2011). Artificial neural networks based vehicle license plate recognition. *Procedia Computer Science*, *3*, 1033–1037.

[31] Kong, H., Audibert, J.-Y., & Ponce, J. (2009). Vanishing point detection for road detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, (pp. 96–103). IEEE.

[32] Langner, F., Knauer, C., Jans, W., & Ebert, A. (2009). Side scan sonar image resolution and automatic object detection, classification and identification. In *OCEANS 2009-EUROPE*, (pp. 1–8). IEEE.

[33] Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011). Brisk: Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, (pp. 2548–2555). IEEE.

[34] Liang, M. & Hu, X. (2015). Recurrent convolutional neural network for object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 3367–3375).

[35] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, (pp. 1150–1157). Ieee.

[36] Luo, Y. & Duraiswami, R. (2008). Canny edge detection on nvidia cuda. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, (pp. 1–8). IEEE.

[37] Ma, L., Crawford, M. M., & Tian, J. (2010). Local manifold learning-based k-nearest-neighbor for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(11), 4099–4109.

[38] Maturana, D. & Scherer, S. (2015). Voxnet: A 3d convolutional neural network for real-time object recognition. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, (pp. 922–928). IEEE.

[39] Meng, H., Bianchi-Berthouze, N., Deng, Y., Cheng, J., & Cosmas, J. P. (2016). Time-delay neural network for continuous emotional dimension prediction from facial expression sequences. *IEEE transactions on cybernetics*, *46*(4), 916–929.

[40] Miao, Q., Wang, G., Shi, C., Lin, X., & Ruan, Z. (2011). A new framework for on-line object tracking based on surf. *Pattern Recognition Letters*, *32*(13), 1564–1571.

[41] Okuda, R., Kajiwara, Y., & Terashima, K. (2014). A survey of technical trend of adas and autonomous driving. In *VLSI Technology, Systems and Application (VLSI-TSA), Proceedings of Technical Program-2014 International Symposium on*, (pp. 1–4). IEEE.

[42] Ors, A. O. (2017). *Choosing the Optimum Mix of Sensors for Driver Assistance and Autonomous Vehicles.*

[43] Ortiz, J. E. C. (2012). Visual servoing for an omnidirectional mobile robot using the neural network-multilayer perceptron. In *Engineering Applications (WEA), 2012 Workshop on*, (pp. 1–6). IEEE.

[44] Ouyang, W., Wang, X., Zeng, X., Qiu, S., Luo, P., Tian, Y., Li, H., Yang, S., Wang, Z., Loy, C.-C., et al. (2015). Deepid-net: Deformable deep convolutional neural networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 2403–2412).

[45] Plantenberg, K. (2016). *Engineering Graphics Essentials Independent Learning Content @ Copyright 2016*.

[46] Quiles, M. G. & Romero, R. A. F. (2006). A computer vision system based on multi-layer perceptrons for controlling mobile robots. In *ABCM Symposium series in Mechatronics*, volume 2, (pp. 661–668).

[47] Ramisa, A., Vasudevan, S., Aldavert, D., Toledo, R., & de Mantaras, R. L. (2009). Evaluation of the sift object recognition method in mobile robots. In *CCIA*, (pp. 9–18).

[48] Ramzan, H., Fatima, B., Shahid, A. R., Ziauddin, S., & Safi, A. A. (2016). Intelligent pedestrian detection using optical flow and hog. *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, *7*(9), 408–417.

[49] Rauf, R., Shahid, A. R., Ziauddin, S., & Safi, A. A. (2016). Pedestrian detection using hog, luv and optical flow as features with adaboost as classifier. In *Image Processing Theory Tools and Applications (IPTA), 2016 6th International Conference on*, (pp. 1–4). IEEE.

[50] Ren, M. & Sun, H. (2005). A practical method for moving target detection under complex background. *Computer Engineering*, *31*(20), 33–40.

[51] Routray, S., Ray, A. K., & Mishra, C. (2017). Analysis of various image feature extraction methods against noisy image: Sift, surf and hog. In *Electrical, Computer and Communication Technologies (ICECCT), 2017 Second International Conference on*, (pp. 1–5). IEEE.

[52] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). Orb: An efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE international conference on*, (pp. 2564–2571). IEEE.

[53] Rybski, P. E., Huber, D., Morris, D. D., & Hoffman, R. (2010). Visual classification of coarse vehicle orientation using histogram of oriented gradients features. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, (pp. 921–928). IEEE.

[54] Salhi, A., Minaoui, B., Fakir, M., Chakib, H., & Grimech, H. (2017). Traffic signs recognition using hp and hog descriptors combined to mlp and svm classifiers. *Traffic*, *8*(11).

[55] Schuldt, C., Laptev, I., & Caputo, B. (2004). Recognizing human actions: a local svm approach. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, (pp. 32–36). IEEE.

[56] Sermanet, P. & LeCun, Y. (2011). Traffic sign recognition with multi-scale convolutional networks. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, (pp. 2809–2813). IEEE.

[57] Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning.

[58] Suard, F., Rakotomamonjy, A., Bensrhair, A., & Broggi, A. (2006). Pedestrian detection using infrared images and histograms of oriented gradients. In *Intelligent Vehicles Symposium, 2006 IEEE*, (pp. 206–212). IEEE.

[59] Suga, A., Fukuda, K., Takiguchi, T., & Ariki, Y. (2008). Object recognition and segmentation using sift and graph cuts. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, (pp. 1–4). IEEE.

[60] Szarvas, M., Yoshizawa, A., Yamamoto, M., & Ogata, J. (2005). Pedestrian detection with convolutional neural networks. In *Intelligent vehicles symposium, 2005. Proceedings. IEEE*, (pp. 224–229). IEEE.

[61] Uijlings, J. R., Duta, I. C., Rostamzadeh, N., & Sebe, N. (2014). Realtime video classification using dense hof/hog. In *Proceedings of International Conference on Multimedia Retrieval*, (pp. 145). ACM.

[62] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. J. (1990). Phoneme recognition using time-delay neural networks. In *Readings in speech recognition* (pp. 393–404). Elsevier.

[63] Wang, C.-C., Thorpe, C., & Suppe, A. (2003). Ladar-based detection and tracking of moving objects from a ground vehicle at high speeds. In *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, (pp. 416–421). IEEE.

[64] Wöhler, C. & Anlauf, J. K. (2001). Real-time object recognition on image sequences with the adaptable time delay neural network algorithmapplications for autonomous vehicles. *Image and Vision Computing*, *19*(9-10), 593–618.

[65] Wöhler, C., Anlauf, J. K., Pörtner, T., & Franke, U. (1998). A time delay neural network algorithm for real-time pedestrian recognition. In *International Conference on Intelligent Vehicle*. Citeseer.

[66] Yamada, Y., Matsuo, T., Shimada, N., & Shirai, Y. (2008). Classification of hand shape with similar contour for sign language recognition. In *The Fourth Joint Workshop on Machine Perception and Robotics (MPR2008). Beijing, China.* Citeseer.

[67] Zeng, X., Ouyang, W., & Wang, X. (2013). Multi-stage contextual deep learning for pedestrian detection. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, (pp. 121–128). IEEE.

[68] Zeng, Y., Xu, X., Fang, Y., & Zhao, K. (2015). Traffic sign recognition using extreme learning classifier with deep convolutional features. In *The 2015 international conference on intelligence science and big data engineering (IScIDE 2015), Suzhou, China*, volume 9242, (pp. 272–280).

[69] Zhan, C., Duan, X., Xu, S., Song, Z., & Luo, M. (2007). An improved moving object detection algorithm based on frame difference and edge detection. In *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on*, (pp. 519–523). IEEE.

[70] Zhang, G. et al. A pedestrian detection method based on svm classifier and optimized histograms of oriented gradients feature. natural computation (icnc). In *2010 Sixth International Conference on*, volume 6.

[71] Zhang, H., Berg, A. C., Maire, M., & Malik, J. (2006). Svm-knn: Discriminative nearest neighbor classification for visual category recognition. In *Computer Vision*

and Pattern Recognition, 2006 IEEE Computer Society Conference on, volume 2, (pp. 2126–2136). IEEE.

[72] Zhao, L. & Thorpe, C. E. (2000). Stereo-and neural network-based pedestrian detection. *IEEE Transactions on intelligent transportation systems*, *1*(3), 148–154.