Department of Mechanical Engineering

Vision-based Detection of Mobile Device Use While Driving

Woo Yew Meng

This thesis is presented for the Degree of
Master of Philosophy (Mechanical Engineering)
of
Curtin University

June 2019

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Human Ethics (For projects involving human participants/tissue, etc) The research presented and reported in this thesis was conducted in accordance with the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research (2007) – updated March 2014. The proposed research study received human research ethics approval from the Curtin University Human Research Ethics Committee (EC00262), Approval Number #RDSE-01-15

Signature: ………………………………………….

Date: ………………………...

Acknowledgements

I would like to thank my supervisor, Tele Tan, for his tireless guidance and encouragement throughout the entire course of the experimental work and in preparation of this dissertation.

I would also like to thank the Department of Mechanical Engineering and Curtin Information Technology Services of Curtin University for their generous provisions and support.

In addition, I am also grateful to my fellow researchers, Richard Leslie Palmer, Alex Kwang Leng Goh and Sawitchaya Tippaya for their interest, advice and assistance in this thesis.

Finally, many thanks to my parents, family members and friends who provided me with their active moral support and encouragement, for I would not have been able to complete this research without them.

Thank you.

Abstract

The use of mobile devices while driving has been correlated to the increase in road accidents. While law enforcement agencies are dependent on the existing traffic monitoring systems to deter user misconduct and monitor for road accidents, they are struggling to visually identify drivers using mobile devices. Currently, there are mostly indirect, non-visual based technologies to identify driver's mobile usage which requires significant additional infrastructure and may need trained expertise to use. This study aims to explore a vision-based detection system which works directly on the existing traffic monitoring platform, with minor setup adjustments.

An in-depth overview on the existing mobile usage detection solutions highlighting their functions and limitations is provided. The image processing literature is also reviewed to explore their capabilities in detecting mobile usage. Subsequently, a comparison of image processing toolkits is presented to identify suitability for prototyping the detection workflow. 2 online accessible image classification datasets are used for testing the detection system which was obtained from UCSD/Calit2 Car Licence Plate, Make and Model database and ImageNet database. Moreover, a data collection of mock drivers' mobile usage was carried out in order to supplement the lack of actual driver mobile usage images.

In this thesis, a hybrid visual detection solution is formulated by integrating multiple algorithms in the sequence of Licence Plate (LP) template matching, knowledge-based distance approximation, face detection and Region based Convolutional Neural Network (R-CNN) classifier. The solution aims to reduce search space and consistently identify mobile device usage under realistic circumstances. In the initial stage, LP template matching is used for vehicle identification and localisation. Subsequently, a knowledge-based algorithm applies the data to approximate LP to driver's face distance within the image. A face detection algorithm further pinpoints the exact driver's face position. Lastly, a region based neural network classifier is used to identify mobile device usage.

While the developed approach is only a theoretical model, it will be elaborated to demonstrate its utility in providing optimal analysis of driver's mobile usage and providing useful support to law enforcers in traffic monitoring. A summary of the experimental performance of the detection will be shown to illustrate the solution's capacity to function accordingly.

This research will prove that the suggested approach has the necessary design to function in a practical and optimal manner. The techniques, logic and outcomes will be specified in detail. Based on the validity of the experimental results, several practical approaches to improve the system are recommended.

Table Of Contents

List Of Figures

LIST OF TABLES

Abbreviations

2D	two-dimensional
ANN	Artificial Neural Network
CCTV	Closed-circuit television
CDPM	Cascade Deformable Part Based Model
CNN	Convolutional Neural Network
CVS	Computer vision systems
FD	Face Detection
FN	false negative
FP	false positive
GPU	graphic processing unit
IoU	intersection-over-union
ISO	International Standards Organisation
LBP	Local Binary Pattern
LP	Licence Plate
LPD	Licence Plate Detection
LP-F	Licence Plate to Driver Face
MCG	Multiscale Combinatorial Grouping
MDUD	Mobile Device Usage Detection
MVFD	Multi View Face Detector
OCR	Optical Character Recognition
RCNN	Region Convolutional Neural Network
ROI	region of interest
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TN	true negative
TP	true positive
WHO	World Health Organisation

1 Introduction

Driver distraction contributes significantly to serious road crashes. Two studies, one in Toronto, Canada [1] and one in Perth, Australia [2] show that crash risk was about four times greater when drivers are interacting with or using mobile related technologies. Some drivers are either unaware or unconcerned to stop using their mobile phone while driving. In Australia, traffic enforcers have gone undercover to spot mobile phone use by drivers and signalling their colleagues ahead of the road to stop them [3]. Furthermore, helmet cameras [4] and roadside speed cameras are being employed to capture drivers using mobile devices. However, this requires significant amount of preparation and policing hours. Traffic enforcers must manually search through hundreds of hours of video to identify the offenders. The detection process is time-consuming, labour intensive and expensive.

Using the existing monitoring infrastructure, this research study aims to help traffic enforcers to avoid spending excessive effort and costs while effectively detecting mobile phone using drivers. By efficiently capturing this errant behaviour in action, it may help change the behaviour of drivers when it comes to using mobile devices while driving. This potentially leads to a reduction in driver distraction and consequently lower fatality rate on the road.

This research aims to experimentally investigate the image processing technologies and techniques with algorithms capable of security and surveillance applications. Experimentations will be conducted on these research software tools as a basis to demonstrate a proof of concept system exhibiting the feasibility of the mobile device usage detection using image processing algorithms. This system is only expected to supplement the existing practices and not as a substitute.

1.1 Motivation

The incentive for performing this research is to help improve driver safety on the road. We view the costs and limitations of the current road user surveillance approaches as a factor to this issue. Hence, we attempt to use image processing algorithms to solve this.

Image processing algorithms have been used in other fields, such as in security [5] and factory automation [6], [7], where thousands of parts and products are inspected with high speed, precision and accuracy.

This image processing algorithm approach is not advocated as a total replacement, but to supplement existing CCTV systems and its users. Traffic law enforcers could use this approach to analyse large volumes of live CCTV feeds and even video recordings of drivers who are distracted by their mobile devices.

1.2 Aims

This research intends to provide a proof of concept for a vision-based detection of mobile device usage by vehicle drivers.

1.2.1 Exploration And Prototyping

Suitable hardware and software are first identified to produce the desired requirements of the system. Different state of the art methods will be investigated and incorporated where it is relevant and potentially useful towards the outcome of the proposed system.

1.2.2 Data Collection

Secondly, datasets of mobile usage by vehicle drivers that are appropriate for testing the system will be acquired. Different methods of generating, classifying and data testing will be used.

1.2.3 Development Of Mobile Device Usage Detection

Thirdly, a system that makes use of data collected and compiled to provide MDUD among vehicle drivers will be investigated and developed. This can serve to demonstrate the viability of using image processing techniques to incorporate and automate a process that is usually handled by human experts, to the extent that it can be operated by first time users.

1.2.4 Assessment

Finally, there is a need to test the methods and examine the outcome of the system. This step is to determine the feasibility and limitations of the proposed system.

1.3 Problem Description And Scope

In this context, using mobile devices while driving is defined as the action of the driver holding his/her phone in any way (i.e. in hand, between shoulder and ear.). This is regardless of whether the vehicle is moving or stationary. The proposed system will focus on detecting mobile device usage during driving.

The use of mobile devices while driving is a major contributor to driver distraction and road accidents globally. According to the WHO’s 2015 Global Status Report On Road Safety [8], more than 100 countries in the world have laws which ban the use of handheld mobile device usage while driving. These include both developed and developing countries, such as Australia, United Kingdom, China and India. Detecting and identifying the mobile device usage by drivers is the primary step to deter risky driving behaviour.

More specifically, this thesis will look at developing image processing methodologies that is on par or even outperform the existing practices in terms of accuracy and processing capabilities. This will need investigation of suitable techniques for locating the moving vehicle, the vehicle driver within the vehicle as well as detecting the presence or absence of the mobile device usage. However, the challenge may come in detecting various possible forms of usage interactions.

Hence, this thesis also attempts to use a realistic dataset which encompasses all possible forms of mobile device usage interactions for training and testing the system. This will require identification of existing datasets or the creation of a new dataset which suits the purpose of the research. The system should be flexible for input of additional data and modular for combination of various image processing techniques.

There are some practical considerations to be made when selecting the type of dataset used for testing the prototype. The required processing capabilities and data storage depends on whether the dataset is in terms of video or image sequences. On the other hand, there are privacy issues when acquiring video data of road users. There are also safety concerns when capturing videos of drivers using their mobile devices.

To verify the effectiveness of the proposed detection module, a comprehensive testing dataset is required. An online search query was performed to locate existing image datasets for the required purposes. A set of LP and non-LP image datasets is located and available for research use. However, no unique image datasets exist for driver mobile usage.

Hence, a setup is created to collect the dataset for driver mobile device usage under various conditions. The setup will consist of a digital video capturing device located at the side of the road aimed towards a stationary car. The test subject is positioned in the driver’s seat.

The setup will simulate various conditions as follows:

A) Driving pose (including possible driver mobile device usage pose)

B) Lighting at different times of the day: 6am, 12pm, 6pm.

C) Occlusions (shadows, tinted glasses, sunglasses, mobile device, hand, face orientation, hair)

The dataset will be collected to be evaluated by the proposed detection pipeline.

The setup is designed to assist users by generating relevant information to reduce user's workload. There is a huge amount of data involved and this potentially overloads a user's processing capability, which may result in incomplete and erroneous detection. In addition, more data can be processed faster and more efficiently. In addition, the system would be made user-friendly, regardless of proficiency level of users to quickly understand, assimilate and translate the information into useful action. In an ideal case, stakeholder consultation and user interface development would be considered, but such endeavours are beyond the range of tasks that the research intends to undertake and instead could be considered in upcoming research projects.

Basically, the final system should be able to capture all mobile device related driver distractions and gives a result summary to the traffic enforcers.

In conclusion, the application is to provide a vision based detection system of driver's mobile device usage for traffic enforcers.

1.4 Thesis Scope

The range of studies to be conducted can be extensive and broad; the focus is on the practical application of a range of image processing algorithms, specifically in the context of licence plate, face and mobile device usage. All the algorithms involve a large bulk of work, so within them, specific knowledge most useful to integrate the three parts together will be distilled. In the following chapters, the details of the work and the processes applied to exhibit its capabilities will be outlined.

In LPD, current studies on the effectiveness of LPD's role in the surveillance industry will be investigated. Some of the techniques by which researchers have used to add value to the existing processes will also be examined. Finally, the usability of LPD for our system will be demonstrated.

In the area of FD, a range of existing research for FD under a variety of realistic conditions will be researched. While a large body of work has already been done in the area, it does not solve the issue of differentiation between driver and passenger faces. This can be resolved by introducing a region of interest given by the relationship of the licence plate location and the vehicle driver location.

In the area of mobile device usage detection, MDUD itself, the image processing related research done will be examined. As there has been some amount of work done in this novelty area, existing methods from other parts of image processing will be introduced and applied.

The integration in LPD, FD and MDUD in the detection system will be discussed. The methodology for determining the effectiveness of the system compared to the current surveillance methodologies will then be evaluated and assessed. Lastly, to present the potential of this research, the results of our proof of concept system will be presented.

1.5 Thesis Structure

This dissertation contributes toward the area of applied image processing. Particularly, it contributes to rigorous testing of established algorithms with experimental data, introduces a novel method to the field and applies a machine learning system to detect a novel object.

The structure of the thesis is as follows:

The different perspectives of image processing research are dealt with in chapter two. This chapter first examines the motivation and background in which the investigation is based on, which provides the context for the application of computer vision technology. The chapter also broadly discusses the types of existing detection mechanisms and image processing techniques. Lastly, this chapter presents a survey of the existing software toolset and selects suitable algorithmic components for the detection module.

Chapter three goes through a method for testing the performance of a selected licence plate detection algorithm. This chapter starts by describing the processes of the LP detection algorithm and the datasets used to test it. The performance of LP detection is then identified and exhaustively reviewed.

Similarly, chapter four details the methods for a novel knowledge-based estimation model and a popular face detection algorithm. The chapter first identifies the detection challenges faced and presents a set of solutions. A suitable dataset and detection criterion is then used to test the solutions. A comparison of their performance is presented and discussed.

Chapter five covers the mobile device usage detection method based on Region Convolutional Neural Network (RCNN). In this chapter, the RCNN algorithm is explained, followed by the introduction of a refined training and testing datasets. Then, the appropriate performance criterion is outlined, and the subsequent results is provided and analysed.

Chapter six presents the discussion on the importance of this work, conclusions and future work considerations.

2 Literature Review

To provide a practical framework, the review attempts to broadly identify with the challenges that traffic enforcement face. How does one visually determine the driver’s identity? How does one differentiate between drivers and passengers? How does one visually determine mobile usage behaviour? The literature review is formulated based on these leading questions.

In this chapter the main papers that generate some perspective for this research will be reviewed, as well as relevant papers in other fields. The chapter is further divided into several main sections. Each section consists of a topic, though there are some studies which consist of several topics which are appropriately rearranged under the same topic.

The readers will be introduced to the different detection mechanisms being developed and used in the surveillance industry and their backgrounds. In addition, the existing imaging systems will be discussed, followed by an introduction to the major image processing techniques such as Licence Plate Detection (LPD), Face Detection (FD) as well as machine learning techniques such as Convolutional Neural Network (CNN). The strategies for mobile device usage detection (MDUD) itself will also be discussed.

Lastly, the relevant techniques and processes which will be used in this research are selected.

2.1 Comparison Of Detection Mechanisms

There are many different methods that are being developed to detect and control mobile device use while driving. One of these methods is to send jamming signals to the mobile phone when vehicle gear shifts are detected [9]. Another technology leverages the existing car stereo infrastructure to discretely identify driver/passenger phone use [10]. Alternatively, the location and movement of mobile phone use can be detected via its built-in GPS and accelerometer sensors [11].

One of the key problems of the existing methods is that it either does not distinguish or pose difficulty in identifying between vehicle driver and other vehicle occupants. It is important to be able to differentiate and classify between the vehicle driver and passengers because of economic and safety implications. Passengers should be allowed to use their mobile devices. Another major factor is the setup cost. Some approaches require additional setup for vehicle manufacturers and vehicle drivers. Furthermore, these may also depend on explicit vehicle driver’s cooperation. Unless there is economic standardisation and legal intervention, these problems would present themselves as barriers to implementation.

Hence, the criteria for selecting an approach to mobile usage detection are to be able to distinguish vehicle driver from other passengers, minimal installation cost and do not require explicit vehicle occupant consent.

Currently, traffic law enforcers use live video cameras to monitor traffic conditions, which can be incorporated with computer vision systems (CVS) to automate the processing of the video stream. CVS use different pattern recognition algorithms and do not require humans to operate. They have been shown to be able to effectively keep track of vehicles in real time [12] . Hence, the visual detection of mobile devices could be used as evidence of distraction.

2.2 Image Processing Techniques

This section reviews recent studies, published in the literature, regarding MDUD based on image processing techniques. The studies are directly linked to this thesis due to their potentially useful techniques for detecting driver's mobile device usage. The following chapters are divided into context specific algorithms.

2.2.1 Background Subtraction

From the input video provided, the region of interest for a moving vehicle can be deduced by comparing the differences between two consecutive image frames [13]. This should help to reduce the search space required for potential region of interest, potentially optimizing computational resources for more important uses.

2.2.2 Licence Plate Recognition

According to the state-of-the-art reviews on Licence Plate Recognition (LPR) [14], there are generally four stages in the following sequence, namely image acquisition, LPD, licence plate segmentation and character recognition.

After acquiring an image, LPD can be carried out by using one or more of the following features: boundary/edge information, global image information, textual features, colour features, and character features. Features can be combined to improve the reliability of the detection although it may be computationally complex [14]. Similarly, licence plate segmentation uses pixel connectivity, projection profiles, and prior knowledge of characters and/or character contours. In character recognition, there are two broad methods: pixel intensity values (such as template matching) and extracted features (such as Gabor filter). Numerous studies [15], [16], [17], [18] have demonstrated real time LPR with a relatively robust detection rate for different licence plate types.

2.2.2.1 Related Work

Over the years, various LPD techniques have been put forward to detect and recognise vehicular licence plates and different image database have been examined. Hence, as noted by a 2015 comparative study [19] , under carefully crafted conditions, a majority of the techniques work well. These techniques are specific to the different stages of the LPD process. A comprehensive literature review is provided on these techniques in the following sections.

2.2.2.2 Licence Plate Detection

LPD starts by determining the location of LP in an image. This is the most critical part of LPD. Like a fingerprint that is unique to a person, there is certain information in an image that is looked for and identified as the LP. In order to automate this identification process, researchers have come up with algorithms that identify these essential LP features, which are generally categorised [14] as follows:

Border/Boundary. The first class of algorithms is the most straightforward approach as it is based on identifying edge features. A licence plate has a rectangular edge of a regular dimension, where its length and width are proportionally pre-determined. This shape is relatively unique to the LP, which is different from, for example, the windshield or the tyres of the vehicle. Edge detection techniques can be trained to detect this shape which fits the appropriate size.

Global Image. There are a set of algorithms which determine the licence plate location by looking at the whole image and distilling the minimum information required [20]. Connected component analysis is one such algorithm which does this by grouping pixels of similar densities into regions [21]. By doing so, the spatial relationship between different components, such as the licence plate location, can be identified. However, this may lead to incomplete licence plate localisation [19] due to the variation in images.

Texture. Licence plates can also be detected by its texture features. The texture of the licence plate is mostly unique, when compared to the texture of other vehicle parts. The techniques require prominent grey scale level differences between the licence plate character and its background. It also means that there is usually a higher concentration of boundary transitions where there are licence plate characters. There are different algorithms [22], that have been researched to exploit the characteristics of these texture features.

Colour. The distinctiveness of the licence plate colour with respect to the vehicle is also utilised as a feature to be detected. Neural network [23], genetic algorithm [17] and fuzzy based [24] techniques are employed. However, this application is limited to certain places due to factors such as illumination and noise sensitivity, as well as the diversity of colour usage [25].

Character. This class of algorithms directly detect characters that match those on a licence plate. The approach maintains relatively high accuracy despite rotation, illumination and partial occlusion [26]. However, the process is time consuming [26] and may detect non LP characters [27].

Hybrid. In order to improve on the accuracy of the LP detection algorithms, most researchers have resorted to hybrid algorithms which uses two or more LP features [28]. Hybrid algorithms are proven to be more robust but are typically expensive to compute. This is because different combinations of machine learning techniques are usually invoked such as neural networks and fuzzy based classifiers [29]. In addition, spatial/frequency domain filtering and connected component labelling [30] are used to identify both edge and colour details of the licence plate. By calculating the covariance of the different LP features used, the position of the licence plate is deduced [31].

2.2.2.3 Licence Plate Segmentation

Assuming that the licence plate is properly located, the next phase is to further separate the detected region into possible individual character regions.

Pixel Grouping. One type of algorithm [32] which detects individual characters are separate groups of pixel. However, the quality of individual characters is dependent on the image quality. Segmentation may fail when multiple connected characters are detected as a single character or a single broken character is detected as separate characters.

Projection Profiling. Another computable feature is where the colour variation between licence plate background and character allows for distinct binary values to be detected when the colours in an image is segregated and transformed into black and white colour. In projection profiling [33], the plate region is scanned vertically and horizontally to separate the character based on binary difference. However, considerable amount of noise [34] or rotation can result in segmentation failure.

Prior Knowledge Of Character. Since LP characters are standardised in a unique manner, certain assumptions/expectations can be made. For segmentation purposes, these algorithms take advantage of those key features such as character-background pixel ratio [35], character dimensions [36], fixed character amount [37]. The method is straightforward but may be location specific as there may be different LP standards in different locations.

Character Outline. One aspect of characters is that they have clear and finite perimeters, which means they can be identified by their shape using contour following algorithms [38]. For example, the fast marching algorithm [39] which is used in the shape driven active contour model [40] can concurrently detect and classify a character. Similar to pixel grouping algorithms, poor image quality could result in detecting broken or misshapen outline.

Hybrid. Multiple algorithms can be combined to improve proper segmentation. For example, fragmented characters joined before segmentation using a histogram-based algorithm whereas overlapped characters and connected characters are differentiated with morphological thickening algorithm and morphological thinning algorithm respectively. In [41], their method automatically finds broken, superimposing and linked shapes and applies the relevant algorithm.

2.2.2.4 Optical Character Recognition

After character segmentation, the extracted parts are classified in a process known as Optical Character Recognition (OCR). Some of the main issues at this stage are non-uniform dimensions [42] , fragmentation, style differences, and rotation [43]. The features used by the OCR algorithms become the basis for the following classifications.

Shape Similarity. An easy and direct approach to OCR is template matching [44]. This method examines the correlation between the extracted characters and the templates and matches the character with the least different template. After scaling the segmented characters to template size, template matching takes place. In a research done by Naito [45] , additional character templates is found to improve inclined character recognition at the increased cost of processing time.

Prominent Features. Identifying salient features in segmented characters and then classifying them is another OCR technique. Some of the feature extraction methods include projection profiling [33, 46], Hotelling transformation [47], ‘chunking and counting’ [48, 49] and ‘character-background’ transition frequency from character contours [50]. Machine learning techniques such as hidden Markov model [21] are employed after feature extraction for classification purposes. While feature extraction can be time consuming, these techniques are relatively more robust compared to template matching.

On the other hand, LP detection can benefit from using a sequence of image frames instead of just relying on individual frames. LP localisation within video streams shows up to 32% improvement over use of single images [51].

2.2.3 Face Detection

In 2001, Paul and Viola [52] used a range of techniques, namely integral imaging, Adaboost, and Haar Cascade to develop a visual object detection process which reduces computation time while achieving high detection accuracy. Their technique allows for the robust and real time processing of frontal upright face detection. In addition to Viola Jones technique, other face detection algorithms use complementary features [53] and hybrid methods of soft computing tools such as Artificial Neural Network, Support Vector Machine (SVM) and Gabor filter to become more efficient and robust. Furthermore, Cascade Deformable Part Based Model (CDPM) can be used to create a more reliable Multi View Face Detector (MVFD) [54]. This helps when the angle of the camera to the driver's face is not perpendicularly aligned. Another challenge in face detection would be occlusion. The face of a driver encounters a variety of possible occlusions, such as sunglasses, fingers, steering wheel and the mobile device itself. Previous studies have shown different techniques for occlusion detection by detecting the motion cue between the face in the video sequence and the occlusion with the background [55], [56]. However, there is a reasonable possibility that the image of the drivers captured would show no or little motion cue detected between the face and the occlusion with the background. In addition, by introducing an occlusion dictionary [57], the Gabor feature-based sparse representation in image processing is robust to occlusion. However, its real time performance is unknown.

2.2.4 Ear Detection

Similarly, the ear can be detected with high precision [58] using cascaded Adaboost. The detection of the ear is crucial as most drivers with mobile device usage would possibly occlude the ear. Also the detection of the ear would help provide further evidence for locating the driver's face [59]. The multi-scale ear detector is suitable for real-time surveillance and non-intrusive biometric applications [58].

However, the absence of ear may be attributed to factors other than the presence of mobile devices. Occlusion by head angle or hair [58] could be false positives instead of actual telecommunication usage.

2.2.5 Poselets

An alternative for detecting driver mobile device usage is by detecting the entire human pose from still images [60]. This takes an integrated approach where a database of poselets is trained instead of merely detecting the face and ears of the driver. It is unclear how well this approach will perform in a driving scenario where the visibility of the driver's posture is limited to the chest level upwards.

2.2.6 Region Proposal

Region proposal [61] is a technique of locating possible objects or object parts within an image. It generally offers a more optimal localisation [62] over the sliding window method. Selective Search [63], Multiscale Combinatorial Grouping (MCG)

[64] and objectness [65] are some of the region proposal techniques that have been formulated.

2.2.7 Artificial Neural Network

Artificial Neural Network (ANN) consists of individual artificial neurons acting as computational units. Multiple weighted inputs are computed by the individual unit's activating function [66] to map the output values [67]. The importance of an input is indicated by its weight, which may be positive or negative. ANN is a type of supervised learning [68] algorithm, where the algorithm learns from a labelled dataset to create a model that approximates the input-output association and predicts the output when given an unseen input.

Training or learning is the process of optimizing the weights to find a function which closely estimates the association between the input and output. One of the methods to identify the best weights [69] is to rely on a function which measures the error and then the network is trained towards minimizing the error. This function is known as cost function [70]. By framing the network learning in terms of an optimisation problem, the gradient descent algorithm [71] is used to find the function minima.

However, as computation of cost function may be expensive due to the amount of data and the size of the network. As a solution, a variant of gradient descent algorithm, Stochastic Gradient Descent (SGD) [72] could be used. SGD involves performing gradient descent on parts of the dataset at a time, thus reducing the computational cost.

2.2.8 Convolutional Layer

The convolutional layer is a set of artificial neurons which act as spatial filters, and are applied to input images to obtain reduced outputs known as feature maps. As object characteristics occur in a specific probability in particular locations within the image, the convolutional layer is able to 'learn' about the objects through its filters. Multiple feature maps [73] are generated by the layer to obtain multiple dimensional information about the input images. This process is also known as feature extraction. This is used in conjunction with an ANN to form a CNN.

2.2.9 Support Vector Machine

As a supervised learning algorithm, SVM classifies new data based on a learned hyperplane. Annotated examples are plotted on a higher dimensional area [74] and the higher dimensional plane which maximally separates two data classes is computed.

2.2.10 Region Convolution Neural Network

Region Convolution Neural Network (RCNN) [75] is a group of image processing algorithms, consisting of a region proposal algorithm, a CNN and a class based SVM. The region proposal algorithm first extracts likely object regions from an image, in which CNN extracts features from and the SVM classifies the object regions based on the CNN features.

2.2.11 Detection Pipeline

A recent work with similar research objectives functions in two stages [76]. Their detection pipeline first locates the vehicle's windshield's region of interest (ROI) and then performs classification of the driver's mobile device usage.

2.3 Survey Of Existing Software Toolsets

While the previous sections describe an academic understanding in the field, the actual resources available for other researchers to experiment and test may vary. Since a single software or algorithm is highly unlikely to fulfil all the objectives perfectly, a 'divide-and-conquer' approach is taken. The detection framework would be multi-staged, similar to the approach described in the previous section. The framework would consist of pre-existing LPR and FD algorithms as their domain area performance are very good and could largely eliminate errors in the detection process. In order to perform the required functions, programs that have such algorithms could be used.

In various image processing applications, such as CellProfiler [77], an image processing system for driver assistance [78], and Haar classifier based detection pipeline [79], image processing modules and libraries are used instead of requiring researchers to develop image processing functions from scratch, which requires additional expertise and time.

In order to identify the current resources available, a comprehensive analysis was conducted and reviewed in the following sections. The result of the survey will significantly aid efforts in achieving the research objectives. Due to the numerous possible variations in different application programming interfaces and software features, the suitable toolsets had to be committed to, after a few trials as continuous revision of the toolsets becomes unjustifiable.

The key criteria in identifying the suitable software for this research are capability, interoperability, cost and transparency.

Capability. The software toolset should be able to perform the necessary functions and processing that is required by the user. As an example, the software may allow for the implementation of user created functions and library toolsets.

Interoperability. There are a diverse number of softwares that may be used by various hardware users; the software toolset must be able to perform on most if not all of them, while complying with their standards and specifications. To be widely used, the software toolset must be capable of reading, writing and transferring data between different operating formats and machines. For example, two softwares created with the same programming language or using a common library toolset should be easier to analyse and to perform communication with each other.

Cost. The licensing and usage cost of a software toolset should be flexible and reasonable for long term user adoption. Generally, open source and free software toolsets are preferred to cater for future adaptation.

Transparency. The software toolset selected would have a proper documentation of the software libraries and functions. This enables users to identify and apply the specific parts of the software.

Generally, there are two types of software models: proprietary versus open source.

Open source software is defined as licenced software providing the right to be examinable, modifiable and sharable, for all intentions and parties. This open source characteristic has allowed some clients to be actively involved in open source development; popular open source platforms may have thousands of client-developers working on the project.

Due to its licensing nature, most open source developers are usually not paid to work on the project; they have no full-time commitment to carry out the project. Moreover, when this is coupled with factors like limited number of developers and clients, it could lead to slow or a complete halt to software updates, thus affecting future client interest and preferences. On the other hand, the free sharing of properly maintained open source softwares can encourage widespread adoption. This combination of attributes is good for further software development and use.

2.3.1 Licence Plate

There are a huge range of toolsets for LPD. In this section a number of those toolsets available were investigated.

One of the LPR system explored is GeoVision GV-LPR[1]. It is the LPR module for GeoVision industrial video management system, originally designed for vehicle parking security. It has an interface which accommodates up to 16 camera channels. This is practical for simultaneously monitoring multiple vehicles on a road. Unfortunately, the software only performs review from network-based surveillance cameras. GeoVision GV-LPR does not accept input from a digital video camera,

[1] http://www.geovision.com.tw/english/Prod_GVLPR.asp

which became one of the critical factors for the lack of adoption. There is no function to access sources of image frame or video input from a digital video camera, which proved vital due to the nature of the research setup. Due to the nature of ethical issues and financial limitations, it was considered impractical for research purposes to procure a network-based camera with live traffic feed.

Another critical feature is that it is closed source and cannot be retooled. For example, LP characters, time stamps and the vehicle image are most probably provided but other details such as the pixel size, LP location and source code may not be available, such as in GeoVision GV-LPR. This impedes the optimisation of LP detection accuracy via improvement of LPR algorithm.

iSpyLPR[2] is another surveillance framework which allows the third party LPR software integration, such as SimpleLPR from Warelogic, to run on Windows platform. Like GeoVision GV-LPR, the software consists of many features including remote monitoring over the internet enabled devices, video recording, multiple camera setup and usage, and motion tracking. Integration of applications is relatively straightforward as iSpyLPR generates details of LP detections such as plate area, plate number and plate location. A simple licence is to be acquired for running iSpyLPR with multiple cameras simultaneously.

As mentioned earlier, SimpleLPR[3] software is a LPR software that works under the iSpyLPR framework. It can deal with images from a database or directly from a memory buffer. It allows easy integration with C++, C# and scripting languages due to its nature as a .NET assembly. However, being only able to work within the Microsoft operating system is one of the limitations. In addition, there are no compilation instructions for modifying the source code that may be required to suit the needs of the research objectives. This is a major consideration as there are limited amount of research time and one would reduce the setup cost for learning and prototyping.

Another prominent open source LPR toolset that was investigated is OpenALPR[4]. The toolset has the basic LPD and recognition source code. Additonally, the OpenALPR library is capable of processing both pictorial and video input. OpenALPR provides step by step instructions on how it can be compiled, which allows for easy development of the research project. In addition, OpenALPR runs on major software platforms such as Windows, Ubuntu Linux, OS X, iOS and Android.

In addition, it also has a commercial version with additional features. The commercial version of OpenALPR has a monitoring interface, database search and

[2] http://www.ispyconnect.com/userguide-alpr.aspx

[3] http://www.warelogic.com

[4] http://www.openalpr.com

alert notification. The software constantly watches the video feeds and documents every licence plate that it sees. The licence plate data is configured on site and by camera. Monitoring can be performed simultaneously on all plates or selectively on specific sites. When a plate of significance is detected, the image of the plate can be reviewed, and the accuracy of the results can be manually determined.

In addition, alerts can be triggered for a specified set of licence plates. When a match is detected, an electronic notification can be submitted with the licence plate details for assessment.

All licence plate data is stored in OpenALPR’s online repository and are searchable via identifiers such as plate number, time, and location.

The software library can be linked and integrated to third party applications. Closed sourced applications require a commercial licence.

OpenALPR requires the operation of OpenCV, which is a widely used open source computer vision toolset. OpenCV has a range of image processing libraries, supports multiple operating systems, such as Windows and Linux, and has features to support real time video processing. This is particularly useful for overall integration of other modules in this research.

As the scalability and extensibility of the software toolset enables the research objectives to be accomplished, OpenALPR is selected as the main option for dealing with LPD. In addition, OpenALPR is well documented and relatively fit for this purpose.

2.3.2 Face Detection

There is a range of face detection software platforms available for one to select. This section explores the more prominent options in detail.

Due to its open source nature, the OpenCV [80] platform has amassed a wide range of image processing related algorithms. OpenCV has a tutorial which showcased its object detection module and coincidentally provides an example of face detection application[5]. OpenCV library enables classifier training in addition to having a pre-trained Haar classifier and LBP classifier,

5

https://docs.opencv.org/2.4.8/doc/tutorials/objdetect/cascade_classifier/cascade_classifier.html#cascade-classifier

Alternatively, the Matlab platform also comes with a set of image processing libraries which includes face detection and a cascade classifier trainer [6]. These features are similar to the OpenCV's version.

Another face detection algorithm [81] was also reviewed as it jointly performs face detection, pose estimation, and landmark estimation in live, unconstrained images. It also runs on Matlab platform, has some pre-trained models to demonstrate its primary features and allows for model training with a relatively small image dataset.

All three software platforms for face detection have similar capabilities and have clear documentation. However, OpenCV is selected as its existing interoperability with OpenALPR allows easier software integration for testing FD and LPD.

2.3.3 Mobile Device Usage Detection

Most of the softwares available for MDUD are in the research stage. There are two general strategies of identifying mobile device usage by the vehicle driver, namely object detection and action recognition.

Bourdev and Malik [82] have developed a 'poselet' model based software to detect human individuals from an input image and identifies different human actions. It is available in C++ and in Matlab. The poselets can be graphically represented and explored in Matlab. The software can only be used for research and educational purposes. An almost exact result can be obtained with the C++ version. Matlab's Parallel Computing Toolbox can also be used to improve processing capacity, available for Matlab users.

However, the poselet software was provided without the model training tools. The poselet detectors were already trained beforehand and stored as .mat files. In order to train a poselet classifier, Matlab's SVM training toolbox and extractHOGfeatures function could be used.

On the other hand, an object detection software that potentially aligns with the aims of this project is the RCNN. As discussed in Section 2.2.10, this method was originally designed to take a pretrained convolutional neural network [83] model developed on ImageNet [84] images and further fine-tunes the model on PASCAL [85] images. Both datasets identify object and background classes for every image. To identify potential objects within an image, RCNN runs Selective Search [63] function which identifies and boxes up about 2000 regions, and tests the network against each box, subsequently performing edge thinning within each class. Caffe, a C++ ConvNet library, is used by RCNN to train the models. Both RCNN and Caffe

[6] http://www.mathworks.com/help/vision/ug/train-a-cascade-object-detector.html?requestedDomain=www.mathworks.com

are available for research purposes under Berkeley Software Distribution licence on Github.

While both the poselets and RCNN software are relatively equal and relevant in achieving research objectives, RCNN was eventually chosen. RCNN is more equipped than poselets for training new detection models. RCNN has a clear, step by step instructions and a more active support forum available on their online repository. However, this selection does not necessarily translate as a recommendation for others. One should examine the different methods or algorithms based on their objectives, specific criteria and resources available.

2.4 Detection Construct

Based on the software toolsets discussed in Section 2.3, there is a need to develop a coherent multistage detection workflow. The sequence of detection based on individual software toolsets capabilities are determined and explained in the following paragraphs.

The LPR is used at the beginning phase of the detection pipeline because a traffic offence usually needs to be identified simultaneously with the vehicle's LP. LP is a standard identification tool for motor vehicle users. Without LP detected, the computer vision system would have to use other metrics to identify the potential traffic offender. In addition, the LP location is a relatively constant feature of the vehicle, which potentially cues to the driver position within the vehicle.

Secondly, face detection software is positioned as the middle part of the detection process. The detection of the driver face would further reduce the region of interest for subsequent analysis. The reduced search space may provide a suitable amount of memory and computation required for the follow up mobile usage detection.

Lastly, the mobile usage detection by RCNN is the final component of detection system. This part attempts to classify driver's mobile detection usage by machine learning, which is the main objective of this research.

3 Licence Plate Detection

While newer LPD algorithms seem to have better detection capabilities, no-free-lunch theorem [86] implies that no single LPD algorithm can perform perfectly for all traffic imaging conditions. Instead, an approach is to identify a LPD algorithm's operating range, through extensive testing and analysis.

The primary contribution of this chapter is the study of a specific LPD algorithm's detection capabilities for different imaging conditions. The hypothesis is that a LPD algorithm can detect LP sufficiently well for use by traffic law enforcement. It should be stated that this hypothesis can neither be formally proven nor disproven; this study aims to demonstrate the usability and limitations of the LPD algorithm under simulated traffic conditions.

In the following section, the requirements and the setup of the LP detection under the proposed system are elaborated. The method for the LPD algorithm to detect and the type of image dataset used is detailed. The algorithm's performance is tabulated and a follow up analysis is performed.

3.1 Requirements

Based on the aims and scope of this thesis, the general requirements for LP detection system to work is as follows:

A) The vision based detection system must be able to reliably detect and identify LP from an angle and distance, similar to that of a CCTV setup.

B) The system must be able the detect LP under various imaging conditions.

C) The system must be able to provide fast LP detection for subsequent processing.

3.2 System Setup

Based on the above requirements, the following setup is proposed. This setup consists of a computing device, an image based algorithm and image datasets. Although using live video feeds would closely simulate the ability of the traffic enforcement CCTV system, they are potentially costly to process due to the inherent sparsity and unpredictable occurrence of the offending behaviour. Hence, an image based system is developed instead. The computing device is a general-purpose workstation that is capable of displaying graphics, as well as performing image processing and analysis. The image processing software component applies LPD algorithm to the images. The algorithm used is detailed in the following section.

3.3 Method For Licence Plate Detection And Identification

After the initial image acquisition process, the first function of OpenALPR[7] is to use trained Local Binary Pattern (LBP) [87] features to detect potential LP areas (Figure 3.1) . These segments are then further processed in subsequent stages.

Binarisation is then applied to detect potential LP areas several times in order to be able to capture all possible licence plate characters that may be missing in a single binary image setting. Basically, this step generates two or more binary images under a range of settings with the Wolf-Jolion [88] and Sauovola [89] multimethod binarisation.

Character analysis is applied after binarisation. This step is for detecting character-sized areas within the plate area detected. It first locates all connected contours in the detected LP area. Then it filters the contours by the possible LP character sizes and the potential alignment of LP characters in a LP. This filtering process takes places numerous times, each time increasing the size of character to be processed. If there are no possible characters detected in a given area, the processing ends and the LP area is removed. If there are possible characters, the LP area is kept for additional computation.

The next step is to locate the licence plate edges, because the detection step taken previously merely indicates the possible LP area, which may vary from the actual LP area. This step attempts to accurately locate the LP borders. Firstly, all the Hough lines [90] within the LP area are located. A series of horizontal and vertical lines is generated as a result. This series and the character height determined by the character analysis are used to determine the edges of the LP.

After locating a possible LP location via plate edges, the candidate area may still be rotated and distorted. At this stage, the de-skew function is employed to reconfigure the candidate area to an expected orientation and dimension.

Next, the character segmentation function attempts to identify and differentiate each individual character. The spaces between characters are searched and defined by a vertical histogram. Only the characters which meet the height requirement are retained in this process.

The OCR function then computes the percentage of which the potential characters match a letter or a number.

Lastly, a post processing step identifies the most probable plate number arrangement. It first generates a list of all possible LP arrangements. This includes selectively adding blank spaces between certain LP characters and removing low quality characters.

[7] https://github.com/openalpr/openalpr/wiki/OpenALPR-Design

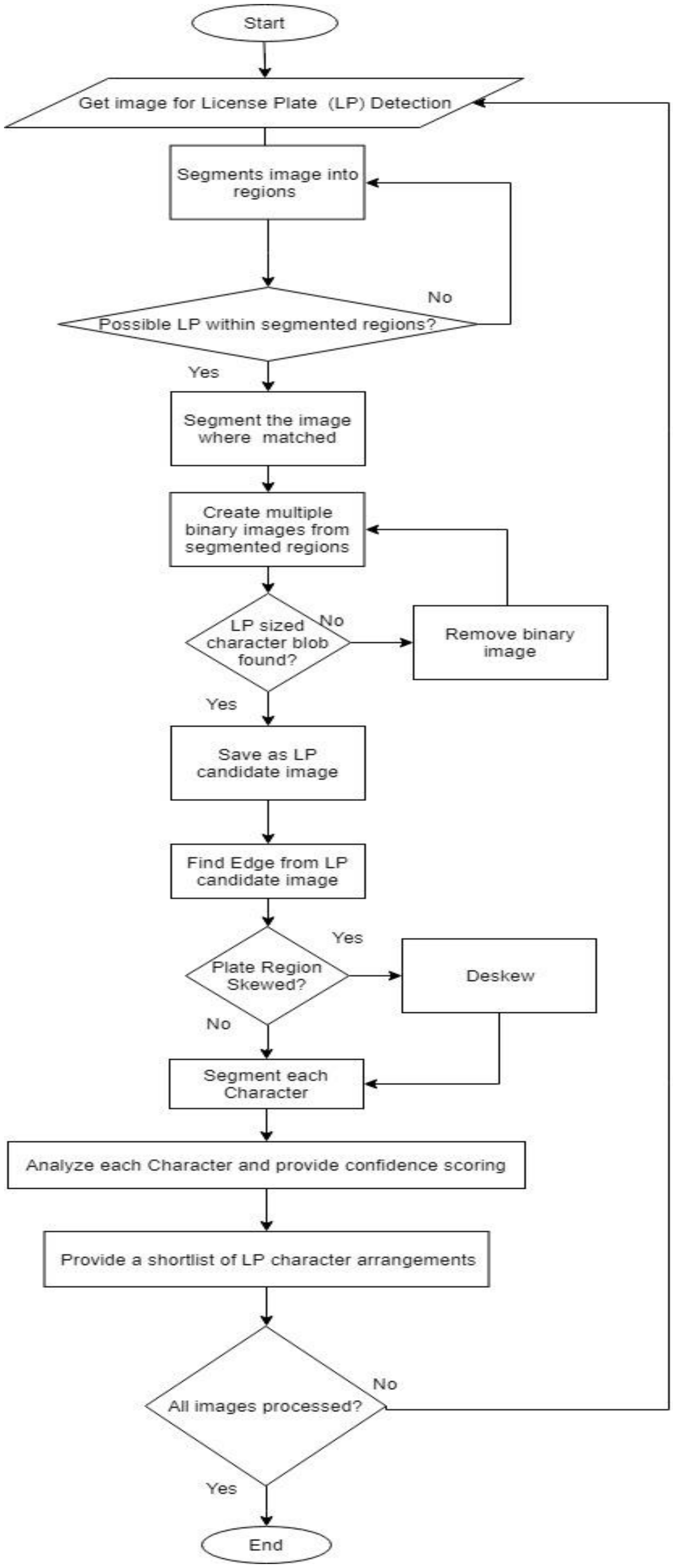

Figure 3.1 : Flowchart for LP detection process using the OpenALPR code

3.4 Dataset

The dataset is separated into positive and negative sets. The positive dataset compose of various LP related images (Figure 3.2), as per obtained from UCSD/Calit2 Car Licence Plate, Make and Model database [91]. The image resolutions of the positive image dataset range from 200 × 50 to 2304 × 1728. The LP sizes are about 3% of the image sizes. In approximately 82% of the images, the background is composed mostly of vehicular features. In the remaining 18%, the image background is largely dominated by road and building features. The minimum LP resolution is about 26 × 7. The camera is aligned in perpendicular to the LP for about 59% of images and at an angle for the other 41% of the images. 76% of the images is rear facing and 24 % are front facing. 15% of LPs within the images are unevenly illuminated. 2% of the LPs are motion blurred. 96 % of the LP images are outdoor scenes whereas 4% are indoor scenes. 95% of the LP images are day scenes whereas 5% are night scenes.

For the negative dataset, it is composed of road scenes (Figure 3.3) as per obtained from ImageNet database [84]. The image resolutions range from 134 × 69 to 2816 × 2112. While 54% of the road scene images consist of character features, only 5% of the road scene images consist of LP character features. The rest of the character features in the images come from road signage and image captions.

Figure 3.2 : Examples of images with LP, under different conditions, taken from UCSD/Calit2 Car License Plate, Make and Model database

Figure 3.3 : Example of various non-LP images, taken from ImageNet database

3.5 Detection Criteria

In order to consistently interpret the results of the detection module, a set of metrics are identified and applied. For positive image set, the result can be either a true positive (TP) or false negative (FN). The results that do not meet the TP criteria for positive images are false negatives (FN). On the other hand, for a negative set of images, the result can either be a true negative (TN) or false positive (FP).

For detections on the positive LP dataset, a result is considered a TP when the detection bounding box covers at least 80% of the LP area of a positive image. All other results are considered FN. For detections on the negative LP dataset, a result is considered a TN when there is no bounding box provided and is considered a FP when there is a bounding box provided.

3.6 Experimental Results

Table 3.1 shows a summary of the test results whereas Table 3.2 details the type of images within the positive image set. The LP is detected in about 68.2% of the positive LP images and no LP is detected in 99.3% of the negative LP images. The positive image sets which have less than 50% TP are grouped under the folder names: day_blurred, day_color (small sample), day_gray_scale, difficult_multi_color, and difficult_tracks_night.

Table 3.1 : Error matrix of LP detection on vehicles under various imaging conditions

		Predicted class	
		LP (%)	No LP (%)
Actual class	LP (n=1072)	68.2	31.8
	No LP (n=1239)	0.7	99.3

Table 3.2 : Positive image sets used for testing LP detection

Image set	Number of images	True Positive (%)	False Negative (%)
baza_slika	509	77	23
day_blurred	7	43	57
day_color (large sample)	138	80	20
day_color(small sample)	67	37	63
day_gray_scale	48	44	56
day_shadows_in_plate	52	69	31
day_very_close_view	122	80	21
difficult_multi_color	20	0	100
difficult_shadows	26	73	27
difficult_tracks	51	45	55
difficult_tracks_night	32	19	81
Total LP images	1072	68	32

3.7 Discussion

As the LP detection is the first component of the proposed detection module, the performance of the LP algorithm is critical to identify the personal vehicle used by potential traffic offenders. This work has demonstrated a range of imaging conditions that are expected for proper LP detection of traffic related images.

One of the factors contributing to the sub-optimal performance of the LP detection module on the dataset can mainly be attributed to difficult images within the datasets. Figure 3.4 shows a sample of challenging images from the different categories within the positive image set that were not detected by the LP module. These types of images contribute to about 21% of the entire positive image dataset and contribute to slightly more than 40% of total FN.

Figure 3.4 : Examples of undetected LP positive images from a variety of difficult positive image data subset

Another factor for the detection failure may be attributed to the imaging conditions such as contrast, tilt or unfamiliar LP area code. As depicted in Figure 3.5, this suggests that the LP algorithm operates within a much more finite range of imaging conditions than initially expected.

Figure 3.5 : Examples of LP positive images undetected by detection module due to subtle imaging conditions

To improve the LP detection capability of the system, one could apply multiple LPD algorithms in parallel, or a better LPD algorithm which work well for images taken under a variety of imaging conditions.

For the continuation of this research, the same LP algorithm will be used, under similar testing conditions. The focus of the next chapter would be on a similar set of images that allows for the existing LP algorithm to reliably detect LP while reasonably simulating the actual imaging factors affecting the subsequent detection stages, namely driver face and mobile device usage detection.

4 Driver Face Detection

The primary contribution of this chapter is the study of a specific LPD algorithm's capabilities for different imaging conditions. The hypothesis is that a LPD algorithm can detect LP sufficiently well for use by traffic law enforcement. It should be stated that this hypothesis can neither be formally proven nor disproven; this study aims to demonstrate the usability and limitations of the LPD algorithm under simulated traffic conditions.

In the following section, the requirements and the setup of the LP detection under the proposed system are elaborated. The method for the LPD algorithm to detect and the type of image dataset used is detailed. The algorithm's performance is tabulated and a follow up analysis is performed.

4.1 Requirements

Assuming that the LPD described in the previous chapter functions reliably, the requirements for driver face detection are as follows:

A) The system needs to be able to reasonably narrow the region of interest for detecting driver face within the motor vehicle.

B) The system needs to reliably detect the driver face, under realistic imaging conditions.

C) The system needs to perform fast driver face detection.

In order to meet the requirements mentioned, the system is setup as described in the following section.

4.2 System Setup

While the requirements for driver face detection is similar to LP detection requirement, the setup for driver face detection differs from the LP detection setup.

The setup consists of an image capturing unit, a computing device, an image processing software and an image analysis software.

The image capturing unit is introduced in this system as there are no suitable dataset that details driver faces within a vehicle. The image capturing unit is a digital video camera which intended to capture the driver's action within the vehicle from a minimum capturing distance outside of the vehicle. This is required to simulate the ability of a CCTV system to record passing vehicles on a road. Initial testing demonstrated the capability of the digital video camera to capture the view of both the vehicle and driver. However, to clearly view the driver within the vehicle, an additional image stream with an optimal setting is calibrated.

The additional image stream is intended to resolve potential image lighting issues. Firstly, the International Standards Organisation (ISO) imaging speed for vehicle interior stream is increased to allow the interior vehicle details to be clearly defined under poor lighting conditions. As seen in Figure 4.1, the increased ISO speed clarifies the interior content of the vehicle at the expense of exterior vehicle details, such as LP definition.

A polarizing lens is used to reduce glare directed by the vehicle windshield onto the video cameras by an external light source. In Figure 4.2, the midday sun introduces unwanted reflections on the vehicle and this is reduced with the application of a light filter.

Figure 4.1 : Comparison between low (left) and high (right) ISO speed and vehicle interior definition

Figure 4.2 : Glare on vehicle windshield (left) reduced by polarizing lens (right)

4.3 Method For Driver Face Detection

After the LP is detected, there are two subsequent components to detect the driver face:

A) Locating vehicle driver’s region of interest via knowledge-based estimation.

B) Performing face detection algorithm on vehicle driver’s region of interest generated in step A.

4.3.1 Method For Identifying Vehicle Driver's Region Of Interest

To further extrapolate driver's region of interest from the given image, a knowledge-based solution is developed. The knowledge model is used to determine the region of interest for detecting the driver's face, which is around the driver's seat. Hence the performance in distinguishing between driver and front seat passenger can also be interpreted as correctly detecting the appropriate region of interest for locating driver face. The knowledge model is based on the general representation of a typical passenger vehicle. For example, the licence plates of passenger vehicles are usually situated at the front and the centre of the vehicle. In addition, the driver seat is a fixed position in the vehicle, though the exact location varies between vehicles of different make and model.

The specification of camera angle, licence plate location, and vehicle dimensions could be used to pinpoint the area where the vehicle driver is most likely to be located.

In addition, the size and orientation of the licence plate detected provide further clarification of the expected vehicle driver's face size and orientation respectively.

The approximate distance from the LP to the driver's location, specifically the driver's face (LP-F), could be estimated by comparing with the dimensions of the vehicle types and model from an existing vehicle reference database.

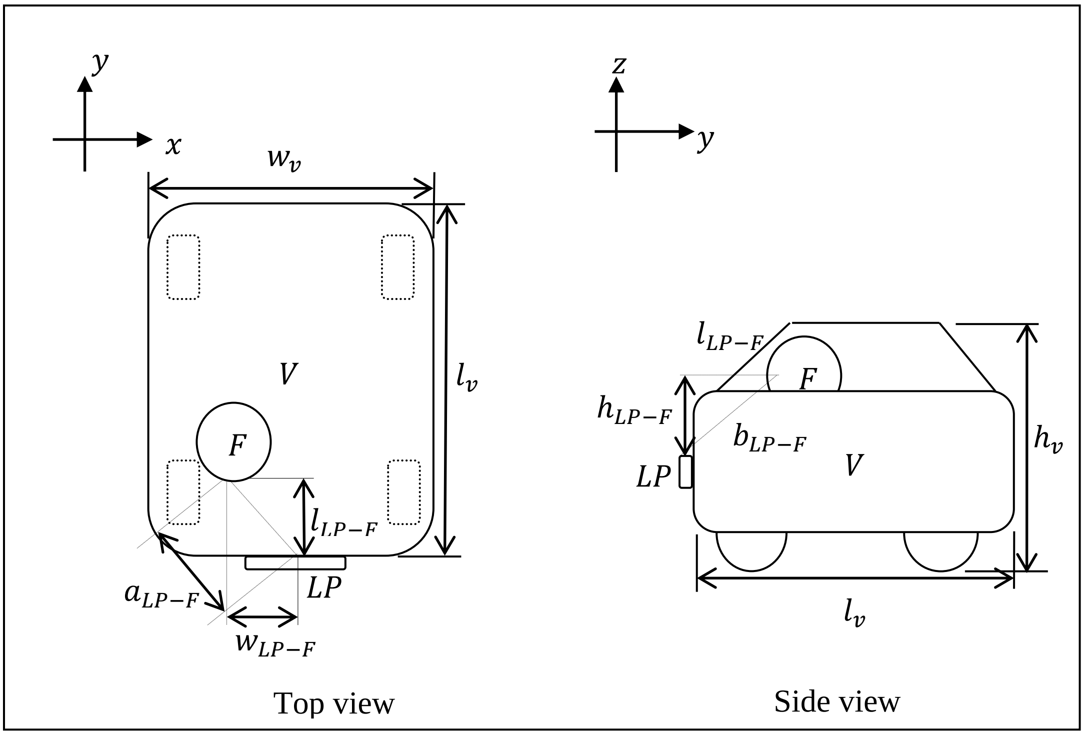

Figure 4.3 : A general view of the LP-F approximation model

A visual representation of the Licence Plate to Driver Face (LP-F) distance model is illustrated in Figure 4.3. In the top view, one can observe the horizontal plane of the vehicle. The lengthwise and widthwise LP-F distances are defined as l_{LP-F} and w_{LP-F} in parallel alignment to the vehicle's length, l_V and width, w_V. Additionally, a_{LP-F} is the shortest LP-F distance in the horizontal plane. Similarly, the LP-F distances visible from the side view is defined as h_{LP-F} , l_{LP-F} and b_{LP-F} respectively. Hence, two equations can be identified as follows:

$$a_{LP-F} = \sqrt{{w_{LP-F}}^2 + {l_{LP-F}}^2}$$

$$b_{LP-F} = \sqrt{{w_{LP-F}}^2 + {h_{LP-F}}^2}$$

Hence, a LP-F distance correlation within a two-dimensional (2D) image can be deduced based on the following assumptions:

A. The limits of the possible search space are based on the minimum and maximum possible dimensions for a variety of passenger vehicles.

B. There is a relatively narrow range for the driver to be situated within the vehicle.

C. The driver seat is located at the right side of the vehicle.

D. The licence plate is highly visible and usually located in the centre of the vehicle front.

E. The average h_{LP-F} is approximately 80% of h_V.

F. The average l_{LP-F} is approximately 40% of l_V.

G. The average w_{LP-F} is approximately 12.5% of w_V.

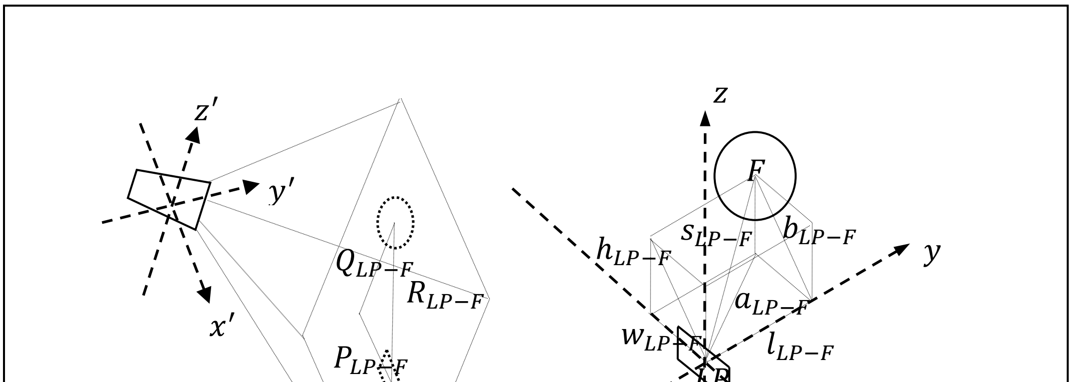

Figure 4.4 : LP-F knowledge model graph representation

Figure 4.4 shows an integrated LP-F representation with respect to the camera's angle. The camera angle is given by z', x' and y' respectively. In addition, a previously undiscovered component s_{LP-F} is revealed through this perspective representation. s_{LP-F} is the overall LP-F shortest distance. Two new equations that can be derived from this new perspective:

$$s_{LP-F} = \sqrt{{b_{LP-F}}^2 + {l_{LP-F}}^2}$$

$$R_{LP-F} = \sqrt{{P_{LP-F}}^2 + {Q_{LP-F}}^2}$$

where P_{LP-F} is the captured LP-F horizontal distance; Q_{LP-F} is the captured LP-F vertical distance and R_{LP-F} is the captured LP-F shortest distance.

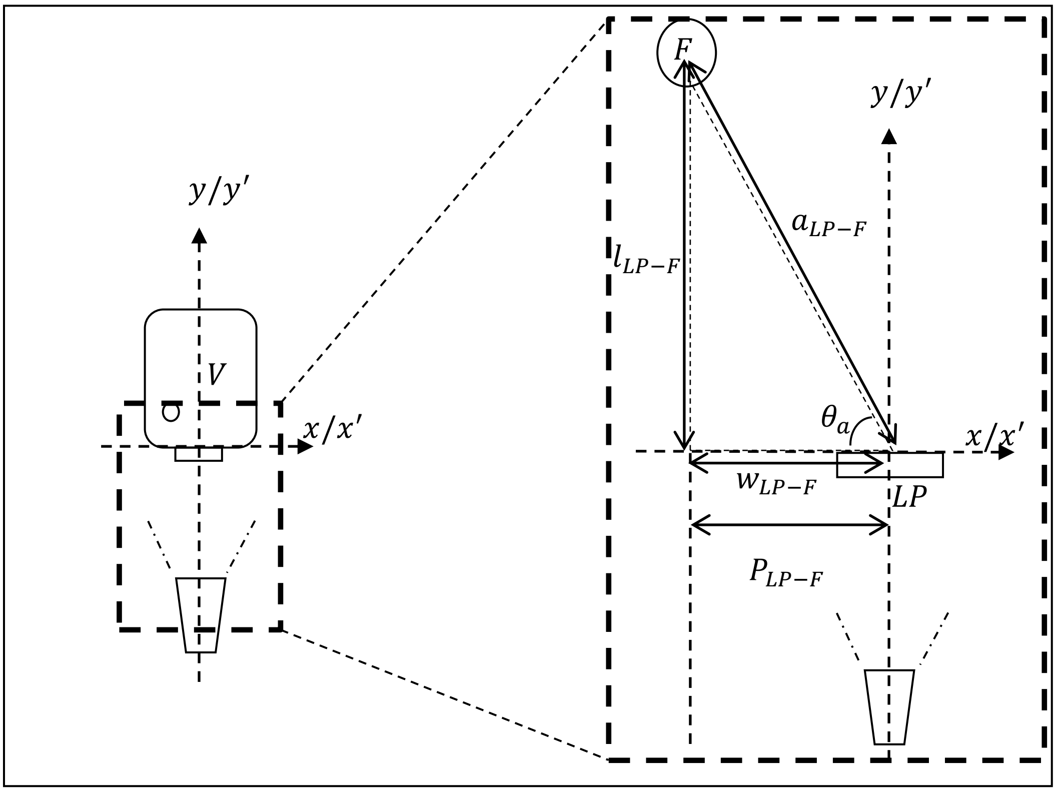

Figure 4.5 : The LP-F horizontal plane with respect to the camera position perpendicular to the vehicle front

To better understand the model, a simplified view of the model under a given scenario is presented. Figure 4.5 shows the top view of the model. w_{LP-F} and l_{LP-F} are measured from a reference image to obtain a_{LP-F}. The centre horizontal angle, θ_a is the angle between w_{LP-F} and a_{LP-F} in the horizontal plane, when the camera is perpendicular to the front and centre face of the LP. The equation is as follows:

$$\theta_a = cos^{-1}\left(\frac{w_{LP-F}}{a_{LP-F}}\right)$$

P_{LP-F} can then be estimated with the following equation

$$P_{LP-F} = a_{LP-F}\, cos(\theta_a - \theta_p)$$

With θ_p being the camera's yaw angle with respect to the centre line passing through the y-axis of the vehicle's LP.

Since the camera is perpendicular to the front and centre face of the LP,

$$\theta_p = 0$$

Once P_{LP-F} is obtained, it can be scaled by comparing the reference image dimensions to the vehicle dimensions within the image.

On the other hand, when the angle of the camera deviates from the perpendicular position, the perceived vehicle dimension would also change. In order to determine the new perceived distance between LP and driver's face, the camera angles relative to the centre and front of the vehicle must be specified.

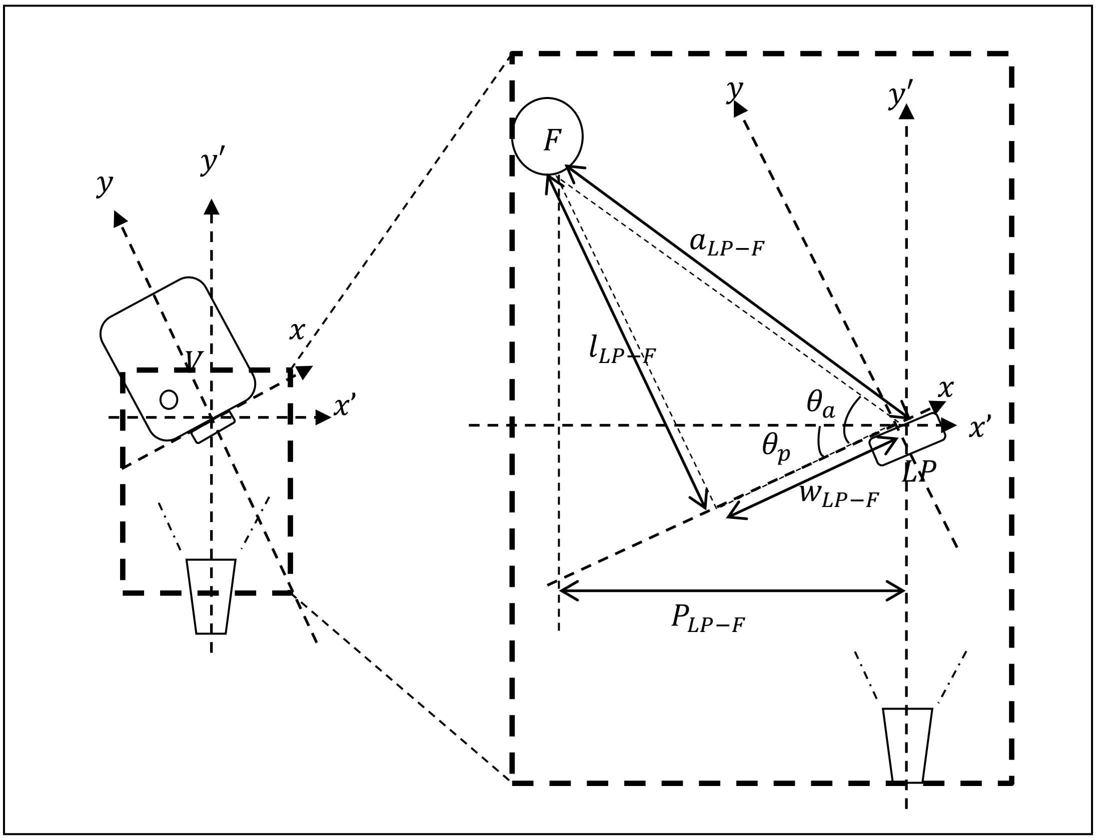

Figure 4.6 : The horizontal plane distance between LP and face when camera is positioned to the left of vehicle front

As shown in Figure 4.6, when the camera is position to the left of the vehicle, the P_{LP-F} can be calculated with the following equation:

$$P_{LP-F} = a_{LP-F}\; cos(\theta_a - \theta_p)$$

With θ_p being the camera's yaw angle with respect to the centre line passing through the y-axis of the vehicle.

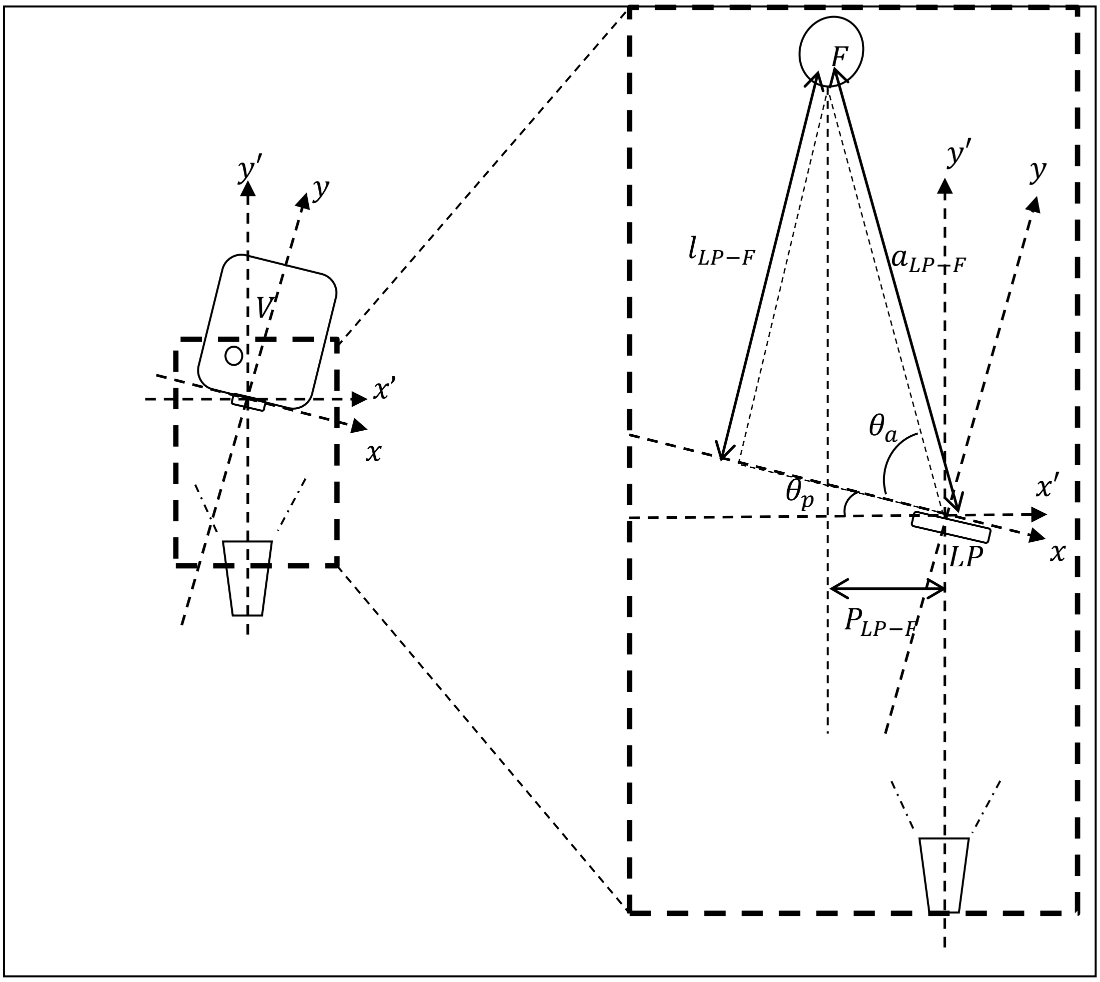

Figure 4.7 : The horizontal plane between LP and face when camera is positioned to the right of vehicle front

As depicted in Figure 4.7, when the camera is positioned on the right side of the vehicle and when θ_p is headed towards the opposite direction, a negative sign can be assigned to it. The perspective difference can be taken into account by the rewriting the equation as follows:

$$P_{LP-F} = a_{LP-F}\ cos(\theta_a + \theta_p)$$

When the sum of centre angle and specified is more than 90 degrees, the driver's face would be on the left side of LP.

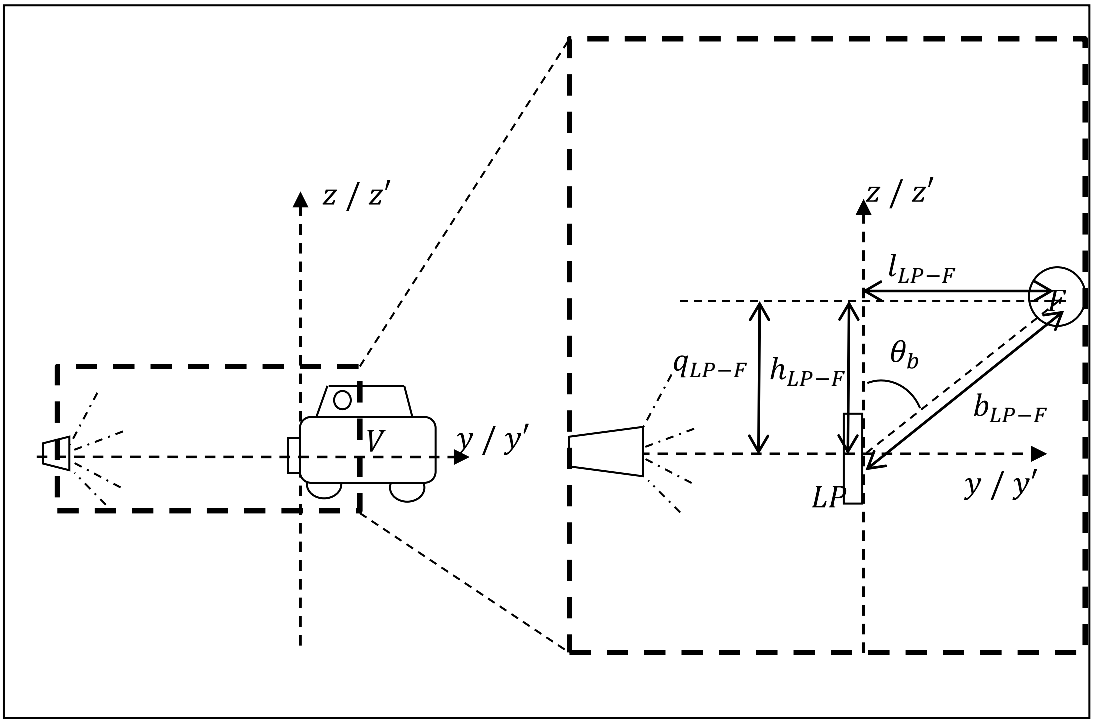

Figure 4.8 : Vertical plane distance between LP and face when camera is positioned perpendicularly to the vehicle front

Similarly, in the vertical plane as depicted by the side view in Figure 4.8, h_{LP-F} and l_{LP-F} are measured from a reference image to obtain b_{LP-F}. The centre vertical angle, θ_b is the angle between h_{LP-F} and b_{LP-F} in the vertical plane, when the camera is perpendicular to the front and centre face of the LP. The equation is as follows:

$$\theta_b = cos^{-1}\left(\frac{h_{LP-F}}{b_{LP-F}}\right)$$

Q_{LP-F} can then be estimated with the following equation

$$Q_{LP-F} = b_{LP-F}\, cos(\theta_b - \theta_q)$$

With θ_q being the camera's pitch angle with respect to the centre line passing through the y-axis of the vehicle's LP.

Since the camera is perpendicular to the front and centre face of the LP,

$$\theta_q = 0$$

Once Q_{LP-F} is obtained, it can be scaled by comparing the reference image dimensions to the vehicle dimensions within the image. The equation can also be applied to the situation shown in Figure 4.9 when the camera is positioned above the perpendicular line to the front and centre face of the LP, θ_q is non-zero.

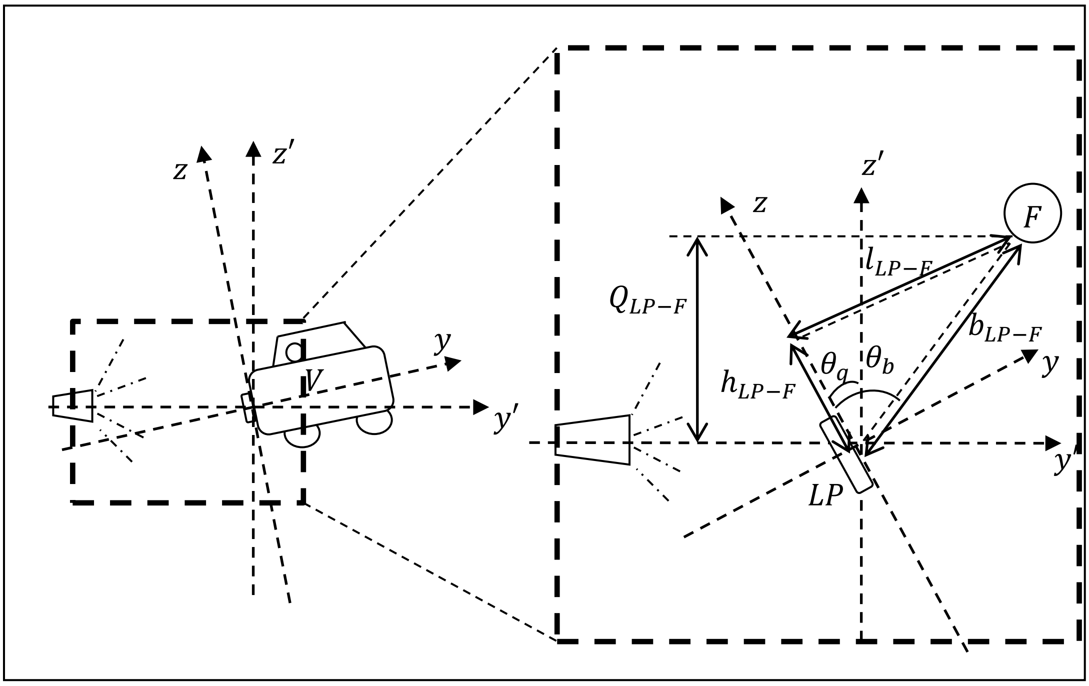

Figure 4.9 : Vertical plane distance between LP and face when camera is positioned higher than the vehicle front

With the above information, the LP-face distance in a vehicle can be estimated. However, there are many different vehicle dimensions. Due to the individual differences between the vehicle dimensions, the LP-face distance will also be different for each vehicle. Based on the LP-F knowledge model, the face detection window and offset distance from LP is calculated based on the maximum and minimum vehicle dimensions, as depicted in Figure 4.10.

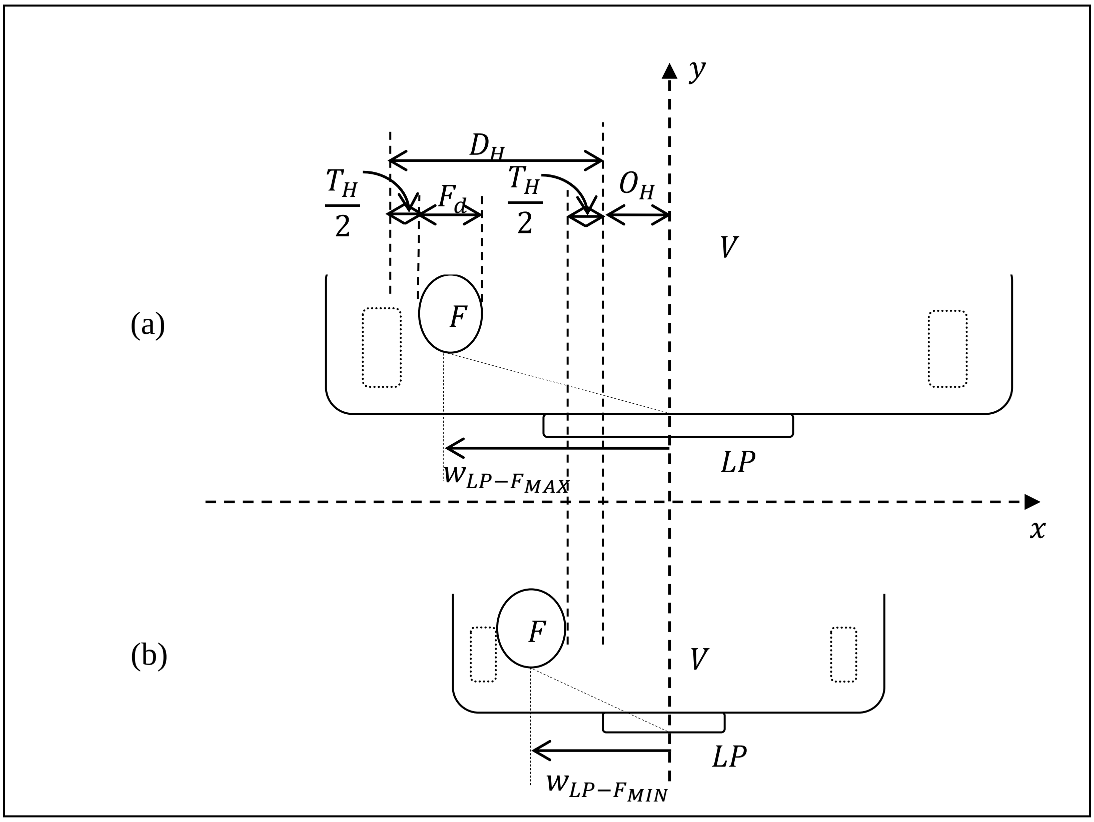

Figure 4.10 : Face detection window width and horizontal offset distance from LP based on (a) maximum and (b) minimum vehicle dimensions

$$O_H = w_{LP-F_{MIN}} - \frac{F_d}{2} - \frac{T_H}{2}$$

$$D_H = w_{LP-F_{MAX}} - w_{LP-F_{MIN}} + F_d + T_H$$

$$T_H = \frac{w_{LP}}{2} \cos\theta$$

where O_H is the horizontal offset distance from LP to detection window, $w_{LP-F_{MIN}}$ is the minimum LP-F width, $w_{LP-F_{MAX}}$ is the maximum LP-F width, F_d is the average face width, T_H is the horizontal allowable tolerance and D_H is the width of the detection window. The position of the vehicle driver's face within the vehicle may vary due to different posture or head angle. This is taken into consideration by nominating a value of T_H.

By using the equations deduced from the knowledge model, an approximate face detection window can be cropped as per Figure 4.11 for the use of the face detection algorithm described in the next section. Figure 4.12 shows a flowchart of LP-F approximation model.

Figure 4.11 : An example of an image (left) where LP-F approximation knowledge model can be applied to obtain a specific face detection region of interest (right)

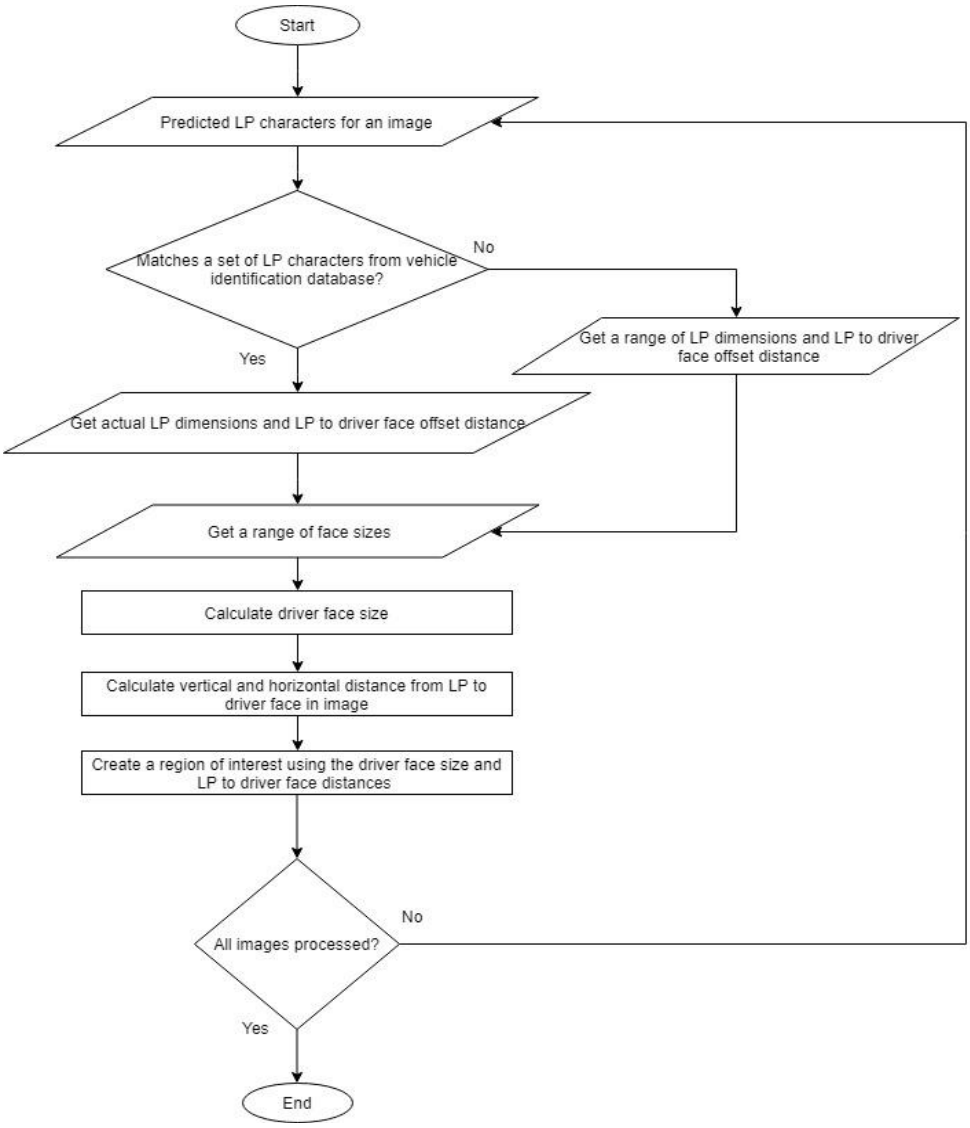

Figure 4.12 : Flowchart of approximating optimal driver face detection window using LP-F knowledge estimation model

4.3.2 Method For Face Detection

Detecting objects, such as faces, in realistic scenario is a complex task to automate and can be computationally expensive. The difficulty lies in its feature variations and distracting background similarities. A method to identify an object feature is by initially training a labelled object dataset against a sequence of basic filters to assemble a general bottom-up object feature, that can be used later to classify the object. Haar cascade classifier is such an example and is used in OpenCV for face detection.

Before the face detection, the image set requires pre-processing. For example in Figure 4.13, grey scale conversion is applied the colour images as the face detection algorithm specifically works with grey scale images. This pre-processing step is important to reduce the algorithmic computation required. The function has three input variables: source matrix, destination matrix and conversion format.

Figure 4.13 : Coloured (left) to grey scale (right) image conversion

In addition, the overall contrast of the converted image may be low for distinguishable features to be detected. As depicted in Figure 4.14, a histogram equalization function is applied to improve the image contrast by balancing out the histogram distribution. Similar to the colour conversion function, this function require two input parameters: source matrix and destination matrix.

Figure 4.14 : The distinctiveness of image features before (left) and after (right) applying histogram equalization

After the pre-processing steps, the OpenCV's object detection function, known as Cascade Classifier, is applied to detect the driver's face. The function loads a pre-trained face classifier and also accepts multiple input settings to determine an optimal trade-off between detection accuracy versus detection efficiency. For example, the scaleFactor parameter determines the magnitude of the input size reduction. Hence, the scale factor is inversely proportional to the number of measurements taken. minNeighbour (Figure 4.15) defines the minimum number of overlapping detections required to accept as a single detected object.

Figure 4.15 : Images show face detection as indicated by coloured circles when minNeighbour is set to 0 (left) and when minNeighbour is set to 2 (right)

Similarly, minSize and maxSize are parameters which specify the minimum and maximum allowable object dimensions respectively. To improve the performance of the detection, the knowledge of the approximate face size with respect to the LP can be used. The expected LP dimensions are approximated from the reference image. The maximum and minimum face dimensions are paired with the expected LP dimensions to generate a size ratio. As a result, the knowledge model estimates the face size based on the detected LP size and passes the face size value to the face detection algorithm to be processed.

By running the Cascade Classifier with a pre-trained face classifier, candidate faces are located at multiple scales within the image. Lastly, a post processing step takes the coordinates from Cascade Classifier and draws a rectangle on the image with a drawing function.

Figure 4.16 shows the process breakdown of OpenCV face detection when provided with a predicted region of interest.

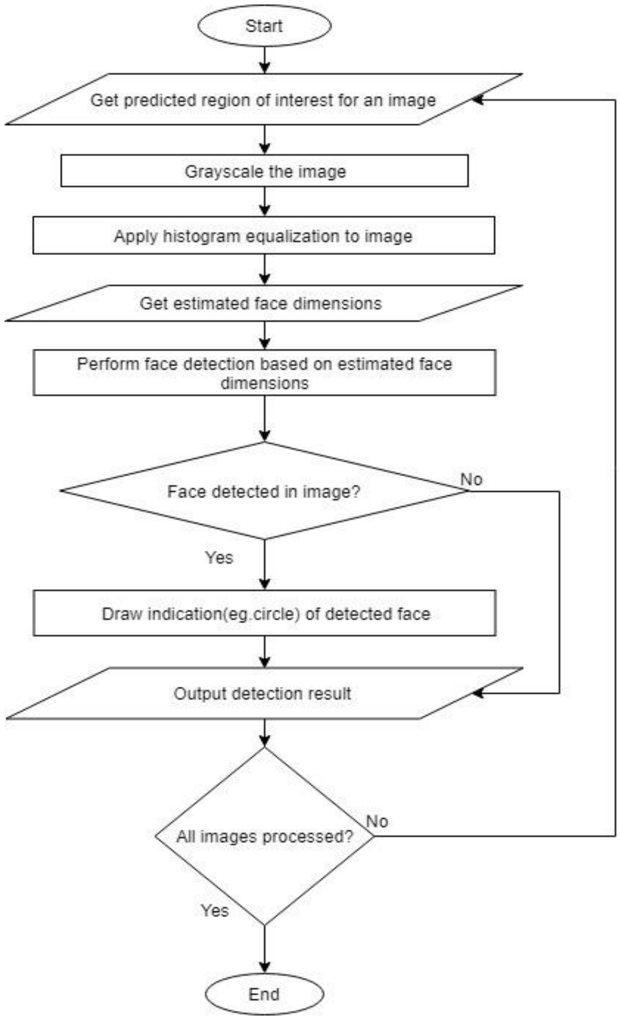

Figure 4.16 : Flowchart of face detection module using OpenCV's face detection algorithm

4.4 Dataset

The positive dataset for consist of 397 images, each containing a vehicle with a driver as shown in Figure 4.17, where 39% of the images have resolutions of 5184 × 3456, and 61% of the images have resolutions of 2848 x 4288. In terms of imaging direction, 47% of the images are captured from directly front and centre of the vehicle and 52% are from taken from the side of the vehicle. The LP sizes range from 0.1% to 1.9 % of the image size. The minimum LP size is 241 × 72. The driver face sizes range from 0.04% to 0.27% of the image sizes. The minimum driver face size is 97 × 77. For 77% of these images, the driver face is partially occluded. Figure 4.18 show that images are captured under different light intensities. About two-thirds of the driver faces within the image dataset are well lit. The remaining driver faces are poorly lit. As depicted in Figure 4.19, these occlusions include sunglasses, steering wheel, mobile devices and driver's hand.

The negative testing dataset consists of 1288 road images and 1307 motor vehicle images, all obtained from ImageNet database [84]. These images are chosen to test the detection module, where 99.5% of these images consist only of the background features similar to positive testing dataset. The image resolution for motor vehicle images ranges from 82 x 61 to 3000 x 1996 whereas the road images ranges from 134 x 69 to 2816 x 2112.

Figure 4.17 : Images showing various driver mobile usage and non-mobile usage pose

Figure 4.18 : Images captured under different lighting intensity at various times of the day

Figure 4.19 : Images showing occlusions that may interfere at various stages of the detection framework. The occlusions include driver's hand, face orientation and vehicle paraphernalia

4.5 Results And Analysis

4.5.1 Experimental Setup

To explore the effectiveness of the face detection module, a prototype is constructed. The prototype consists of LPD, LP-F knowledge model and the face detection algorithm, as depicted in Figure 4.20.

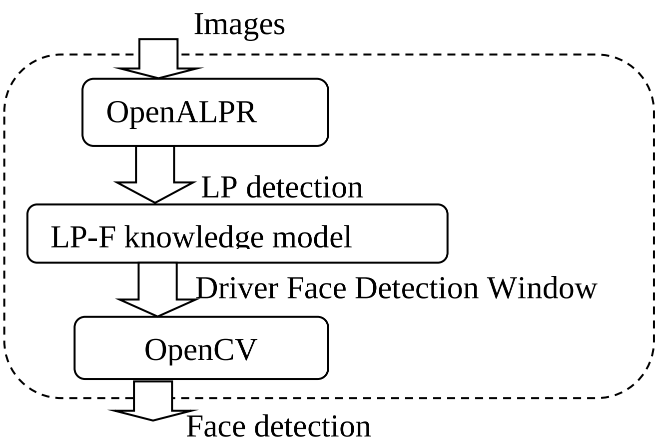

Figure 4.20 : Experimental setup of the driver face detection module

While the module is developed with the three functions working back to back to generate the final output, the results for LP detection, knowledge model and face detection are also presented separately. This is in order to determine the performance of the individual functions and their impact on the overall module performance.

4.5.2 Detection Criteria

The LP-F knowledge model is not a classifier since it passively receives input from LP detection stage. Thus, its evaluation criteria are different from that of LP and face detection. The knowledge model is evaluated based on the correctness of the model output. A correct detection is when the knowledge model accurately determines and displays the region of the driver's face in the detection window for the positive image set. The model output for the negative image set is ignored.

For the face detection, the result is TP when it meets the evaluation criteria as follows:

1) A circular bounding box located within vehicle region.

2) At least 1 circular bounding box which covers at least 50% of face.

3) The circular bounding box between 100 by 100 and 200 by 200, in pixels.

The evaluation for FP is similar to TP, with the exception that the circular bounding box does not cover an actual driver's face. The evaluation criteria for TN are the same as FN.

Lastly, the overall performance of the detection module is determined by both the amount of detected LPs and faces over the number of images in the image sets.

4.5.3 Results

The following results are manually verified and tabulated according to the criteria specified in the previous section, 4.5.2.

Table 4.1 : Error matrix of LP detection

		Predicted class	
		LP (%)	No LP (%)
Actual class	LP (n = 397)	96.5	3.5
	No LP (n = 2595)	0.2	99.8

Table 4.1 illustrates the LP detection by the LP detection sub module. For the positive image set of 397 images, OpenALPR has a 96.5% TP detection rate. Only 0.2% out of 2595 images from the negative image set was detected as LP.

Table 4.2 : Driver Face Detection Window performance

	Driver Face Detection Window	
	Correct (%)	Incorrect (%)
Correctly detected as LP (n = 383)	95.6	4.4

Table 4.2 show that of the 383 LP detected images from the positive image set, the driver face detection window is accurate for 95.6% of them.

Table 4.3: Face detection error matrix

		Predicted class	
		Face (%)	No Face (%)
Actual class	Face (n = 383)	71.0	29.0
	No Face (n = 6)	0.0	100.0

Table 4.3 display the face detection results of LP-detected images, for both positive and negative image sets. 71% of the faces were detected in the positive set whereas the detector achieved 100% TN for the negative set.

Table 4.4 : Overall Driver Face detection module error matrix

		Predicted class	
		Vehicle driver's LP and face (%)	Not vehicle driver's LP and face (%)
Actual class	LP and face (n = 397)	68.5	31.5
	No LP and face (n = 2595)	0.0	100.0

Table 4.4 provides an overview of the overall performance of the detection module. 31.5% of the positive image set is FN, where either the LP, face or both were not detected. Conversely, 100% TN was achieved for the negative image set.

4.6 Discussion

One of the issues faced is the sensitivity of the LP detection to the lighting situation around the LP area, as shown in Figure 4.21 and Figure 4.22. As a direct result of the variations in LP detection bounding box sizes, the LP-F detection window also varies as depicted in Figure 4.23. A possible reason for the variation is the contrast differences around the LP areas may have affected the detection and binarisation of the LP plates, part of LP detection process as described previously in the Section 3.3. There is no direct or easy workaround to this issue but the LP detection issue could be resolved by using a more robust LP algorithm. Another potential solution could be to re-arrange the order of detection sequence, for example, by detecting the LP characters first before the LP area detection.

Figure 4.21 : Differences in LP detection bounding boxes due to slight lighting variation

Figure 4.22 : A subtle lighting variation which resulted in dissimilarities in LP detection predictions

(a) (b) (c) (d)

Figure 4.23: Variation between LP detections (a, c) affecting the corresponding detection windows (b, d) predicted by the LP-F knowledge model

On the other hand, Figure 4.24 presents another major factor affecting the performance of LP detection; the LP angle. The LP detector is challenged when the LP is not horizontally level. However, this problem could be easily remedied by re-aligning the LP angle, either by using the knowledge of the camera-to-vehicle angle or by manual adjustment.

Figure 4.24 : The left image depicts an undetected LP in an image taken at an angle to the vehicle front whereas the right image shows LP detection in an image captured from a perpendicular angle to the vehicle front

Lastly, the capability of the detection module is affected by the face detection sub-module. Hence, issues such as driver face angle (Figure 4.25) and face occlusion remain critical factors in face detection. In order to overcome the angle and occlusion problem, the image capturing device can be positioned at various angles relative to the vehicle. The face location could then be deduced by combining the face detections from each image.

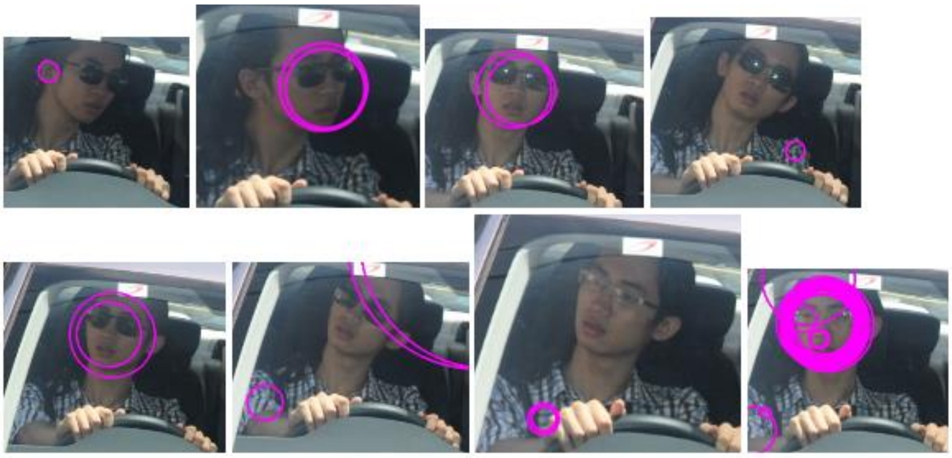

Figure 4.25 : Various levels of face angle affecting face detection performance

5 Mobile Device Detection

While the region of interest provided by the detection modules described in Chapter 4 and Chapter 5 attempts to narrow down the location of the mobile device, there are several challenges specific to detecting mobile device usage in images.

Firstly, the size of mobile devices vary and are usually smaller than the driver's face, hence an image dataset of sufficient quality is required. Secondly, the devices in use tend to be partially or heavily hidden, mainly by the driver's hands. The shadows from the driver's hands and face also contribute to the occlusion. Thirdly, the perceived shape of mobile devices can differ substantially for different image capturing angles and driver poses.

All these issues make the task of handcrafting visual features time consuming and potentially unfeasible. Hence, this module is dependent on a detection model which can learn the dynamic visual properties and classify mobile devices under realistic circumstances.

The following section details the key techniques for creating a mobile device detection model. Subsequently, the performance of the detection model is evaluated in the Results section. In the Discussion section, recommendations of possible improvements will be presented.

5.1 Requirements

The mobile device detection system needs to be able to fulfil the following requirements:

1) It needs to be able to reliably locate the mobile device, for various realistic, challenging driver poses and imaging settings.

2) The detection model needs to be able to reliably detect and identify a variety of mobile device angles and shapes.

5.2 System Setup

Similar to the setup for detecting driver faces, the MDUD setup consist of an image capturing device, a computing device and an image analysis software. The image capturing device is set to capture a variety of mobile device usage poses within the vehicle, which has not yet been captured and compiled into dataset. The computing device is a workstation capable of running image processing algorithms on large quantities of images. The techniques used in the image analysis software are addressed in the next section.

5.3 Method

The algorithm chosen for visually detecting mobile device usage is based on several basic image processing techniques that were addressed in Chapter 2, which culminated into a hybrid algorithm known as RCNN [75]. Hence, the aim of this module is to determine the practicality of the new object detector without focusing on optimizing its performance.

The custom dataset is to be readily trained and tested by the existing RCNN setup with no or minimal adjustments. For the purpose of applying RCNN to mobile usage detection, an image dataset of vehicle driver posing various mobile and non-mobile usage behaviours is used for training and testing. This will be described in the section 5.4 Dataset.

In order to identify potential objects within the image, a generic region proposal technique known as Selective Search [63] is applied.

For the purpose of mobile usage detection, Selective Search is applied to an image dataset of vehicle driver posing various mobile and non-mobile usage behaviours, which will also be described in Section 5.4 Dataset.

Since the proposed parts come in different shapes and sizes, an image warping function [75] converts each segment to a compatible format. Positive examples are defined when there is an intersection-over-union (IoU) of at least 0.5 between proposed segment and the object class bounding box in a given image. Conversely, the negative examples are defined when the IoU is less than 0.5. The sorted examples are then passed on to the classification function.

In the process of feature extraction, a specific implementation [83] of Caffe's [92] deep convolution network computes and outputs 4096 attributive vectors for each proposed segment. Similarly, these features are passed on to the classification function.

Then, a class specific SVM is employed to compute the best hyperplane from the proposed parts with its corresponding annotations and distilled features. From the identified hyperplane, possible mobile device usages can be classified and located within the proposed segments in new images.

Similarly, for the testing phase, region proposal followed by feature derivation are carried out on the test images. Using the trained classifier, each derived attribute is graded. Finally, greedy non-maximum suppression [75] is employed to remove proposed image parts which have an IoU overlap with a higher rated proposed part greater than the trained limit.

5.4 Dataset

Unlike the image datasets described in the previous chapters which include licence plates and driver faces, the dataset used in this chapter highlights specifically on mobile device usage. In order to simulate the real-world images of driver mobile usage, various orientations and poses have been re-captured with higher pixel definition. The changes of the image background are limited as more emphasis is given to the object variability with respect to the driver's hands.

Using LabelImg [93], the location and area of the mobile device within the images were annotated and saved in a .xml file. Figure 5.1 and Figure 5.2 show annotation examples of various possible mobile device usages.

To visually determine mobile device usage, the object definition is broadened to include the driver's body part in contact with the mobile device. In Figure 5.2, the object's region of interest in green bounding box which covers the driver's hand together with the mobile device. The aim is to increase the likelihood of detection as the mobile device characteristic may be like the background and occluded, making it a challenge to detect. This technique is similar to the preliminary study [75] done to add a context layer around the proposed object which improved the detection performance.

Due to limitations of image capturing device, this dataset for testing mobile usage detection focuses on capturing good quality images featuring mobile device usage and its proximal background. The positive dataset comprises of 1128 images. The image size ranges from 517 x 415 to 716 x 487. The sizes of mobile device itself range from 0.2% to 3.2% of the image sizes whereas the region of interest for mobile device usage behaviour is 0.7% to 7.4% of the image sizes. The driver sizes range from 26% to 63% of the image background and the vehicle paraphernalia form about 15% to 54% of the image background. Lastly, 16% to 31% of the image background is composed of the transparent vehicle window section, which has miscellaneous, irrelevant features. Some of the features overlap and are partially occluded. All mobile devices are partially occluded.

Figure 5.1 Example of an annotated positive mobile device usage image

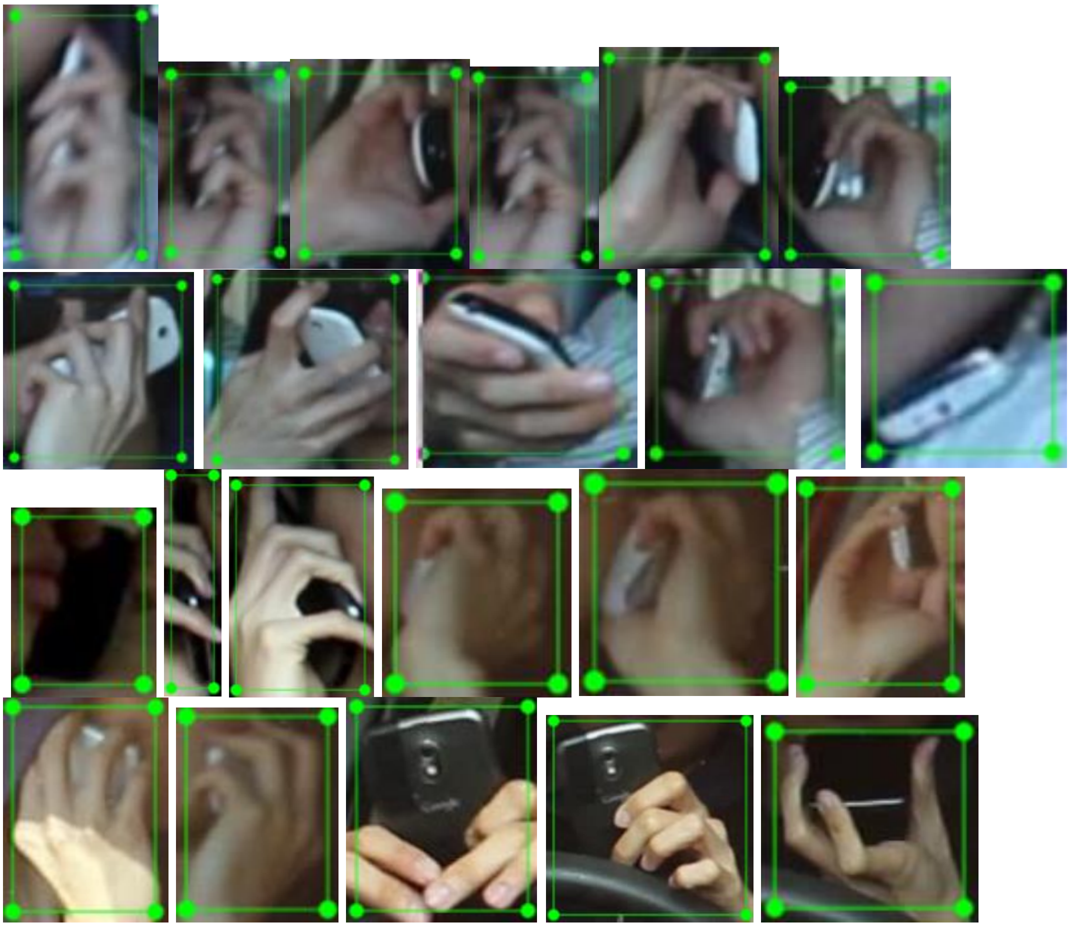
Figure 5.2 : Various examples of mobile device usages annotated with green bounding boxes

In addition, a set of negative images were captured from similar scenarios as the positive image. However, the driver/actor poses without the mobile device as seen in Figure 5.3. The negative testing dataset comprises of 1239 images. Image resolution ranges from 570 x 479 to 674 x 492. About 99.5% of these images have the size of 611 x 484. The driver sizes range from 39% to 49% of the image background and the vehicle paraphernalia form about 30% to 37% of the image background. Finally, 22% to 23% of the image backgrounds compose of the transparent vehicle window section, which has miscellaneous, irrelevant features.

Figure 5.3 : Examples from negative test set

5.5 Results And Analysis

5.5.1 Experimental Setup

To determine the viability of the MDUD module, a mobile usage detection image dataset is trained and tested with the RCNN algorithm. The training and testing stages are represented in Figure 5.4 and Figure 5.5.

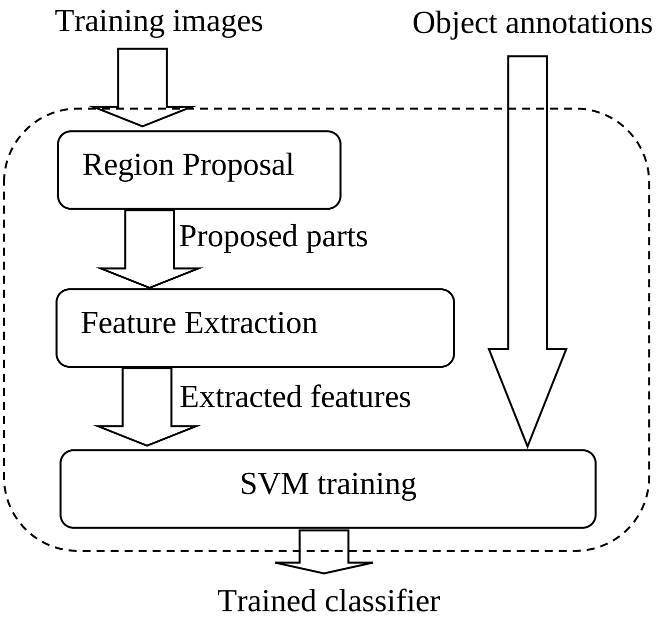

Figure 5.4: Training sub-module for mobile usage detection module, based on RCNN algorithm

The first stage to the mobile usage detection module is the training of a classifier. As seen in Figure 5.4, images from the training set and the object annotations are passed through the RCNN training algorithm. For each image, positive examples are taken from the proposed parts that has an equal or greater than 0.5 IoU with the annotated object bounding boxes and negative examples are from the remaining proposed parts [75]. Finally, the SVM training uses these examples and object annotations to train a classifier. The training specifications are detailed within the RCNN research paper.

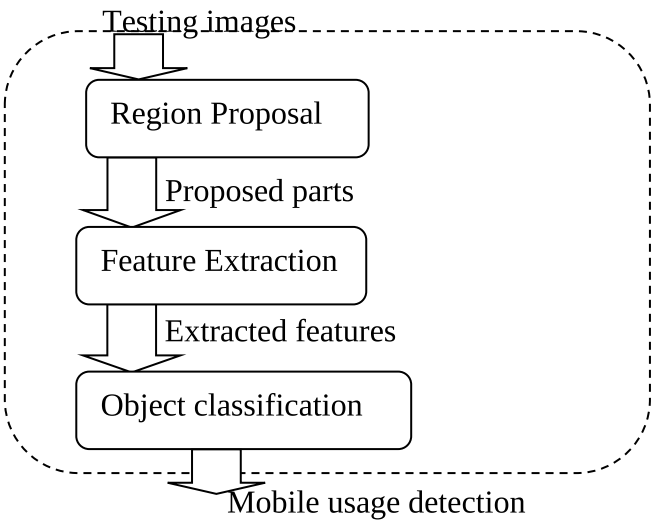

Figure 5.5 : The testing sub-module for mobile usage detection module, based on RCNN algorithm

Next, the testing images are run through the testing module. Similar to the training module depicted in Figure 5.4, the testing images have to go through the region proposal and feature extraction algorithms as shown in Figure 5.5. The trained classifier then computes the extracted features from a proposed part to evaluate the likelihood of mobile usage detection.

Preliminary tests initially showed that the trained classifier have low TP. By removing the greedy non-maximum suppression and score threshold filters, all possible mobile device usage detections were revealed. As shown in Figure 5.6, the classifier correctly provides a bounding box covering mobile device usages but was dismissed by the result filters due to their sub-zero score, resulting in sub optimal detection performance.

Figure 5.6 : An example of a detection output for an image generated by the RCNN testing algorithm without greedy non-maximum suppression and score thresholding

Hence, to capture the extent of the capabilities that this detection module can offer, the detection output has been modified to show all possible mobile device usage output and their ratings. Based on the preliminary analysis of the detection output, the evaluation criteria are reformulated and described in the next section.

5.5.2 Detection Criteria

The premise of this section is to describe an evaluation criterion for measuring the feasibility of the MDUD model. While the RCNN software does provide a similar set of performance metric, some modifications to the presentation of the metric are required. This is to be consistent with the detection criteria as described in the earlier chapters.

The detection is to be considered TP if the following requirements are met:

A) The detection output must show the target image, instead of a blank output.

B) The score for the first detection is more than -0.5.

C) The bounding box for the first detection output captures at least 50% of the mobile device usage in the image.

On the other hand, the detection is considered FN when there are signs of negative detection output on a positive image set. The signs of a negative detection output are:

A) A blank output is shown instead of the target image.

B) The score for the first detection is less than or equal to -0.5.

C) The bounding box in the first detection output captures less than 50% of the mobile device usage in the target image.

For a negative image set, a detection is counted as FP when the first detection output is a non-blank target image with a scoring above -0.5. Hence, the bounding box for the FP captures a false mobile device usage.

A TN detection is when there is a blank output shown or when the score for first detection is less than or equal to -0.5, on a negative image dataset.

A summary of criteria applicable to MDUD is provided in Table 5.1.

Table 5.1 : Summary of criteria for MDUD model trained by RCNN

		Detection	
		Positive	Negative
Image set	Positive	det1 score>-0.5 and correct bbox	blank result or det1 score<=-0.5 or (det1 score>-0.5 and wrong bbox)
	Negative	det1 score>-0.5 and wrong bbox	blank result or det1 score<=-0.5

5.5.3 Results

Based on the criteria specifications in the previous section, the detection outputs from the module are inspected manually and presented as shown in the following Table 5.2.

Table 5.2 : Error matrix of MDUD module

		Predicted class	
		Mobile (%)	No Mobile (%)
Actual class	mobile (n=1128)	76	24
	No mobile (n=1239)	43	57

Table 5.2 shows the MDUD by the trained MDUD module. The model achieved a 76% TP detection rate for a positive image set of 1128 images and a 57% TN detection rate for the negative set of 1239 images.

5.6 Discussion

Figure 5.7 : A variety of proposed parts identified as mobile device usage

As shown in Figure 5.7, the results demonstrate that the MDUD classifier can detect a wide range of features for the same object. The detection scores of the proposed part are based on the object characteristics detected by the trained classifier. It was observed that there is variability in the detection scores for different object poses.

However, the detection scoring of proposed parts can differ greatly even when the parts are visually similar. In the following Figure 5.8, there is a slight rotation of the proposed parts and this result in differences in scores. The classifier may have only been trained on proposed parts similar to (a) which is then able to generalise the detection in different proposed parts in (b) and (c).

(a) (b) (c)

Figure 5.8 : Similar looking proposed regions with different detection scores

Hence, a straightforward method to improve the performance of the modules would be to add more training images of variable tilts.

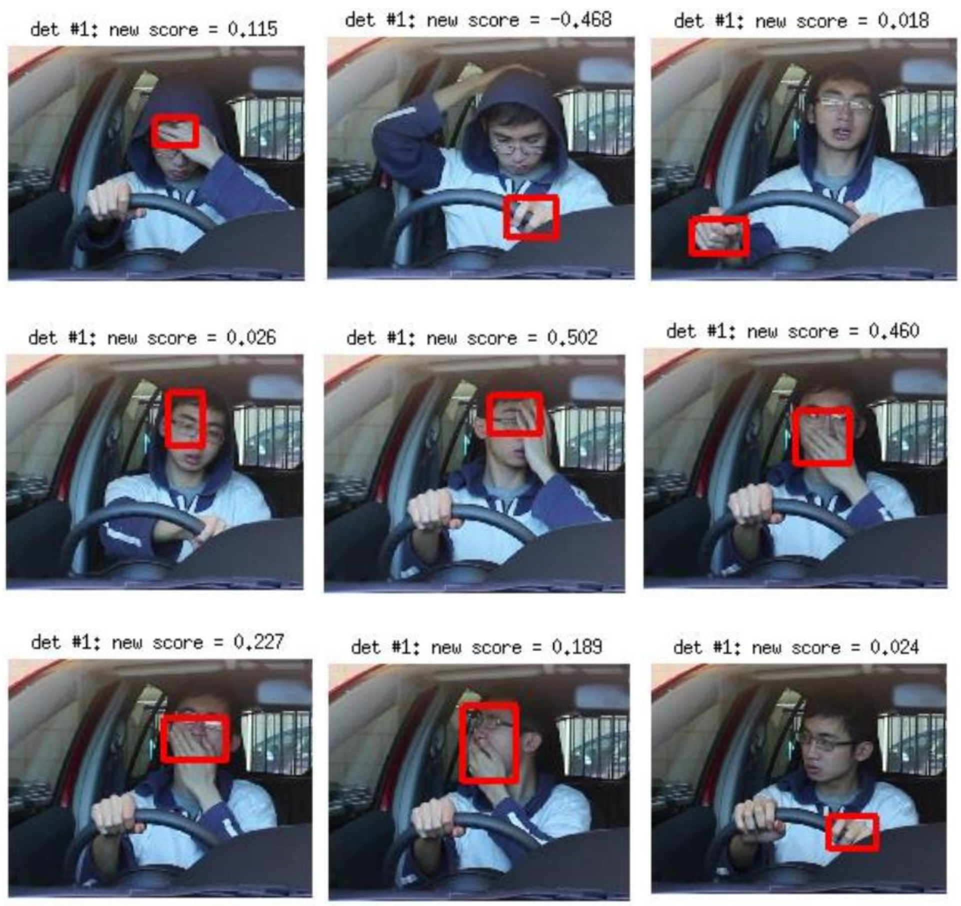

Figure 5.9 : Examples of positive detection of negative image set

The detection of driver's hand and face by the module as seen in Figure 5.9 indicates that the trained classifier is unable to properly discriminate mobile device usages in the absence of mobile devices. This is due to the object features including the driver's hands and face. As the classifier is only negatively trained on the proposed parts of the positive image set with less than IoU overlap of 0.5, this could have impacted the detection performance of the classifier. Based on this understanding, a possible solution would be to expand the negative training of the classifier to the images of drivers without mobile devices.

The computational time for feature extraction of mobile device usage based on the Matlab's Tic Toc function [94] is estimated to be about 450 seconds per image. This time consuming task can be speeded up by dividing the task into subtasks and running them in parallel. For example, an image can be segmented and a subtask is performed on individual segments by multiple individual processing units concurrently. By adding more individual processing units to simultaneously run tasks may proportionately increase the speed of the tasks, to a certain extent. Furthermore, creating and using customised processing units such as graphics processing unit (GPU) [80], which are tailored to specific use can further accelerate task completion.

6 Conclusion And Future Directions

The initial section compares the initial objectives to the actual outcome. Another section addresses the limitations of the detection framework. Next, a section is dedicated to provide an overview for the body of work done in the course of this thesis. Finally, directions for future research are presented.

6.1 Objectives And Outcomes

This chapter discusses the outcomes of the design with respect to the *initial objective statements*.

Exploration And Prototyping

Suitable hardware and software are first identified to produce the desired requirements of the system. Different state of the art methods will be investigated and incorporated where it is relevant and potentially useful towards the outcome of the proposed system.

An extensive research was carried out on the various developments of traffic law enforcement technologies and related alternatives as detailed in chapter 2. The review provided a sound understanding of the capabilities and limitations of each technique. In addition, a thorough knowledge of the problem scope was obtained, which allowed for a realistic expectation of a workable detection system. Hence, to develop a minimum viable prototype, a set of criteria was formed with a focus on the readily available options in line with the area of research.

Based on the analysis of the traffic enforcement systems, the manual review of live video feeds was singled out as a possible improvement area for implementing a system for traffic violation detection while requiring minimal adjustments to the existing infrastructure. Furthermore, the development and application of computer vision technology in multiple areas such as LP detection [14], face detection [52], pose detection [60] and convolution neural networks was helpful in the development of the detection framework. For prototyping purposes, image datasets were used in place of live video feeds.

From the review of the accessible detection algorithms, the availability of the source code was a prime factor for selection, as well as the interoperability among algorithms and product cost. Hence, the toolsets selected were largely open source, which supports interfacing and integration between detection algorithms. By having an overview of the current techniques and relevant methods, a working prototype was constructed based on the optimal selection of detection algorithms in the subsequent chapters. Thus, the objectives of acquiring relevant knowledge and resources to determine a suitable detection module have been accomplished.

Data Collection

Secondly, dataset of mobile usage by vehicle drivers that are appropriate for testing the system will be acquired. Different methods of generating, classifying and data testing will be used.

In Chapter 3, the types of data suitable for testing the system were determined. This was first done by evaluating the existing datasets created by similar research in the field. While there are some curated datasets for common objects such as vehicles, faces and mobile devices, a new dataset was created specifically for mobile device usage in vehicles. The dataset covers a comprehensive array of realistic poses, lighting and occlusions. Through numerous trials and errors, the datasets of an acceptable quality were developed.

As described in Chapter 4, 5 and Chapter 6, the datasets were chosen to exhaustively test the capabilities of the detection algorithms under realistic situations. In Chapter 4, to test the LP detection, the positive image set consists of a variety of vehicles with LPs under different imaging circumstances, taken from UCSD/Calit2 Car Licence Plate, Make and Model database [91] and the negative image set consists of the road images taken from ImageNet database [84], which is the background that the vehicle would typically be captured against. In Chapter 5, as the LP-F knowledge model and the face detection module were tested, the positive dataset was created specifically with drivers acting out different poses in a vehicle, captured from 3 different angles. In Chapter 6, as training and testing was required for detecting mobile device usage by image, a similar image dataset to that in Chapter 5 was recaptured. The main difference is that the latter images recaptured were of higher resolution and include more varied examples of mobile device usage.

Thus, the datasets used are regarded sufficient to rigorously test the prototype.

Development Of MDUD Framework

Thirdly, a system that makes use of data collected and compiled to provide MDUD among vehicle drivers will be investigated and developed. This can serve to demonstrate the viability of using image processing techniques to incorporate and automate a process that is usually handled by human experts, to the extent that it can be operated by first time users.

For prototyping purposes, the detection components are constructed and tested separately. In order to analyse the performance and limitations of the detection components, the performance of the components has been rigorously tested via various challenging datasets. The results reveal that the detection components work sufficiently well to deal with some of the scenarios it may face in the real world.

However, additional developments and assessments are necessary for the detection modules to become integrated, automated and user friendly. Despite these limitations, the work in this project should be of interest to traffic law enforcers.

Assess

The last objective will be to test the methods and examine the outcome of the system.

Investigations in Chapters 4, 5 and 6 highlights the potential synergy among OpenALPR's LP detection algorithm, LP-F knowledge model, OpenCV's face detection algorithm and Region Convolutional Neural Network. While the performance of the techniques used can be further improved, the results of this project suggest the practicality of partially automated detection to assist traffic law enforcers.

6.2 Limitations

In summary, the detection system has several working limitations.

The detection system was tested on still image datasets. Hence, it needs to be further tested and developed further before real time usage can be incorporated. In addition, unlike 3D image datasets, the detection system determines the relative object distance by applying perspective calibration algorithms on 2D image datasets. The perspective calibration can only provide an estimate and may lead to erroneous extrapolation.

The first part of the detection module identifies vehicle drivers via their vehicle LP. If the detection system is unable to correctly identify the LP, the detection framework fails immediately. This issue occurs in images taken under challenging imaging conditions, such as blurred or small images.

The second part of the detection module depends on a knowledge based driver window detection model, which has a constant specifying the maximum and minimum possible vehicle and face sizes. This constant will vary according to locality and is subject to change over time. This will require manual revision to the constant.

The RCNN model constructed in this thesis is only trained on lightly occluded vehicle driver actions; it may not work on situations where the face or mobile devices are heavily occluded.

6.3 Conclusion

This study is an early attempt at determining the feasibility of a vision-based solution for this challenging problem. It identified a systematic sequence of existing computer vision techniques. This work is significant as it adds to the body of knowledge for other researchers to build upon.

Multiple toolsets were chosen to form a detection ensemble, namely OpenALPR for LP detection, OpenCV for face detection and RCNN for training and testing MDUD. In addition, LP-F knowledge model was created to complement the limitations of the face detector.

To test the algorithms, suitable images were acquired. While some datasets are already available from other similar research, new image datasets are generated to fill the special needs of this research. Next, the detection algorithms are gauged using the realistic and challenging datasets. The datasets consist of a variety of imaging conditions and object features. Following that, their capabilities, limitations and possible improvements are discussed.

In hindsight, the objectives of the research are ambitious. Although it achieves the minimum objectives of exploring the possibility of vision based traffic violation detection, there is a substantial amount of work to be done before it can be serviceable to an end user. Nevertheless, it can act as a valuable reference for potential researchers and developers to aid them in their decision-making processes, thus saving time and resources.

6.4 Directions For Further Research

This section describes the possible areas where this research can be further explored. The recommendations vary in their benefits and costs to each researcher, depending on their area of interests and access to available resources.

6.4.1 Detection Algorithm Optimisation

To improve the quality of the detection system, the exploration of more robust algorithms is a good consideration. For example, a recent LP detector trains on a deep CNN variant [95] has a comparatively better performance on a LP dataset than OpenALPR. For face detection, an algorithm that is comparatively more robust to face orientations [81] could be used. Similarly, a RCNN variant with YOLO [96] showed an increased performance on a large generic object dataset and it could be re-purposed for MDUD.

Alternatively, the mobile usage detection framework can be entirely replaced by a series of deep convolutional neural network. Faster RCNN [97] has been trained on Face Detection [98] with greater accuracy and speed than the original RCNN. A similar training could be extended to detect LP, differentiate driver from passengers and detect mobile device usage. However, these CNN based detection modules would also need to overcome unexpected FPs and FNs [99] .

6.4.2 Optimizing Driver's Region Of Interest Localisation

Currently, the driver's region of interest within the image is located using the existing LP-F knowledge model. Hence, one can either seek to improve the existing model or find alternatives to achieve the same objective.

A new model can be developed by training a classifier on the common vehicle features such as the wind breaker, rear view mirror or steering wheel. From the detected vehicle feature, the driver's location could be further pinpointed.

6.4.3 Improvement To RCNN Trained Detector

As mentioned in the section 5.6, the RCNN framework could be expanded to include the negative training of the classifier to the images of drivers without mobile devices, thus potentially improving its performance.

6.4.4 Model Training And Testing

The purpose of using different datasets is to independently test the capabilities of each algorithm. The study is limited as there is no a unifying dataset to test all the algorithms. Hence, sourcing or creating a common testing dataset for testing this detection module is a possible direction for future extension of this work.

A logical step would be to consult and collaborate with traffic law enforcement agencies to obtain the testing datasets. This would greatly improve the understanding of the strengths and limitations of existing imaging hardware and allow the software to be tested. It would also help establish the image capturing quality baseline and standards.

Another way of acquiring data is through generative models [100]. An abundance of scenarios of variable illumination and shadows can be generated to train and test the detection algorithm.

6.4.5 User Centric Interface

Although the algorithms used are for research purposes, it can be more user friendly for the current research. The research setup requires the algorithm to be reassembled each time it is modified. In addition, some algorithm components have specific operating requirements which result in fragmented installations and troubleshooting issues. Since this is an investigation, the downtime cost is one-off. However, an effective deployable system would need the algorithms assembled and performing under a coherent interface.

6.4.6 Additional Research

An option would be to research new alternatives to the detection framework. For example, algorithms based on action or gesture recognition could potentially complement or substitute the object detection module of the detection framework. Action or gesture recognition implies a more contextual approach to image processing. In addition, action or gesture recognition algorithms may tackle multiple image frames at a time. Another example would be utilizing existing camera systems with higher resolution or increased processing capabilities to help detect the traffic violations. Another area of research would be to determine the framework's performance on live video feeds. The potential sparsity, data volume and scale differences are some of the potential challenges of real time video object detection. Further review is required and would be a good area for future research.

Bibliography

[1] D. A. Redelmeier and R. J. Tibshirani, "Association between Cellular-Telephone Calls and Motor Vehicle Collisions," *New England Journal of Medicine*, vol. 336, no. 7, pp. 453-458, 1997. Available: http://www.nejm.org/doi/full/10.1056/NEJM199702133360701 [Accessed September 12, 2014]

[2] S. P. McEvoy *et al.*, "Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study," *Bmj*, vol. 331, no. 7514, p. 428, 2005. Available: https://www.bmj.com/content/331/7514/428 [Accessed September 16, 2014]

[3] N. Cox. 2014. *WA Police go undercover in traffic sting* [Online]. Available: http://www.watoday.com.au/wa-news/wa-police-go-undercover-in-traffic-sting-20140611-zs41g.html [Accessed July 28, 2014]

[4] AAP. 2014. *WA police bust over 1000 texting drivers* [Online]. Available: http://www.sbs.com.au/news/article/2014/01/30/wa-police-bust-over-1000-texting-drivers [Accessed July 30, 2014]

[5] O. P. Popoola and K. Wang, "Video-based abnormal human behavior recognition—A review," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Online vol. 42, no. 6, pp. 865-878, 2012. Available: https://www.researchgate.net/profile/Oluwatoyin_Popoola/publication/270790763_Video-Based_Abnormal_Human_Behavior_Recognition-A_Review/links/5703bade08aeade57a25ad4e/Video-Based-Abnormal-Human-Behavior-Recognition-A-Review.pdf [Accessed July 18, 2014]

[6] X. Liming and Z. Yanchao, "Automated strawberry grading system based on image processing," *Computers and Electronics in Agriculture*, vol. 71, pp. S32-S39, 2010. Available: https://www.sciencedirect.com/science/article/pii/S016816990900204X [Accessed July 18, 2014]

[7] M. H. Karimi and D. Asemani, "Surface defect detection in tiling Industries using digital image processing methods: Analysis and evaluation," *ISA transactions*, vol. 53, no. 3, pp. 834-844, 2014. Available: https://www.sciencedirect.com/science/article/abs/pii/S001905781300205X [Accessed July 19, 2014]

[8] WHO. 2015. *Global status report on road safety 2015* [Online]. Available: https://www.who.int/violence_injury_prevention/road_safety_status/2015/Section_2_GSRRS2015.pdf?ua=1 [Accessed November 2, 2015]

[9] M. F. Oyaski, "Method of preventing cell phone use while vehicle is in motion," Patent US6771946B1, 2004. Available: https://encrypted.google.com/patents/US6771946.

[10] J. Yang *et al.*, "Detecting driver phone use leveraging car speakers," presented at the Proceedings of the 17th annual international conference on Mobile computing and networking, Las Vegas, Nevada, USA, 2011.

[11] X. Zhou and W. Curry, "System and Method for Preventing Cell Phone Use While Driving," Patent US 2011/0294520 A1, Dec. 1, 2011, 2011. Available: http://www.google.com/patents/US20110294520.

[12] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, "A real-time computer vision system for vehicle tracking and traffic surveillance," *Transportation Research Part C: Emerging Technologies*, vol. 6, no. 4, pp. 271-288, 1998. Available:

https://www.sciencedirect.com/science/article/pii/S0968090X98000199 [Accessed July 17, 2014]

[13] N. A. Mandellos, I. Keramitsoglou, and C. T. Kiranoudis, "A background subtraction algorithm for detecting and tracking vehicles," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1619-1631, 3// 2011. Available: http://www.sciencedirect.com/science/article/pii/S0957417410007116 [Accessed July 21, 2014]

[14] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic License Plate Recognition (ALPR): A State-of-the-Art Review," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 311-325, 2013. Available: https://ieeexplore.ieee.org/abstract/document/6213519 [Accessed July 21, 2014]

[15] L. Hsi-Jian, C. Si-Yuan, and S.-Z. Wang, "Extraction and recognition of license plates of motorcycles and vehicles on highways," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, vol. 4, pp. 356-359 Vol.4.

[16] W. Jia, H. Zhang, and X. He, "Region-based license plate detection," *Journal of Network and computer Applications*, vol. 30, no. 4, pp. 1324-1333, 2007. Available: https://www.sciencedirect.com/science/article/pii/S1084804506000762 [Accessed July 19, 2014]

[17] S. K. Kim, D. W. Kim, and H. J. Kim, "A recognition of vehicle license plate using a genetic algorithm based segmentation," in *Image Processing, 1996. Proceedings., International Conference on*, 1996, vol. 1, pp. 661-664: IEEE.

[18] H. Bai and C. Liu, "A hybrid license plate extraction method based on edge statistics and morphology," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, vol. 2, pp. 831-834 Vol.2.

[19] I. S. Ahmad, B. Boufama, P. Habashi, W. Anderson, and T. Elamsy, "Automatic license plate recognition: A comparative study," in *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2015, pp. 635-640.

[20] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," *Progress in brain research*, vol. 155, pp. 23-36, 2006. Available: https://www.sciencedirect.com/science/article/pii/S0079612306550022 [Accessed July 24, 2014]

[21] D. Llorens, A. Marzal, V. Palazón, and J. M. Vilar, "Car License Plates Extraction and Recognition Based on Connected Components Analysis and HMM Decoding," in *Pattern Recognition and Image Analysis: Second Iberian Conference, IbPRIA 2005, Estoril, Portugal, June 7-9, 2005, Proceedings, Part I*, J. S. Marques, N. Pérez de la Blanca, and P. Pina, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 571-578.

[22] Y. S. Soh, B. T. Chun, and H. S. Yoon, "Design of real time vehicle identification system," in *Systems, Man, and Cybernetics, 1994. Humans, Information and Technology., 1994 IEEE International Conference on*, 1994, vol. 3, pp. 2147-2152: IEEE.

[23] X. Shi, W. Zhao, and Y. Shen, "Automatic License Plate Recognition System Based on Color Image Processing," in *Computational Science and Its Applications – ICCSA 2005: International Conference, Singapore, May 9-12,*

2005, Proceedings, Part IV, O. Gervasi *et al.*, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1159-1168.

[24] F. Wang, L. Man, B. Wang, Y. Xiao, W. Pan, and X. Lu, "Fuzzy-based algorithm for color recognition of license plates," *Pattern Recognition Letters*, vol. 29, no. 7, pp. 1007-1020, 5/1/ 2008. Available: http://www.sciencedirect.com/science/article/pii/S0167865508000548 [Accessed July 23, 2014]

[25] C. Shyang-Lih, C. Li-Shien, C. Yun-Chung, and C. Sei-Wan, "Automatic license plate recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 1, pp. 42-53, 2004. Available: http://www.academia.edu/download/13159645/automatic_license_plate_recognition.pdf [Accessed July 20, 2014]

[26] J. Matas and K. Zimmermann, "Unconstrained licence plate and text localization and recognition," in *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, 2005, pp. 225-230.

[27] S. Draghici, "A Neural Network Based Artificial Vision System for Licence Plate Recognition," *International Journal of Neural Systems*, vol. 08, no. 01, pp. 113-126, 1997. Available: http://www.worldscientific.com/doi/abs/10.1142/S0129065797000148 [Accessed July 19, 2014]

[28] L. Wangchao and L. Shaofa, "A Hybrid License Plate Extraction Method for Complex Scenes," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, vol. 2, pp. 324-327.

[29] J. A. G. Nijhuis *et al.*, "Car license plate recognition with neural networks and fuzzy logic," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 1995, vol. 5, pp. 2232-2236 vol.5.

[30] M.-L. Wang, Y.-H. Liu, B.-Y. Liao, Y.-S. Lin, and M.-F. Horng, "A Vehicle License Plate Recognition System Based on Spatial/Frequency Domain Filtering and Neural Networks," in *Computational Collective Intelligence. Technologies and Applications: Second International Conference, ICCCI 2010, Kaohsiung, Taiwan, November 10-12, 2010. Proceedings, Part III*, J.-S. Pan, S.-M. Chen, and N. T. Nguyen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 63-70.

[31] F. Porikli and T. Kocak, "Robust License Plate Detection Using Covariance Descriptor in a Neural Network Framework," in *2006 IEEE International Conference on Video and Signal Based Surveillance*, 2006, pp. 107-107.

[32] K. Kanayama, Y. Fujikawa, K. Fujimoto, and M. Horino, "Development of vehicle-license number recognition system using real-time image processing and its application to travel-time measurement," in *Vehicular Technology Conference, 1991. Gateway to the Future Technology in Motion., 41st IEEE*, 1991, pp. 798-804: IEEE.

[33] C. A. Rahman, W. M. Badawy, and A. Radmanesh, "A Real Time Vehicle's License Plate Recognition System," in *AVSS*, 2003, p. 163.

[34] M.-S. Pan, J.-B. Yan, and Z.-H. Xiao, "Vehicle license plate character segmentation," *International Journal of Automation and Computing*, journal article vol. 5, no. 4, pp. 425-432, 2008. Available: http://dx.doi.org/10.1007/s11633-008-0425-0 [Accessed July 18, 2014]

[35] C. Busch, R. Domer, C. Freytag, and H. Ziegler, "Feature based recognition of traffic video streams for online route tracing," in *Vehicular Technology Conference, 1998. VTC 98. 48th IEEE*, 1998, vol. 3, pp. 1790-1794 vol.3.

[36] I. Paliy, V. Turchenko, V. Koval, A. Sachenko, and G. Markowsky, "Approach to recognition of license plate numbers using neural networks," in *Proc. IEEE Int. Joint Conf. Neur. Netw*, 2004, vol. 4, pp. 2965-2970.

[37] Q. Gao, X. Wang, and G. Xie, "License plate recognition based on prior knowledge," in *2007 IEEE International Conference on Automation and Logistics*, 2007, pp. 2964-2968: IEEE.

[38] K.-B. Kim, S.-W. Jang, and C.-K. Kim, "Recognition of Car License Plate by Using Dynamical Thresholding Method and Enhanced Neural Networks," in *Computer Analysis of Images and Patterns: 10th International Conference, CAIP 2003, Groningen, The Netherlands, August 25-27, 2003. Proceedings*, N. Petkov and M. A. Westenberg, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 309-319.

[39] J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Proceedings of the National Academy of Sciences*, vol. 93, no. 4, pp. 1591-1595, 1996. Available: https://www.pnas.org/content/93/4/1591.short [Accessed July 22, 2014]

[40] A. Capar and M. Gokmen, "Concurrent Segmentation and Recognition with Shape-Driven Fast Marching Methods," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, vol. 1, pp. 155-158.

[41] S. Nomura, K. Yamanaka, O. Katai, H. Kawakami, and T. Shiose, "A novel adaptive morphological approach for degraded character image segmentation," *Pattern Recognition*, vol. 38, no. 11, pp. 1961-1975, 2005. Available: https://www.sciencedirect.com/science/article/pii/S0031320305001251 [Accessed July 24, 2014]

[42] P. Comelli, P. Ferragina, M. N. Granieri, and F. Stabile, "Optical recognition of motor vehicle license plates," *IEEE transactions on Vehicular Technology*, vol. 44, no. 4, pp. 790-799, 1995. Available: https://ieeexplore.ieee.org/abstract/document/467963/ [Accessed July 20, 2014]

[43] K. Miyamoto, K. Nagano, M. Tamagawa, I. Fujita, and M. Yamamoto, "Vehicle license-plate recognition by image analysis," in *Industrial Electronics, Control and Instrumentation, 1991. Proceedings. IECON '91., 1991 International Conference on*, 1991, pp. 1734-1738 vol.3.

[44] H. Yo-Ping, L. Shi-Yong, and C. Wei-Po, "A template-based model for license plate recognition," in *Networking, Sensing and Control, 2004 IEEE International Conference on*, 2004, vol. 2, pp. 737-742 Vol.2.

[45] T. Naito, T. Tsukada, K. Yamada, K. Kozuka, and S. Yamamoto, "Robust license-plate recognition method for passing vehicles under outside environment," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 6, pp. 2309-2319, 2000. Available: https://ieeexplore.ieee.org/abstract/document/901900/ [Accessed July 22, 2014]

[46] Y. Dia, N. Zheng, X. Zhang, and G. Xuan, "Automatic recognition of province name on the license plate of moving vehicle," in *Pattern Recognition, 1988., 9th International Conference on*, 1988, pp. 927-929: IEEE.

[47] H. A. Hegt, R. J. d. l. Haye, and N. A. Khan, "A high performance license plate recognition system," in *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, 1998, vol. 5, pp. 4357-4362 vol.5.

[48] F. Aghdasi and H. Ndungo, "Automatic licence plate recognition system," in *AFRICON, 2004. 7th AFRICON Conference in Africa*, 2004, vol. 1, pp. 45-50: IEEE.
[49] T. Nukano, M. Fukumi, and M. Khalid, "Vehicle license plate character recognition by neural networks," in *Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004. Proceedings of 2004 International Symposium on*, 2004, pp. 771-775: IEEE.
[50] K. Mi-Ae and K. Young-Mo, "A simple OCR method from strong perspective view," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, 2004, pp. 235-240.
[51] K. Kluwak, J. Segen, M. Kulbacki, A. Drabik, and K. Wojciechowski, "ALPR - Extension to Traditional Plate Recognition Methods," in *Intelligent Information and Database Systems: 8th Asian Conference, ACIIDS 2016, Da Nang, Vietnam, March 14-16, 2016, Proceedings, Part II*, T. N. Nguyen, B. Trawiński, H. Fujita, and T.-P. Hong, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 755-764.
[52] P. A. Viola and M. J. Jones, "Robust Real-Time Face Detection," in *ICCV*, 2001, vol. 2, p. 747: Citeseer.
[53] L.-L. Huang and A. Shimizu, "A multi-expert approach for robust face detection," *Pattern Recognition*, vol. 39, no. 9, pp. 1695-1703, 2006. Available: https://www.sciencedirect.com/science/article/pii/S003132030500453X [Accessed July 23, 2014]
[54] J. Orozco, B. Martinez, and M. Pantic, "Empirical analysis of cascade deformable models for multi-view face detection," in *International Conference on Image Processing*, 2013, vol. 5.
[55] P. Smith, N. da Vitoria Lobo, and M. Shah, "Resolving hand over face occlusion," *Image and Vision Computing*, vol. 25, no. 9, pp. 1432-1448, 2007. Available: https://www.sciencedirect.com/science/article/pii/S0262885606003556 [Accessed July 25, 2014]
[56] D.-T. Lin and M.-J. Liu, "Face occlusion detection for automated teller machine surveillance," in *Advances in Image and Video Technology*: Springer, 2006, pp. 641-651.
[57] M. Yang and L. Zhang, "Gabor Feature Based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary," in *Computer Vision – ECCV 2010*, vol. 6316, K. Daniilidis, P. Maragos, and N. Paragios, Eds. (Lecture Notes in Computer Science: Springer Berlin Heidelberg, 2010, pp. 448-461.
[58] S. M. Islam, M. Bennamoun, and R. Davies, "Fast and fully automatic ear detection using cascaded adaboost," in *Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on*, 2008, pp. 1-6: IEEE.
[59] L. Nunes and M. A. Recarte, "Cognitive demands of hands-free-phone conversation while driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 5, no. 2, pp. 133-144, 2002. Available: https://www.sciencedirect.com/science/article/pii/S1369847802000128 [Accessed July 23, 2014]
[60] W. Yang, Y. Wang, and G. Mori, "Recognizing human actions from still images with latent poses," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 2030-2037: IEEE.

[61] I. Endres and D. Hoiem, "Category-independent object proposals with diverse ranking," *IEEE transactions on pattern analysis and machine intelligence,* vol. 36, no. 2, pp. 222-234, 2013. Available: https://ieeexplore.ieee.org/abstract/document/6544186/ [Accessed July 25, 2014]
[62] C. Gu, J. J. Lim, P. Arbeláez, and J. Malik, "Recognition using regions," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 1030-1037: IEEE.
[63] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International journal of computer vision,* vol. 104, no. 2, pp. 154-171, 2013. Available: https://link.springer.com/article/10.1007/s11263-013-0620-5 [Accessed July 18, 2014]
[64] P. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 328-335.
[65] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE transactions on pattern analysis and machine intelligence,* vol. 34, no. 11, pp. 2189-2202, 2012. Available: https://ieeexplore.ieee.org/abstract/document/6133291/ [Accessed November 2, 2014]
[66] F. Agostinelli, M. Hoffman, P. Sadowski, and P. Baldi, "Learning activation functions to improve deep neural networks," *arXiv preprint arXiv:1412.6830,* 2014. Available: https://arxiv.org/abs/1412.6830 [Accessed January 3, 2015]
[67] C. Gershenson, "Artificial neural networks for beginners," *arXiv preprint cs/0308031,* 2003. Available: https://arxiv.org/abs/cs/0308031 [Accessed February 2, 2014]
[68] C. Donalek, "Supervised and Unsupervised learning," in *Astronomy Colloquia. USA*, 2011.
[69] A. Ng, "CS229 Lecture notes - Supervised Learning," Available: http://cs229.stanford.edu/notes/cs229-notes1.pdf
[70] G. Montavon, *Neural Networks: Tricks of the Trade : Second Edition / edited by Grégoire Montavon, Geneviève B. Orr, Klaus-Robert Müller*. Berlin, Heidelberg: Berlin, Heidelberg : Springer Berlin Heidelberg : Imprint: Springer, 2012.
[71] A. Govan, "Introduction to optimization," in *North Carolina State University, SAMSI NDHS, Undergraduate workshop*, 2006.
[72] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*: Springer, 2010, pp. 177-186.
[73] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks,* vol. 8, no. 1, pp. 98-113, 1997. Available: http://www.cs.cmu.edu/afs/cs/user/bhiksha/WWW/courses/deeplearning/Fall.2016/pdfs/Lawrence_et_al.pdf [Accessed September 26, 2014]
[74] M. M. Adankon and M. Cheriet, "Support vector machine," in *Encyclopedia of biometrics*: Springer, 2009, pp. 1303-1308.
[75] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.

[76] B. Xu and R. P. Loce, "A machine learning approach for detecting cell phone usage," 2015, vol. 9407, pp. 94070A-94070A-8.

[77] A. E. Carpenter *et al.*, "CellProfiler: image analysis software for identifying and quantifying cell phenotypes," *Genome biology*, vol. 7, no. 10, p. R100, 2006. Available: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2006-7-10-r100 [Accessed September 23, 2018]

[78] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, and W. Seelen, "An image processing system for driver assistance," *Image and Vision Computing*, vol. 18, no. 5, pp. 367-376, 2000. Available: https://www.sciencedirect.com/science/article/pii/S0262885699000323 [Accessed September 3, 2014]

[79] J. Cho, S. Mirzaei, J. Oberg, and R. Kastner, "Fpga-based face detection system using haar classifiers," in *Proceedings of the ACM/SIGDA international symposium on Field programmable gate arrays*, 2009, pp. 103-112: ACM.

[80] K. Pulli, A. Baksheev, K. Kornyakov, and V. Eruhimov, "Real-time computer vision with OpenCV," *Communications of the ACM*, vol. 55, no. 6, pp. 61-69, 2012. Available: http://lvelho.impa.br/ip08/reading/rt-ocv.pdf [Accessed December 2, 2014]

[81] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2879-2886: IEEE.

[82] L. Bourdev and J. Malik, "Poselets: Body part detectors trained using 3d human pose annotations," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 1365-1372: IEEE.

[83] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[84] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248-255: IEEE.

[85] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303-338, 2010. Available: https://link.springer.com/article/10.1007/s11263-009-0275-4 [Accessed November 2, 2014]

[86] Y.-C. Ho and D. L. Pepyne, "Simple explanation of the no-free-lunch theorem and its implications," *Journal of optimization theory and applications*, vol. 115, no. 3, pp. 549-570, 2002. Available: https://link.springer.com/article/10.1023/A:1021251113462 [Accessed December 19, 2014]

[87] L. Liu, H. Zhang, A. Feng, X. Wan, and J. Guo, "Simplified Local Binary Pattern Descriptor for Character Recognition of Vehicle License Plate," in *Computer Graphics, Imaging and Visualization (CGIV), 2010 Seventh International Conference on*, 2010, pp. 157-161.

[88] C. Wolf, J. M. Jolion, and F. Chassaing, "Text localization, enhancement and binarization in multimedia documents," in *Pattern Recognition, 2002.*

Proceedings. 16th International Conference on, 2002, vol. 2, pp. 1037-1040 vol.2.
[89] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern recognition*, vol. 33, no. 2, pp. 225-236, 2000. Available: https://www.sciencedirect.com/science/article/pii/S0031320399000552 [Accessed December 19, 2014]
[90] V. Kamat and S. Ganesan, "An efficient implementation of the Hough transform for detecting vehicle license plates using DSP'S," in *Real-Time Technology and Applications Symposium, 1995. Proceedings*, 1995, pp. 58-59.
[91] L. Dlagnekov and S. Belongie, "UCSD/Calit2 Car License Plate, Make and Model Database," ed. Available: http://vision.ucsd.edu/car_data.html
[92] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675-678: ACM.
[93] Tzutalin. 2015. *LabelImg Github*. Available: https://github.com/tzutalin/labelImg
[94] MathWorks. 2014. *tic*. Available: https://www.mathworks.com/help/matlab/ref/tic.html [Accessed September 14, 2018]
[95] S. Z. Masood, G. Shu, A. Dehghan, and E. G. Ortiz, "License plate detection and recognition using deeply learned convolutional neural networks," *arXiv preprint arXiv:1703.07330,* 2017. Available: https://arxiv.org/abs/1703.07330 [Accessed September 10, 2018]
[96] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.
[97] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91-99.
[98] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 650-657: IEEE.
[99] J. Su, D. V. Vargas, and S. Kouichi, "One pixel attack for fooling deep neural networks," *arXiv preprint arXiv:1710.08864,* 2017. Available: https://ieeexplore.ieee.org/abstract/document/8601309/ [Accessed January 20, 2019]
[100] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672-2680.

Appendices

Appendix 1 Invitation to participate in data collection

Appendix 2 Participant Information Sheet

Appendix 3 Participant Consent Form

Appendix 4 OpenALPR, Knowledge Model & Face Detection algorithm

Appendix 1 - Invitation to participate in data collection

Email Subject: Participants needed for data collection

Are you tired of seeing vehicle drivers on their phones while driving? Do you agree that driving while using a mobile device is distracting and potentially dangerous?

I am a Masters by Research student and my research objective is to develop and evaluate a vision based system capable of visually detecting drivers using mobile devices while driving. I am looking for participants who can spare 45 minutes of your time to take part in a preliminary data collection exercise.

If you are interested and would like to be involved, please email Yew Meng at <yewmeng.woo@postgrad.curtin.edu.au> for more information.

Thank you.

This study has been approved by the Curtin University Human Research Ethics Committee (Approval Number RDSE-01-15). The Committee is comprised of members of the public, academics, lawyers, doctors and pastoral carers. If needed, verification of approval can be obtained either by writing to the Curtin University Human Research Ethics Committee, c/- Office of Research and Development, Curtin University, GPO Box U1987, Perth, 6845 or by telephoning 9266 2784 or by emailing hrec@curtin.edu.au

Appendix 2 - Participant Information Sheet

Participant Information Sheet

Project Title:

Study of Vision-based Detection of Mobile Device Use While Driving

Contents

Purpose of Research

This project aims to develop a framework that is able to visually identify specific instances of the driver using mobile devices while driving. The integration of template matching of licence plate, face detection algorithms and measure of symmetry in faces is used to detect specific instances of mobile device usage while driving. The results from this research are expected to show significant benefit in automating the visual based detection of driver mobile device usage.

Equipment

The equipment used is a digital image capturing device which captures high resolution images. The digital camera basically consists of a lens, image sensor and a memory card. The digital camera will be positioned in front of the stationary parked car. In addition, a mobile phone will be provided to the participant. The procedure is safe and non-invasive.

Your Role

At the beginning, we will give you an introduction to the research and the experiment you are about to go through. Before entering the driver's seat, you will be given a prompt card which informs you to carry out several postures as follows when prompted:

1) Using your *left* hand, position the mobile phone next to the *left ear*, with eyes looking *straight forward*.

2) Using your *left* hand, position the mobile phone next to the *left ear*, with eyes looking *towards the left*.

3) Using your *left* hand, position the mobile phone next to the *left ear*, with eyes looking *towards the right*.

4) Using your *left* hand, position the mobile phone in *front of the mouth*, with eyes looking *straight forward*.

5) Using your *right* hand, position the mobile phone next to the *right ear*, with eyes looking *straight forward*.

Next, you will be seated in the driver's seat of a stationary parked car. The digital camera will be setup in front of the car.

Then, the experimenter will prompt you to act out the postures.

You will be given time to read the prompted action in the cue card and then act out the postures. Each posture should last only 10 seconds.

We will be recording the data and progress at all times, and will be available for questions. The data collection will take approximately 30 minutes.

Consent to participate

Your involvement in the research is entirely voluntary. When you have signed the consent form we will assume that you have agreed to participate and allow the use your data in this research. You maintain the right to withdraw at any stage of the experiment without prejudice.

Confidentiality

The information you provide will be kept separate from your personal details which will be kept confidential. The data recordings will not have your name or any other identifying information on it. The data collected during the trial will be transferred via hard disk to our secure computer at the Dept of Mechanical Engineering, upon which all data relating to the trial recorded on the local computer disk will be deleted.

Further Information

This research has been reviewed and given approval by Curtin University Human Research Ethics Committee.

Approval number: **RDSE-01-15**

If you would like further information about the study at a later point, please feel free to contact Assoc Prof Tele Tan on t.tan@curtin.edu.au.

Thank you very much for your involvement in this research, your participation is greatly appreciated. Please keep this letter for your information.

This study has been approved by the Curtin University Human Research Ethics Committee (Approval Number **RDSE-01-15***). The Committee is comprised of members of the public, academics, lawyers, doctors and pastoral carers. If needed, verification of approval can be obtained either by writing to the Curtin University Human Research Ethics Committee, c/- Office of Research and Development, Curtin University, GPO Box U1987, Perth, 6845 or by telephoning 9266 2784 or by emailing hrec@curtin.edu.au.*

Appendix 3 - Participant Consent Form

Department of Mechanical Engineering

Curtin University

Study of Automatic Vision-based Detection of Mobile Device Usage in Drivers of Vehicles

I, ______________________________ (the participant) have read the information on the attached letter. I agree that the purpose and operation of the data capture process has been explained to me to my satisfaction and I have been given a written outline of the project.

I agree to participate in this activity, understanding that I may withdraw at any time, without prejudice to myself.

I understand that all information provided is treated as confidential.

I agree for data collection to be recorded with the following devices: Digital image capturing device.

I agree that the data and research gathered for this study may be published provided names or any other Information that may identify me/us is not used.

My participation in this project will not prejudice or impact on my studies or other roles at Curtin University.

Name ______________________________ Signature ______________________________

Date ______________________________

Investigator ______________________________ Signature ______________________________

Appendix 4 - OpenALPR, Knowledge Model & Face Detection algorithm

Publication of this appendix has been removed due to copyright restrictions.