<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Vision Algorithms and Systems for Vehicle and Pedestrian Detection in Night Time Conditions Using Vehicle Mounted Cameras</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>O’Malley, Ronan</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2011-08-24</td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/2390">http://hdl.handle.net/10379/2390</a></td>
</tr>
</tbody>
</table>

A dissertation presented by

Ronan James O’Malley

to

Electrical and Electronic Engineering
College of Engineering and Informatics
National University of Ireland, Galway

in fulfillment of the requirements for the degree of

Doctor of Philosophy

in the subject of

Electronic Engineering

Supervisors
Dr. Martin Glavin and Dr. Edward Jones

Research Director
Prof. W. G. Hurley

August 2011
# Table of Contents

Title Page ......................................................... i  
Table of Contents ............................................. iii  
Abstract ................................................................ vi  
Acknowledgements ................................................. x  
List of Figures ..................................................... xii  
List of Tables ....................................................... xv  
List of Abbreviations .............................................. xvi  

1 Introduction  
1.1 Motivation and Background ...................................... 1  
1.2 Objectives and Contributions .................................... 4  
1.3 Chapter by Chapter Overview .................................. 5  

2 Vehicle Lamp Identification  
2.1 Introduction ..................................................... 7  
2.2 State of the Art .................................................. 8  
   2.2.1 Advanced Driver Assistance System (ADAS) Landscape 8  
   2.2.2 Literature Review ......................................... 9  
2.3 Camera Configuration ............................................ 12  
   2.3.1 Exposure Control ........................................... 14  
   2.3.2 Colour Configuration ...................................... 19  
   2.3.3 Road Reflections ............................................ 19  
2.4 Red Rear-Lamp Identification .................................. 20  
   2.4.1 Regulation to Device Dependent Colour Space .......... 20  
   2.4.2 Adaptation to Real World Conditions .................. 23  
2.5 Headlamp Identification ....................................... 26  
2.6 Results ......................................................... 28  
2.7 Conclusions .................................................... 30  

3 Visual Vehicle Detection  
3.1 Introduction .................................................... 35  
3.2 State of the Art ................................................ 36
Table of Contents

3.3 Symmetry Analysis ................................................. 39
    3.3.1 Spatial Features ........................................... 39
    3.3.2 Cross-Correlation Symmetry Analysis ...................... 40
3.4 Perspective Distortion Correction ............................... 42
3.5 Kalman Filter Tracking ........................................... 44
3.6 Multiple Vehicle Distinction .................................... 47
3.7 Results ............................................................. 49
    3.7.1 Data Capture ............................................... 49
    3.7.2 Cross-Correlation Results ................................ 51
    3.7.3 Perspective Correction Results ............................ 52
    3.7.4 Video Processing Results ................................. 55
        3.7.4.1 Rear-Lamp Detection Results ....................... 55
        3.7.4.2 Headlamp Detection Results ....................... 58
    3.7.5 Tracking and Tracking Based Detection Results .......... 59
    3.7.6 False Positives and Detection Failures .................. 60
3.8 Conclusions ....................................................... 61

4 Infrared Pedestrian Detection .................................... 67
    4.1 Introduction .................................................... 67
    4.1.1 Far-Infrared Technology .................................. 68
    4.1.2 System Overview .......................................... 69
    4.2 State of the Art ............................................... 70
    4.3 Region Of Interest (ROI) Generation ......................... 75
        4.3.1 Morphology Based Clothing Compensation ............ 75
        4.3.2 Feature Based Region Growing ....................... 78
    4.4 Pedestrian Classification .................................... 83
        4.4.1 Histogram of Oriented Gradients (HOG) Features .... 83
        4.4.2 Support Vector Machine Classifier ................... 85
    4.5 Target Tracking ................................................ 86
    4.6 Results ........................................................ 87
        4.6.1 Data Capture ............................................. 87
        4.6.2 Performance Metrics .................................... 88
        4.6.3 Classifier Performance ................................. 90
        4.6.4 Video Processing Results ............................... 92
    4.7 Conclusions .................................................... 93

5 Fusion of IR and Visible Systems .................................. 96
    5.1 Introduction .................................................... 96
    5.2 Image Registration ............................................ 98
    5.3 Infrared Vehicle Masking ..................................... 98
    5.4 Visual Pedestrian Highlighting ................................ 100
    5.5 Conclusions .................................................... 102
## 6 Conclusions and Future Work

6.1 Project Summary and Conclusions ........................................ 104
6.2 Primary Contributions ...................................................... 106
6.3 Suggestions for Future Work .............................................. 107

References ................................................................. 109

A Projective Transformation .............................................. 120

B Distance Estimation ....................................................... 123

C Publications ............................................................. 125
   C.1 Journals ................................................................. 125
      C.1.1 Published ......................................................... 125
      C.1.2 In Preparation .................................................. 126
   C.2 Book Chapter ........................................................ 126
   C.3 Conferences .......................................................... 126
Abstract

Statistics show that a disproportionate amount of road fatalities occur at night-time despite the greatly reduced volume of traffic on roads. This thesis describes a system aimed at addressing this issue by automatically detecting the two main categories of road users: other road vehicles and pedestrians. A vehicle detection system using a visible spectrum camera and a pedestrian detection system using an infrared camera are presented, followed by a fusion of these systems.

An image processing system to detect and track vehicles using their lamps is presented. As the appearance of vehicle lamps in video can vary depending on camera hardware, a camera configuration process is implemented. To identify red tail- and brake-lamps, parameters for a red colour threshold are derived from automotive regulations. Higher intensity headlamps are identified by utilising a seeded region growing technique.

To obtain vehicle location data from identified lamps, they are paired using bilateral symmetry analysis. This is assessed by means of image cross-correlation, a shape and size independent approach to symmetry analysis. However, images of vehicle headlamps and tail-lamps suffer from perspective distortion during road manoeuvres, such as turning, engaging road bends and overtaking. A projective image transformation corrects for this perspective distortion, ensuring consistent detection performance through these road manoeuvres.

A night-time pedestrian detection system, based on the processing of Far-Infrared video, is presented. A pre-processing step is introduced, which compensates for distortion caused by clothing, using vertically-biased morphological closing. Regions Of Interest (ROIs) are identified using a region growing technique with high intensity seeds and feature-based stopping criteria. Histogram of Oriented Gradients (HOG) features are calculated from a database of images to train a Support Vector Machine (SVM) classifier.

Finally, advantages are attained by implementing a cooperative fusion of the visual
vehicle detection and IR pedestrian detection systems. Vehicle location extracted from the visual system is used to mask the IR frame, removing warm vehicle parts such as lamps, tyres and exhaust pipes. This reduces the probability of false positives in pedestrian detection. Pedestrians detected in IR video are highlighted in the visual image, making pedestrians much more visible for human video consumers.

This research aims to contribute to a safer road environment at night for both drivers and pedestrians.
I hereby declare that the work contained in this thesis has not been submitted by me in pursuance of any other degree

........................................................ Date.......................................... ..

Ronan O’Malley
This research was funded by the Irish Research Council for Science, Engineering and Technology (IRCSET) Embark Initiative.
Acknowledgments

Completing this doctorate has been a tremendous challenge. I have grown and developed many new skills which I have no doubt will change the course of my life. There are many people I need to thank.

Firstly, to my mother and father, Breeda and Padraic. They have always taken an interest in what I was undertaking and given me their unwavering support in every way possible. This would not have been possible without them and I will never be able to thank them enough. My thanks also to my sister Laura and my brothers Luke and Ross, for their support.

I wish to sincerely thank my supervisors Dr. Martin Glavin and Dr. Eddie Jones for their dedication, their patience, their help and their guidance. They motivated me when I was down, provided constructive feedback and always maintained a good sense of humour. They are shining examples of how to supervise and support a postgraduate student. My thanks to the other staff of Electrical & Electronic Engineering in NUI Galway in particular, Mary Costello, Martin Burke, Myles Meehan, Paddy Melia, Pat McGrath, Dr. Fearghal Morgan, Prof. Gearóid ÓLaighin and Prof. Ger Hurley.

I would like to acknowledge the Irish Research Council for Science Engineering and Technology (IRCSET) for their financial support. I would also like to express my appreciation to those who provided the use of cameras and equipment, Valeo Vision Systems / Connaught Electronics, Reg Farrell Engineering, Tyco Healthcare and in particular the Community Knowledge Initiative (CKI), NUI Galway.

Great thanks to my friend Gerard O’Malley for always being willing to help me capture video data.

A special thanks to the past and present members of the Connaught Automotive Research Group, Diarmaid Ó Cualain, Dr. Ciarán Hughes, Dr. Rob McFeely, Pat Hurney, Lorcan Browne, Myles Friel and Dr. Dermot Campbell. Their advice and support has been invaluable. I would also like to thank all my other friends in EEE who made my time in Nuns Island a pleasure, in particular Dr. Sean Finn, Dallan Byrne, Dr. Martin O’Halloran, Dr. Raquel Conceição, Dr. Brian McGinley, Shane...
Lynch, Migle Makelyte, John Maher, Seamus Cawley and all the other EEE postgrads. I know I wouldn’t have made it through without the time we spent in the canteen!

Finally, to my girlfriend and partner Lorraine. For her love and kindness I will be grateful forever. She has always been patient and her enthusiasm keeps me going every day. I could not have achieved this without her.
List of Figures

2.1 Typical Bayer filter GRGB pattern. ................................. 13
2.2 Video frames captured with automatic exposure and a lower static exposure. ......................................................... 15
2.3 Extreme test conditions for static exposure setting. ............. 18
2.4 A circular polarising filter placed in front of the camera lens is used to reduce the intensity of reflections. ......................... 20
2.5 CIE 1931 xy chromaticity diagram showing the regulation limits of light for red rear automotive lamps and the PAL colour gamut. .... 21
2.6 Rear-lamp regulation red colour region transformed into RGB space and RGB scatter plot of pixels from database of tail- and brake-lamp images. ................................................................. 23
2.7 Rear-lamp regulation red colour region mapped to conical HSV colour space and conical HSV scatter plot of pixels from database of tail- and brake-lamp images. ................................................................. 24
2.8 Histogram illustrating the distribution of saturation ($S$) values from database of captured tail- and brake-lamp images and a Gaussian curve fit to the data................................................................. 25
2.9 The stages of region growing thresholding of a vehicle headlamp image. ................................................................. 28
2.10 Low exposure night video frames, showing the rear tail- and brake-lamps of vehicles at various distances and in various lighting conditions and resulting binary images from the corresponding HSV colour threshold and morphological closing................................................................. 32
2.11 Low exposure night video frames, showing headlamps of vehicles at various distances and in various lighting conditions and resulting binary images from the corresponding region growing threshold and morphological closing................................................................. 33
2.12 False positive tail-lamp and headlamp identification, caused by background and infrastructure light sources. .................... 34
3.1 Flow chart outlining the proposed structure of the vehicle detection system. ................................................................. 36
3.2 The distributions and Gaussian curve fits of adjoining angles and area ratios, calculated from a database of 300 images of vehicle lamps.

3.3 Histograms and Gaussian curve fit of cross-correlation values ($\max(\gamma)$) from database of rear-lamp and headlamp vehicle images.

3.4 A synthetic test image of two identical circles, the image after an artificial perspective warping and the resultant image after automatic perspective correction.

3.5 Three oncoming vehicles at similar distances, successfully distinguished from each other.

3.6 Forward facing video capture configuration, camera mounted behind rear view mirror.

3.7 GUI developed to collate vehicle detection results.

3.8 Result of rear-lamp cross-correlation pairing process.

3.9 Result of headlamp cross-correlation pairing process.

3.10 A valid vehicle tail-lamp pair and similarly sized tail-lamps from different vehicles.

3.11 Images of headlamps and the correlation between them ($\max(\gamma)$) from a target vehicle through various stages of a 35m radius road bend and the perspective corrected versions of these images and the correlation between them.

3.12 A graph of the correlation before and after perspective correction between headlamps of a target vehicle during a video segment of on a road bend of radius 35m.

3.13 Tail-lamp and headlamp vehicle detection under extreme perspective distortion.

3.14 Successful rear-lamp vehicle detection result frames, representative of results from the video data set.

3.15 Successful headlamp vehicle detection result frames, representative of frames from the video data set. Detection of a target is highlighted by the bounding quadrilateral around the headlamps.

3.16 A sequence of images from a video segment showing the detection and tracking of an approaching, overtaking vehicle.

3.17 Results of successful tracking based detection when a vehicle lamp is distorted because of a turn signal (indicator) lamp.

3.18 False positive detection of vehicles, caused by incorrect matching of background urban light sources and incorrect matching between lamps from different vehicles.

4.1 Examples images of a pedestrian at 35m with dipped headlights with minimal ambient lighting in the visible spectrum and far-infrared.

4.2 Flow chart outlining the complete structure of the IR pedestrian detection system.
List of Figures

4.3 IR images of two closely positioned pedestrians whose torsos are insulated by clothing and corresponding 3D intensity plots. 79
4.4 IR image of a pedestrian and seed regions for region growing, produced by a high threshold, near the maximum image intensity. 80
4.5 The various stages of region growing ROI generation. 82
4.6 Result of region growing ROI generation in a complex urban scene with three pedestrian ROIs and one non-pedestrian ROI. 82
4.7 Examples of pedestrians and non-pedestrians from the database training set. 83
4.8 Images from the various stages of generating a Histogram of Oriented Gradients (HOG) feature vector. 85
4.9 FLIR PathFindIR® automotive far-infrared microbolometer sensor and the sensor mounted on a host vehicle during data capture. 88
4.10 ROC curve for SVM pedestrian classifier. 91
4.11 A montage of results frames displaying successful pedestrian detection, representative of video processing results. 95

5.1 A flow chart outlining the structure of the fusion of the visual and IR systems. 97
5.2 Data capture set up with far-IR microbolometer and visible camera. 98
5.3 Image registration processes with a rectangular calibration target. 99
5.4 A false positive pedestrian detection caused by warm vehicle parts such as tyres and lamps. 100
5.5 Expansion of vehicle lamp bounding box in visual domain to encompass the rest of the vehicle. 100
5.6 Vehicle masking in IR pedestrian detection system through fusion with the visual vehicle detection system. 101
5.7 Pedestrian highlighting in visual image through fusion with IR pedestrian detection system. 102

A.1 A projective transformation is formed by combining the transformations of the bounding quadrilateral to a unit square and an arbitrary rectangle to a unit square. 121

B.1 Model of camera and parameters used to estimate distance. 124


## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Comparison between different sensor modalities for object detection.</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>HSV red colour threshold values for tail-lamp identification</td>
<td>26</td>
</tr>
<tr>
<td>2.3</td>
<td>Results from rear-lamp identification</td>
<td>29</td>
</tr>
<tr>
<td>2.4</td>
<td>Results from headlamp identification</td>
<td>30</td>
</tr>
<tr>
<td>3.1</td>
<td>Statistical analysis of Gaussian curve fit of heuristic features from vehicle lamp image database</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>Statistical analysis of Gaussian curve fit of correlation values from vehicle lamp image database</td>
<td>41</td>
</tr>
<tr>
<td>3.3</td>
<td>Summary of rear-lamp vehicle detection results - by environment type</td>
<td>56</td>
</tr>
<tr>
<td>3.4</td>
<td>Summary of rear-lamp vehicle detection results - by distance</td>
<td>57</td>
</tr>
<tr>
<td>3.5</td>
<td>Rear-lamp detection performance comparison of different camera’s</td>
<td>58</td>
</tr>
<tr>
<td>3.6</td>
<td>Headlamp vehicle detection results summary</td>
<td>59</td>
</tr>
<tr>
<td>4.1</td>
<td>Parameters used to assess ROI region growing stop point</td>
<td>81</td>
</tr>
<tr>
<td>4.2</td>
<td>Parameters used for Histogram of Oriented Gradient feature extraction</td>
<td>85</td>
</tr>
<tr>
<td>4.3</td>
<td>Confusion matrix for SVM pedestrian classifier at chosen operating point, displaying true positive (TP), false positive (FP), false negative (FN) and true negative (TN) rates</td>
<td>91</td>
</tr>
<tr>
<td>4.4</td>
<td>Pedestrian detection video processing results summary</td>
<td>92</td>
</tr>
</tbody>
</table>
List of Abbreviations

ACC  Automatic Cruise Control
ACW  Advanced Collision Warning
ADAS Advanced Driver Assistance Systems
ANSI American National Standards Institute
CIE  Commission Internationale de l’Eclairage
CMOS Complementary Metal Oxide Semiconductor
EV   Exposure Value
FIR  Far-Infrared
GRGB Green-Red-Green-Blue
GUI  Graphical User Interface
HDR  High Dynamic Range
HOG  Histogram of Oriented Gradients
HSV  Hue-Saturation-Value
IR   Infrared
ISO  International Standards Organisation
LED  Light Emitting Diode
PAL  Phase Alternate Line
RBF  Radial Basis Function
RGB  Red-Green-Blue
RMS  Root Mean Squared
ROC  Receiver Operating Characteristics
ROI  Region Of Interest
SVM  Support Vector Machine
VRU  Vulnerable Road User
Chapter 1

Introduction

1.1 Motivation and Background

Automotive safety is an important factor in vehicle design and technology. Vehicle safety assessment programmes, such as the Euro NCAP, have made both manufacturers and consumers more safety-conscious with regards to road vehicles. Recently the focus has extended beyond the occupants of the vehicle and has turned towards Vulnerable Road Users (VRUs) and other road vehicles. The OECD define VRUs as “those unprotected by an outside shield, namely pedestrians and two-wheelers” [1].

Vehicle collision statistics confirm that night conditions are an important area of focus for road safety and collision prevention. In the EU (EU-15) alone, there were more than 43,000 fatal road accidents in 2006 [2]. Almost one third of these road fatalities (32.3%) occurred during the hours of darkness [3], while 45.5% of pedestrian fatalities were in darkness [4]. A similar pattern is reflected in road accident statistics from the USA, with 47.2% of fatal accidents occurring in darkness [5] and 70% of pedestrian fatalities occurring at night (6pm to 6am) in 2008 [6]. In Japan, 55% of all road fatalities occur during night-time [7].

These statistics confirm that the hours of darkness account for a disproportionate amount of road collisions and fatalities, particularly considering that only 28% of
traffic volume occurs during night hours [8]. It has been estimated that per vehicle mile, the road fatality rate is 3 to 4 times higher in darkness than in daylight [9].

A review by VTI, the Swedish National Road and Transport Research Institute [9], provides an in-depth analysis of the numerous factors that contribute to the imbalance between day-time and night-time road accident rates. They found that human vision is not well adapted to night conditions. Visual acuity, contrast sensitivity, spatial resolution, distance perception and reaction time, all deteriorate as overall light levels decrease. VTI state that numerous factors contribute to driver performance at night including low luminance [10], tiredness [11], alcohol [11] and glare from oncoming headlamps [12].

Recent EU regulations place the onus on automotive manufacturers to protect VRUs by introducing new passive safety standards [13] on a phased basis from 2011 to 2019. These include provisions such as specifications for collision tests and geometric constraints for vehicle panels and windscreens. However, a vehicle will be granted immunity from a large section of the requirements if it is equipped with a collision avoidance system. This is based on a European Commission study [14] that found that VRU protection can be significantly improved by a combination of passive and active measures. While passive safety systems aim to reduce injury on impact, active measures aim to avoid or mitigate a collision. Automotive manufacturers may soon also be legally obliged to make blind zones visible to the driver in EU, USA and Japan, with vision systems providing an effective way to meet the legislative requirements [15].

Demand for Advanced Driver Assistance Systems (ADAS) is expected to increase as consumers grow increasingly safety conscious and insurance companies and regulators begin to recognise the positive impact such systems can have on accident rates. Concurrently, vision systems are becoming increasingly feasible in road vehicles due to advances in technology and lower costs. Some vehicles are currently equipped with vision systems used to aid parking, reversing and viewing of blind zones. Greater
functionality can be extracted from these cameras by using them to detect other road vehicles. The detection of other road vehicles is a core component of many ADAS such as Automatic Cruise Control (ACC) [16], Advanced Collision Warning (ACW) [17], overtaking vehicle monitoring [18], automatic headlamp dimming [19] and automatic blind zone monitoring [20]. Standard colour cameras are often used for these functions as they are low cost, readily available and may already be present on the vehicle. The full-colour video produced can be simultaneously used for other ADAS utilising colour data and for driver display.

However, visual spectrum cameras are limited when it comes to the task of detecting pedestrians in low light conditions, as they can only observe objects illuminated by vehicular lighting, street lighting or moonlight. Removing the near-infrared blocking filter from a visual camera can extend its low light pedestrian detection capabilities towards the near-infrared part of the spectrum (700nm-1000nm), when the target is illuminated with an IR light source. It has been found that Far-Infrared (FIR) thermal night vision technology yields substantially greater detection distances for drivers detecting pedestrians than these near-infrared systems [21]. Far-infrared technology is well suited to the task of night-time automotive pedestrian detection as thermal radiation from humans peaks in the 8-14µm spectral band, therefore no illumination of the target or scene is required. Pedestrians are generally warmer than the background environment, especially at night, so they appear with higher intensity in FIR imagery, thus aiding automatic identification.

This thesis presents a system for the automatic detection and tracking of other road vehicles and pedestrians in low light conditions.

Other road vehicles are detected in the visual spectrum by searching for their most distinguishing features at night: rear lamps and headlamps. These automotive lamps are identified by image processing analysis of low exposure road video from a regular colour camera. Rear-lamps are extracted using colour thresholds derived from global regulations and adapted to an electronic imaging environment. Headlamps are
extracted using high intensity seeds and a region growing technique. Vehicles are classified through symmetry analysis of these lamps, including correction for perspective distortion caused by road bends or relative position. Detected vehicles are tracked between frames using the Kalman filter.

Pedestrians are detected in far-infrared thermal night vision road video from an automotive microbolometer sensor. A technique is introduced to compensate for distortion caused by well insulating clothing, aiding extraction of pedestrian candidates by a shape-feature based region growing technique. ROIs are classified by a Support Vector Machine (SVM) trained using Histogram of Oriented Gradients (HOG) features and pedestrians are tracked using the Kalman filter.

A fusion algorithm is also presented which combines the vehicle and pedestrian detection systems in a low light road user detection system. It is shown that vehicle detection information can lower the probability of false positives in pedestrian detection while infrared image data and pedestrian detection data can be used to enhance a visual camera view for human consumption.

1.2 Objectives and Contributions

The objective of this thesis is the development of an automotive image processing system to detect other road vehicles and pedestrians in low light conditions. More specifically, the primary contributions of this thesis are:

1. A configuration process for a visual spectrum colour camera is proposed, which optimises the appearance of automotive lamps for identification. This ensures that device independent rear-lamp colour boundaries (specified in regulations) can be applied to a device dependent camera colour space and that the appearance of vehicle lamps in captured imagery is optimised for detection.

2. An algorithm for identifying vehicular rear-lamps from low exposure images is
presented. A red colour threshold is derived from international regulations for rear-lamps and adapted to real world conditions in the HSV colour space to identify rear-lamps.

3. A region growing algorithm is proposed for the identification of vehicular head-lamps.

4. A vehicle detection system is proposed, based on the symmetry analysis of lamp pairs using cross-correlation. Automatic correction for perspective distortion is implemented and vehicles are tracked using the Kalman filter.

5. An infrared pedestrian detection system is presented, which compensates for well insulating clothing worn by pedestrians. Pedestrians are detected using region growing, Histogram of Oriented Gradients (HOG) features and a Support Vector Machine (SVM) classifier before being tracked using the Kalman filter.

6. A night-time road user detection ADAS is described, which intelligently combines the visual vehicle detection and IR pedestrian detection systems. Knowledge of the location of detected vehicles is utilised to mask the corresponding IR image and lower the probability of false positives. The IR image of detected pedestrians is fused with the corresponding visual image to create an enhanced image, where pedestrians are highly visible.

1.3 Chapter by Chapter Overview

The thesis contains six chapters in total, including this introductory chapter. The contents of the remaining chapters are as follows:

In Chapter 2, a system for identifying vehicular rear-lamps and headlamps is presented. A camera configuration process ensures optimal appearance of vehicle lamps for identification. Regulations specifying the red colour boundaries for rear
automotive lamps are converted to the colour space utilised by the camera equipment and adapted to real world conditions. A region growing threshold is presented for the identification of headlamps, based on the growing of high intensity seeds to edge detection boundaries.

A vehicle detection system is proposed in Chapter 3 based on the pairing of identified lamps. Lamps are paired using cross-correlation analysis of bilateral symmetry. A projective image transformation is introduced which automatically compensates for perspective distortion of vehicle lamp pairs, typically encountered when not viewing the target vehicle at a right angle. Multiple detected vehicles are tracked concurrently using the Kalman filter.

A far infrared pedestrian detection system is described in Chapter 4. A morphological technique to compensate for clothing-based distortion of pedestrians is introduced. This aids segmentation of pedestrians wearing well insulating clothing considerably. A feature-based region growing algorithm produces Regions Of Interest (ROIs) which are subsequently classified using HOG features and an SVM classifier. Detected pedestrians are tracked using the Kalman filter.

Chapter 5 presents a fusion of the visual vehicle detection and IR pedestrian detection systems. The locations of vehicles found in the visible environment are used to reduce the likelihood of false positives in IR pedestrian detection by masking vehicular hotspots. Pedestrian location data from the IR domain is used to enhance a forward facing visual image for driver consumption.

Finally in Chapter 6, the project is summarised, conclusions are drawn and potential directions for future work are considered.
Chapter 2

Vehicle Lamp Identification

2.1 Introduction

One of the most significant categories of on-road entities that drivers need to be aware of are other road vehicles. Driver assistance systems that automatically track the positions of other road vehicles can be used for numerous functions, such as: collision warning, blind spot monitoring and automatic cruise control. A camera is an inexpensive and versatile sensor to utilise for this task, but at night the amount of visual information available in video is extremely limited compared with day time. Many useful background cues such as road markings, lane boundaries and the horizon line as well as target cues such as the vehicle body structure itself or shadows are lost or highly diminished [22]. The most salient vehicle features in dark environments are the forward and rear facing lamps [22].

While driving at night, vehicles on the road are primarily visible by red coloured rear facing lamps and clear coloured headlamps. While all vehicle lamps will differ in appearance, they must adhere to automotive regulations which provide a set of characteristic features that can be utilised by image processing systems for identification.

This chapter describes a system for identifying these automotive lamps from on-road video data. The chapter is divided into four parts: firstly a review of research
Chapter 2: Vehicle Lamp Identification

in the area is presented. Secondly, a camera configuration process is described. This ensures that the appearance of lamps in video data is optimal for identification and is not affected by ambient lighting conditions. This is followed by an image processing algorithm to identify rear facing vehicle lamps in video frames. Finally an image processing algorithm to identify forward facing vehicle headlamps is described.

2.2 State of the Art

2.2.1 Advanced Driver Assistance System (ADAS) Landscape

Many different sensor technologies can be utilised for on-vehicle object detection. Table 2.1 outlines the advantages and disadvantages of the most common sensors used for ADAS. This table has been reproduced from a review conducted by Gandhi and Trivedi [23].

Many commercial and design factors will also influence the presence of these sensors on road vehicles. In this research, for night vehicle and pedestrian detection, a monocular visual camera was used for vehicle detection and a monocular thermal IR camera was used for pedestrian detection. Monocular systems were chosen as it is envisaged that stereo systems will only be feasible on-board premium vehicles for cost reasons. Two-dimensional systems that produce imagery were chosen (as opposed to one dimensional systems such as radar) as they can be used to convey information to the driver in an intuitive manner, that is easy to understand and consume: visual media are driver centric.

The characteristics of the chosen sensors allow for robust detection performance in the near- and medium-ground. The large fields of view ensure that objects at close range but not directly in front of the vehicle can be detected. For long range object detection it is envisaged that these systems could be supplemented by a narrow field of view, long range system such as radar.
Table 2.1: Comparison between different sensor modalities for object detection.

<table>
<thead>
<tr>
<th>Sensor Type</th>
<th>Field Of View</th>
<th>Detection Range</th>
<th>Illumination</th>
<th>Hardware Cost</th>
<th>Algorithmic Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectilinear Camera</td>
<td>Med.</td>
<td>Low/Med.</td>
<td>Passive reflective, needs ambient light</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Omni Camera</td>
<td>Large</td>
<td>Low</td>
<td>Passive reflective, needs ambient light</td>
<td>Med.</td>
<td>High</td>
</tr>
<tr>
<td>Near IR</td>
<td>Med.</td>
<td>Med</td>
<td>Active, works in dark</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Radar</td>
<td>Small</td>
<td>High</td>
<td>Active, works in dark rain, fog</td>
<td>Med.</td>
<td>Low</td>
</tr>
<tr>
<td>Laser Scanner</td>
<td>Large</td>
<td>Med.</td>
<td>Active, works in dark</td>
<td>High</td>
<td>Low</td>
</tr>
</tbody>
</table>

An extra level of complexity must be considered when dealing with on-vehicle vision systems, as background data will not be still. This means that image processing techniques commonly used with static camera systems, such as background subtraction are usually not relevant.

2.2.2 Literature Review

Vehicular headlamps appear as some of the brightest regions in a frame of nighttime automotive video, so it is common for image processing techniques for lamp detection to begin with some form of thresholding. Grey scale or brightness thresh-
Chapter 2: Vehicle Lamp Identification

Holding is a common starting point [19, 24, 25, 26, 27]. Oncoming headlamps are identified by thresholding and applying a top-hat filter in [19]. The resultant pixels are then grouped and labeled to analyze characteristics such as area, position and shape. Further analysis is required as there are many potential candidate light sources that are not vehicle lamps, such as street lamps, infrastructure lighting or background light sources. Region growing to edge pixels has been used to detect vehicles in night vision [28].

Colour filtering is a useful tool in automotive machine vision [29] and has been used in the past to improve detection rates of road signs [30] and to distinguish vehicle bodies from the background [31]. Employing a red colour filter has been shown to be an effective way to identify rear facing automotive lamps. Many different colour spaces with widely varying parameters have been used to detect red coloured light regions from images, but have been based on subjective colour boundaries. No known previous research detects automotive rear-lamps based on the red colour limits set out by UN regulations. The most common approach makes use of the Red-Green-Blue (RGB) colour space [32, 33, 34, 35, 36]. Separate RGB thresholds for brightness and redness are implemented by Sukthankar [35] while Betke et al. [32] analyze only the red channel of the RGB data. Bright non-red light sources however can have a significant red channel component, so analysis of the red channel alone is not sufficient for red colour thresholding. The RGB colour space has also been used for detection of red vehicle lamps in daylight conditions [37, 38].

The RGB colour space is not ideal for the task of colour thresholding, it is difficult to set and manipulate colour parameters due to high correlation between R, G and B channels [39]. Previous research in the area has opted to use other colour spaces for a red colour threshold to overcome this difficulty. Cabani et al. use the L*a*b colour space to detect red vehicle lamps [40], because it requires only two colour threshold operations compared to three for RGB and is hence more computationally efficient. The claimed advantage is somewhat negated as the data must first undergo a colour
space transformation. The red colour threshold parameters were also chosen subjectively. Grey scale algorithms can also be applied using only the lightness channel $L$. The YCbCr space is used to detect redness of lamps in [25] and [41] where the red-difference chrominance channel $Cr$ is analysed to distinguish which bright areas are red light sources. Manual HSV colour thresholds are defined in [42] for identification of tail lamps and headlamps.

The appearance of lamps in captured video and imagery is highly dependent on the camera settings and sensor characteristics and varies widely in previous work. It has been observed by some that rear-lamps usually appear as white regions with red surroundings as they can partially saturate image sensors [32, 34]. The perimeter of the saturated white regions is analysed in [34] and if the surrounding area is comprised mostly of red pixels it is regarded as a potential rear-lamp. Kim et al. [43] encountered this blooming effect while detecting vehicle headlamps. They extracted bright regions using thresholding and knowledge of the vehicle size.

The task of detecting vehicle lamps can be aided considerably by utilising novel and non-standard camera hardware. Current commercial automatic headlamp dimming systems use custom novel hardware filters. Non-Bayer patterns with combinations of red and clear filters have proven effective in detection of red lamps and differentiating between tail-lamps and other light sources [44, 45]. Another commercial system uses two coated lenses (one blocking red light, one allowing only red light to pass) to focus incoming light onto different parts of a single imaging sensor [46, 47]. While custom hardware can offer performance advantages, these cameras do not produce full-colour images, so they cannot be reused for other ADAS that utilise full-colour image processing and have limited use in terms of displaying video to the driver.

Physical assumptions about lamps have been used in previous work to filter potential light candidates and aid the detection process. It has been assumed that lamps are circular or elliptical in shape [24, 36, 48]. However the shape of lamps is not
specified in the regulations and as LED automotive lamps become more prevalent, lamps are now being manufactured in substantially different shapes.

Optical flow has been used to detect approaching vehicles in day and night situations [49]. This technique is suited to dark environments where there is ample contrast between vehicle lamps and background. Many non-vehicle light sources such as street lamps can be easily removed using a simple assumption about vehicle position, such as masking the top segment of the image [50].

2.3 Camera Configuration

Lighting conditions in road environments can vary dramatically at night as ambient illumination can commonly consist of a combination of: vehicle lamps, street lights, infrastructure lighting and moonlight. Any system for identifying automotive lamps at night must be able to adapt to these conditions or mitigate their impact.

It is beneficial to use standard colour cameras for the task of detecting other vehicles as they are low-cost, readily available and may already be present on the vehicle performing other functions, such as display of blind zones to the driver. The colour video produced by these cameras can be simultaneously used for other ADAS utilising colour data and for full colour display purposes. Furthermore, camera systems are passive, meaning that interference with similar systems on other vehicles is not an issue, as is sometimes the case with active systems such as radar. The performance of camera based systems can however be affected by weather conditions.

CMOS image sensors are low cost and are widely available, but they have an extremely limited dynamic range in comparison to the human eye. While a human eye is sensitive from 0.1 lux to 10,000 lux, a CMOS camera sensor is normally only sensitive in the range of 6 to 15 lux. Standard colour digital video cameras use a Bayer filter [51] to separate incoming light into red, green and blue channels. A Bayer filter consists of a matrix of colour filters over sensor pixels, one quarter blue, one quarter
red and one half green, imitating the physiology of the human eye. A typical GRGB (Green-Red-Green-Blue) Bayer Filter pattern is shown in Figure 2.1. This filtering process reduces the resolution of a digital image captured from an image sensor to one third of the full resolution of the sensor.

![Figure 2.1: Typical Bayer filter GRGB pattern.](image)

The characteristics of camera systems have a substantial effect on the properties of captured images. The colour and intensity of pixels in captured images representing rear-lamps will vary depending on the camera and its configuration. This means results from the literature cannot be accurately replicated or verified and previous techniques cannot be easily ported to different camera hardware. Suitable configuration of the camera hardware has not been addressed directly in any known prior work in this area.

A two stage camera configuration is proposed in this section: firstly, a static expo-
sure level is set based on an ANSI photography standard. This ensures the incoming light does not saturate the camera sensor, ensuring maximum colour and shape information from vehicle lamps is available in the captured image frame. Secondly, colour consistency is considered, so that the device-dependent digital colour data can be compared to the device-independent visible light colour boundaries for tail-lamps defined in automotive regulations. These configurations also ensure that the appearance and more specifically the colour, intensity and shape of lamps in images are not affected by ambient lighting conditions. These configurations are derived using widely accepted and commonly used photography terms.

2.3.1 Exposure Control

Control of camera exposure is essential to the process of photographing light sources such as headlamps and tail-lamps of road vehicles. Too low an exposure setting and the target appears too dark, such that that important image detail, including vehicle lamps, can be lost. Too high an exposure setting and the intensity of the incoming light causes saturation in the image and the appearance of lamp colour, shape and position can be distorted. This phenomenon is commonly referred to as “blooming”. Exposure is controlled automatically by default in most cameras. Automatic exposure control algorithms are generally optimised for daylight operation and are not suitable for detecting light sources at night. In images captured with automatic exposure enabled, lamps typically appear over-saturated. Examples of this are displayed in Figure 2.2 (a) and (c). This blooming effect (where lamps appear as large saturated regions) is undesirable, as the shape of lamps can be distorted and all colour information lost in saturated pixels from rear-lamps. Some previous research has attempted to extract vehicle location from detection of these bloomed regions [34],[43]. Under automatic exposure control, the appearance of vehicle lamps can also change significantly between varying lighting conditions. The exposure is automatically adjusted to compensate for overall scene light-level changes, such as moving between
illuminated and unilluminated environments. Previous work in traffic-signal detection at night has utilised a low static exposure value [52], but research in automotive lamp detection in has not addressed the issue of exposure control.

Figure 2.2: Video frames captured with: (a), (c), automatic exposure control enabled and (b), (d), the same scenes captured using a lower static exposure \( EV_{100} = 10 \).

High Dynamic Range (HDR) cameras have been used to detect traffic signals (traffic lights) at night [53]. HDR cameras fuse multiple images with different exposures, aiming to produce an image that captures detail from a large dynamic range of luminances, from low intensity regions to high intensity regions. Since only the high intensity parts of the image are of interest for this application, some of this
HDR functionality can be imitated with standard hardware by setting a low exposure with regular camera hardware. Setting a low static camera exposure value can ensure vehicle lamps appear as separate or distinct regions in captured images with full colour information. There are several advantages to this process over the use of a HDR camera. Firstly, regular cameras typically allow for more precise control of exposure than HDR. Secondly, a low exposure setting results in only high intensity objects being visible in the resulting image and lower intensity background objects do not appear. This aids the vehicle lamp identification process and reduces the amount of false positives as pixels that are not from light sources are close to zero in intensity and are unlikely to be identified as vehicle lamps. Example images captured with a low exposure configuration are displayed in Figure 2.2 (b) and (d). This greatly assists the detection process and ensures the appearance of target vehicles is not significantly influenced by ambient illumination conditions. These distinct lamp image regions are more representative of the true lamp shape and also allow for a more accurate determination of the vehicle’s size and position than an enlarged bloomed region.

Note how the rear-lamps in Figure 2.2 (a) saturate the image sensor, resulting in the red rear-lamps appearing white towards the centres of the lamps, whereas in Figure 2.2 (b) the colour of the red lights is preserved. It should be noted that in Figure 2.2 (b) much of the lower intensity background detail from the scene is not present, due to the lower exposure setting. In Figure 2.2 (c), oncoming headlamps cause severe blooming when captured with automatic exposure control, resulting in corruption of position data for that vehicle and lamp pairs can even become merged together in some cases. When captured with the proposed low static exposure configuration, such as Figure 2.2 (d), the lamps appear as separate, distinct regions from which the vehicles location can be determined.

Camera exposure is defined in terms of three parameters: exposure time \((\text{shutter speed})\), aperture \((F\text{-number})\) and sensitivity \((\text{ISO value})\). An Exposure Value \((\text{EV})\)
Chapter 2: Vehicle Lamp Identification

...can be determined from these variables. This value can be transferred between cameras to ensure equivalent levels of exposure between different hardware. Exposure Value ($EV$) is defined by

$$EV_{ISO} = \log_2 \frac{N^2}{t}$$  \hspace{1cm} (2.1)

where $N$ is the camera aperture, $t$ is the exposure time and $ISO$ is the camera sensor sensitivity to light. The aperture is the size of the opening in the camera system where light is allowed in, this is typically described in terms of the focal length of the lens. While the $ISO$ value has no bearing on $EV$, it is essential information needed to reproduce image capture conditions and is therefore commonly added as a subscript. In the vehicle lamp identification system presented later in this chapter, an exposure value of $EV_{100} = 10$ ( $EV = 10$ at $ISO = 100$) was adopted from the ANSI photography standard exposure guide [55] for neon lights and other bright signs. This setting is the one which best approximates the scenario presented in this chapter. Direct observation of illuminants is quite an unusual imaging scenario. It has been observed that this exposure configuration ensures vehicle lamps consistently appear as distinct regions (not merged due to blooming), representative of the true lamp shape (not enlarged due to blooming) and that rear-lamps appear with full colour information (not saturated). This setting is comparable to the heuristically set static EV of 12.3 set in [52] for traffic signal detection. The result of this low exposure configuration is that lower intensity detail, which is not of interest to this system, is diminished in the resultant scene. This reduction in low intensity detail further simplifies the lamp identification process. While it may not be entirely practical to reuse these images for human consumption, it is envisaged that in an embedded implementation it may be possible to extract both standard and low exposure images from the same sensor if required.

The regulations specify that the minimum acceptable intensity of tail-lamps is 4 cd while the maximum acceptable intensity of brake-lamps is 185 cd. To investigate...
the suitability of the chosen static exposure setting for retaining rear-lamp colour information, images of two extreme situations were considered: brake-lamps at 3m and tail-lamps at 50m. Figure 2.3 displays these images.

![Figure 2.3: Extreme test conditions for static exposure setting (EV$_{100}$ = 10). (a) A brake-lamp captured at 3m, (b) Tail-lamps captured at 50m.](image)

With this exposure configuration, brake-lamps at a distance of 3m will partially saturate in the image, with typically 10-15% of pixels saturating. Although saturation of any number of pixels is not desirable, a certain amount is tolerated to maximise the range of distances the lamp identification algorithm can operate. Experiments have shown that if exposure is lowered until there is no saturation in the lamp image at close range, the maximum operating distance of the system is greatly reduced.

Ambient lighting conditions are usually a significant factor in photography when configuring a camera. Colour and intensity of ambient light are usually important factors to consider and video cameras typically implement automatic exposure and white balance algorithms by default, to compensate for changes in ambient lighting. Objects in most scenes are visible because they reflect some of this ambient light. However when photographing a light source such as a vehicle lamp, the lamp is primarily visible because of light emitted from its internal light source. The intensity of ambient light reflected from the surface of the lamp is negligible by comparison. Therefore, with a static exposure configuration, changes in ambient lighting intensity, such as driving from an illuminated to unilluminated area of road, do not have a
significant effect on the appearance of the lamps.

2.3.2 Colour Configuration

To ensure that the device dependent digital colour data in captured images can be evaluated with regards to device independent visible light colour boundaries for rear-lamps defined in automotive regulations, camera equipment must be configured with this in mind. The camera is configured to disable all automatic colour processing operations including automatic white balance, as this artificially distorts colour information. The grey world assumption [56] used for many automatic white balance algorithms is not a valid assumption in night conditions. A manual white balance is set to the reference white point of the colour gamut of the camera. The camera hardware used for this thesis conforms to the PAL colour specification, for which the white point is the CIE Standard Illuminant D65 (6500K). As the influence of reflected light from external light sources on the captured pixel colour of the tail- and brake-lamps is negligible, this static white balance configuration ensures that ambient scene illumination is not a significant factor in colour representation. This is an important step as ambient road lighting conditions at night can vary dramatically in colour, from orange-yellow street lamps to blue-clear moonlight.

2.3.3 Road Reflections

Headlamp and tail-lamp reflections can be prevalent on the road surface, particularly in wet weather conditions. This could lead to false positive identification of lamps in some instances, especially if the reflections are high in intensity. In [19] headlamps are distinguished from reflections using a Support Vector Machine (SVM) classifier. In this system the issue of reflections is addressed at an earlier stage by equipping the video camera with a circular polarising filter. Light reflected from non-metallic surfaces becomes polarised, so this filter greatly reduces the intensity of
reflections in video data. This effect is shown in Figure 2.4.

![Image](image_url)

Figure 2.4: (a) A circular polarising filter placed in front of the camera lens is used to reduce the intensity of reflections. (b) Without the filter, bright light reflections are prevalent on a wet road surface. (c) Reflections are almost completely removed by the filter.

### 2.4 Red Rear-Lamp Identification

In this section an image processing algorithm for identifying red rear vehicle lamps from frames of automotive video is described. Bright objects such as street lamps, traffic signals, turn signal lamps, headlamps and reflections from road infrastructure are filtered out, while retaining the rear-lamps of target vehicles.

#### 2.4.1 Regulation to Device Dependent Colour Space

Previous rear-lamp detection image processing systems manually define subjective colour and brightness thresholds. In this thesis, colour boundaries from automotive regulations are utilised to derive image threshold parameters.

The trichromatic coordinates specifying the regulation colour for red rear-lamps [57] are defined in terms of the CIE-1931 [58] tristimulus coordinates, $x$, $y$ and $z$. This colour space was derived from human observations and is device independent.
The regulation red colour limits for rear-lamps are:

Limit towards yellow: \( y \leq 0.335 \) \hspace{1cm} (2.2)

Limit towards purple: \( y \geq 0.980 - x \) \hspace{1cm} (2.3)

These specifications can be observed overlaid on the CIE xy chromaticity diagram in Figure 2.5(a), with the area of acceptable visible light encompassed by the two linear inequalities. Figure 2.5(b) shows the PAL colour gamut and the acceptable rear light colour region overlaid on the CIE xy chromaticity diagram.

Figure 2.5: CIE 1931 xy chromaticity diagram showing (a) the regulation limits of light for red rear automotive lamps and (b) the PAL colour gamut.

To derive colour threshold parameters from these limits, the enclosed region is transformed to the PAL RGB colour space (the gamut of the camera hardware).
The thresholds specified in the regulations and the chromaticity diagram are in two-dimensional xy space. This is a 2D representation of the 3D xyY colour space, the luminance dimension (Y) is omitted from the diagram. For conversion to RGB, all luminance values are transformed. The transformation of this three-dimensional CIE xyY region to PAL RGB is conducted through the CIE XYZ colour space [59]. The first step is to convert the xyY region to XYZ using the following equations:

\[
X = \frac{xY}{y} \quad (2.4)
\]

\[
Z = \frac{(1 - x - y)Y}{y} \quad (2.5)
\]

\[
Y_{(XYZ)} = Y_{(xyY)} \quad (2.6)
\]

The luminance dimension (Y) in these spaces is the same, hence it has the same symbol. This XYZ space is then converted to PAL RGB (2.7) - (2.9).

\[
[r \ g \ b] = [X \ Y \ Z][M]^{-1} \quad (2.7)
\]

Where \([M]^{-1}\) is the inverse transformation matrix. This can be calculated from the reference primary coordinates of the colour system, which for PAL RGB results in:

\[
[M]^{-1} = \begin{pmatrix}
3.2404542 & -0.9692650 & 0.0556434 \\
-1.5371385 & 1.8750108 & -0.2040259 \\
-0.4985314 & 0.0415550 & 1.0572252
\end{pmatrix} \quad (2.8)
\]

The final RGB values are calculated using the gamma value of the colour space \(\gamma\):

\[
R = r^{\gamma}, \ G = g^{\gamma}, \ B = b^{\gamma} \quad (2.9)
\]
Chapter 2: Vehicle Lamp Identification

For PAL RGB the value of $\gamma$ is 2.2. The resulting regulation colour region in RGB space is shown in Figure 2.6 (a). Observe the “tail” segments splitting from the main region. These artefacts are the result of clipping at the maximum $R$ level. This is because the regulation colour region specified extends outside the colour gamut of the most common RGB spaces used in digital imaging such as sRGB, PAL and NTSC as is visible in Figure 2.5 (b).

![Figure 2.6](image)

(a)

(b)

Figure 2.6: (a) Rear-lamp regulation red colour region transformed into RGB space and (b) RGB scatter plot of pixels from database of tail- and brake-lamp images.

A database of 300 tail- and brake-lamp images was created to observe the colour distributions of rear-lamp pixels. Figure 2.6 (b) shows an RGB distribution of pixels from this database. It can be observed from this scatter plot that red rear-lamp pixels do not conform directly to the derived regulation region.

2.4.2 Adaptation to Real World Conditions

To adapt the regulation colour region to real world images it is firstly converted from RGB to the Hue-Saturation-Value (HSV) colour space [60], as it is more intuitive to adjust and manipulate the threshold parameters than RGB. The HSV colour space is more representative of the way humans observe colour than the commonly used
RGB space, which was created mainly for electronic representation of colour and has highly correlated channels. HSV is best represented as an inverted cone, with Hue (tint) as the angle (0-360°), Saturation (shade) as the radius (0-1) and Value (tone) as the perpendicular height (0-1).

Hue describes the shade of a colour and where that colour is found in the colour spectrum. Saturation describes how pure the hue is with respect to a white reference. The Value dimension is a relative measure of the amount of light illuminating a colour. The colour red is located around the hue value of 0°. The resulting regulation HSV region is shown in Figure 2.7 (a).

![Figure 2.7](image)

Figure 2.7: (a) Rear-lamp regulation red colour region mapped to conical HSV colour space. The highlighted section represents the derived colour threshold value limits. (b) Conical HSV scatter plot of pixels from database of tail- and brake-lamp images.

The $H$ colour threshold limits were extracted directly from the limits of this distribution. The $V$ component of the distribution spans the entire range of possible values. However it is undesirable to allow the entire range of $V$ values through the colour threshold, as hue and saturation are inaccurate and unpredictable at very low levels of $V$ and many background pixels would be allowed through. Therefore the lowest $V$ values from the threshold are blocked. Figure 2.7(b) shows a HSV scatter plot of pixels taken from the database of tail- and brake-lamp images.
While the regulations specify fully saturated colour for rear-lamp pixels, the saturation component of a colour is somewhat dependent on the intensity of the incident light. It can be deduced from Figure 2.7(b) that pixels from the database of lamp images occupy a range of saturation values. The $S$ threshold limit is derived from a histogram of these saturation levels, displayed in Figure 2.8. A Gaussian curve was fit to the histogram data and the threshold was established at $0.4645$, the lower 95.4% normal probability point ($\mu - 2\sigma$).

![Figure 2.8: Histogram illustrating the distribution of saturation ($S$) values from database of captured tail- and brake-lamp images and a Gaussian curve fit to the data ($\mu = 0.7841, \sigma = 0.1598, \mu - 2\sigma = 0.4645, RMSE = 16.78$).](image)

Final rear-lamp colour threshold values are presented in Table 2.2 and overlaid on Figure 2.7 (a). The input frame is median filtered to reduce noise

$$\hat{f}(x, y) = \text{median}_{(s,t) \in S_{xy}} \{g(s, t)\}$$

before performing a colour threshold with these values.

$$L_{(x,y)} = \begin{cases} 1 & \left( H_{\text{min}} \leq H_{(x,y)} \leq H_{\text{max}} \right) \& \left( S_{(x,y)} \geq S_{\text{min}} \right) \& \left( V_{(x,y)} \geq V_{\text{min}} \right) \\ 0 & \left( H_{\text{min}} > H_{(x,y)} > H_{\text{max}} \right) \left( S_{(x,y)} < S_{\text{min}} \right) \left( V_{(x,y)} < V_{\text{min}} \right) \end{cases}$$
This produces a binary image indicating the position of red lamp pixels in the image.

Table 2.2: HSV red colour threshold values for tail-lamp identification

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hue ($H$)</td>
<td>342°</td>
<td>9°</td>
</tr>
<tr>
<td>Saturation ($S$)</td>
<td>0.4645</td>
<td>1.0</td>
</tr>
<tr>
<td>Value ($V$)</td>
<td>0.2</td>
<td>1.0</td>
</tr>
</tbody>
</table>

The binary image resulting from the red colour threshold is morphologically closed to remove noise.

$$f \circ b = (f \oplus b) \ominus b$$ (2.12)

This also merges closely located lamp segments which may have been fragmented by the colour threshold. Connected pixel regions are then labelled and considered as rear-lamp candidates.

### 2.5 Headlamp Identification

In this section the algorithm for identifying automotive headlamps in frames of automotive video is described. Headlamps remain among the brightest objects in video captured with the fixed low exposure configuration. Headlamps are identified in the video data using seeded region growing \[61\]. A region growing technique is defined by the method used to choose seeds, the inclusion criteria for neighbouring pixels and the stop criteria, to determine when the region has finished growing \[62\]. For this application, seeds are found by applying a very high fixed threshold $T$, near the maximum intensity value of the image.

$$g(x, y) = \begin{cases} 
1 & (f(x, y) > T) \\
0 & (f(x, y) \leq T) 
\end{cases}$$ (2.13)
Chapter 2: Vehicle Lamp Identification

The threshold is then systematically lowered; neighbouring pixels that satisfy the threshold are included in the region as it grows. Growth stops when the border of the growing region closely matches pixels from a Sobel edge detection of the original image. The Sobel edge detection is formed by combining horizontal and vertical gradient images.

\[
G_y = \begin{pmatrix}
-1 & -2 & -1 \\
0 & 0 & 0 \\
+1 & +2 & +1
\end{pmatrix} * A \tag{2.14}
\]

\[
G_x = \begin{pmatrix}
-1 & -2 & +1 \\
0 & 0 & 0 \\
-1 & +2 & +1
\end{pmatrix} * A \tag{2.15}
\]

\[
G = \sqrt{G_x^2 + G_y^2} \tag{2.16}
\]

The threshold for the Sobel edge detector is calculated automatically based on an RMS estimate of the image noise (the mean of the magnitude squared) \[63\]. It was found that there is typically a significant intensity difference between headlamps and the surrounding pixels. Hence, there is usually a large margin of error for setting the threshold. It is therefore not a parameter that has a significant effect on results.

Similar to the rear-lamp identification process, the resulting binary image is morphologically closed to remove noise and reconnect nearby regions which may have become separated in the thresholding process. This technique is more flexible than using a single threshold. By using region growing only the brightest regions are detected, due to the high seed threshold, while the region’s shape is preserved, due to the growing step. Seeds are sometimes found in other high intensity background objects, such as street-lamps. Street lamps were found to be a common source of false positives, so they are addressed by masking the top third of the image, similar to \[50\]. The various stages of the region growing headlamp identification method are
displayed in Figure 2.9.

Figure 2.9: The stages of region growing thresholding of a vehicle headlamp image: (a) the original video frame containing headlamps, (b) seed regions produced by applying a high static threshold, (c) result of a Sobel edge detection of (a) clearly showing the headlamp boundaries, (d) binary image produced by growing seed regions to the edge boundary.

2.6 Results

In this section results will be presented demonstrating the effectiveness of the system at identifying automotive lamps from low exposure on-road video captured with the proposed camera configuration. Video data was captured using a Sony Handycam CX6EK, configured as described in Section 2.3.

A set of 50 test image frames containing rear-lamps was extracted from on-road video. The results of the rear-lamp identification process from these images are pre-
Chapter 2: Vehicle Lamp Identification

Rear-lamps are identified with a high degree of success, with a low false positive rate.

Table 2.3: Results from rear-lamp identification

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Images</td>
<td>50</td>
</tr>
<tr>
<td>No. Rear-Lamps</td>
<td>124</td>
</tr>
<tr>
<td>No. Rear-Lamps Identified</td>
<td>120</td>
</tr>
<tr>
<td>Rear-Lamp Identification Rate</td>
<td>96.77%</td>
</tr>
<tr>
<td>No. False Positives</td>
<td>13</td>
</tr>
<tr>
<td>False Positives Per Image</td>
<td>0.260</td>
</tr>
</tbody>
</table>

Low exposure night video frames showing the rear vehicle lamps and binary images resulting from the rear-lamp identification process are presented in Figure 2.10. These images were captured in various lighting conditions and encompass vehicles at a range of distances. Vehicular rear-lamp light sources are successfully identified in all cases, while other common light sources such as street lamps and signs are excluded. This is particularly notable in Figure 2.10 (g) and (h), where a traffic signal, headlamps, street lamps and various reflections are all filtered out and the tail-lamps with a non-standard shape are successfully identified. In Figure 2.10 (c) and (d) a vehicle’s brake-lamps are successfully identified despite the visibility of an indicator lamp. Lamps from vehicles at greater distances are identified correctly in Figure 2.10 (b) and (f).

A set of 50 test image frames containing vehicle headlamps was also extracted from on-road video. The results of the headlamp identification process from these images are presented in Table 2.4. As with rear-lamps, headlamps are identified with a high degree of success, with a low false positive rate.

A montage of headlamp identification results, encompassing vehicles at different distances and in different lighting conditions is presented in Figure 2.11.

The main causes of false positives are background and infrastructure lighting; hence they occur most frequently in urban and built up environments. False positives are more common in the headlamp detection process than the rear-lamp detection.
Table 2.4: Results from headlamp identification

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Images</td>
<td>50</td>
</tr>
<tr>
<td>No. Headlamps</td>
<td>132</td>
</tr>
<tr>
<td>No. Headlamps Identified</td>
<td>125</td>
</tr>
<tr>
<td>Headlamp identification Rate</td>
<td>94.70%</td>
</tr>
<tr>
<td>No. False Positives</td>
<td>35</td>
</tr>
<tr>
<td>False Positives Per Image</td>
<td>0.70</td>
</tr>
</tbody>
</table>

process, as other intense light sources such as street lamps can sometimes be detected in the region growing process. Lamp reflections from wet surfaces caused a negligible amount of false positives as the intensity of these reflections was greatly diminished by the polarising filter. Examples of false positive rear-lamp and headlamp identifications are shown in Figure 2.12.

Failure to identify lamps usually occurs when a vehicle is at a large distance from the camera. At large distances, rear-lamps can appear faint and are sometimes not sufficiently intense to satisfy the colour threshold. At large distances headlamps may appear very small in size and sometimes cannot be distinguished from background noise or may blend with other headlamps.

2.7 Conclusions

In this chapter a process for identifying tail-lamps and headlamps from automotive imagery using a standard colour camera has been presented. A camera configuration procedure, with specifications for camera exposure and colour configuration, is presented to ensure that there is minimal saturation of automotive lamps in captured imagery. It also allows for pixel values to be compared to device-independent colour boundaries for rear-lamps specified in regulations and ensures that the appearance of lamps are unaffected by ambient lighting conditions. Finally, this camera configuration allows for the reproduction and verification of results and portability between different camera platforms.
Rear-lamps are identified by using a red colour threshold derived directly from automotive regulations. Threshold boundaries are converted to a device-dependent colour space and adapted for real world conditions in the HSV colour space, where parameters can be adjusted intuitively. Headlamps are identified by a region growing technique where high intensity seeds are grown to an edge detected boundary. Results are presented demonstrating the ability of the technique in identifying tail-lamps and headlamps in on-road imagery. These results show lamps are identified effectively with minimal false positive, at a range of distances and in different ambient lighting conditions. The intensities of reflections are diminished by fitting the camera with a polarising filter, ensuring that the majority of reflections are not identified as vehicle lamps.

While this vehicle lamp identification process is effective, there will inevitably be false positives due to background light sources such as street lighting, building lighting and traffic signals. Also, to extract useful information from the location of these lamps, it is necessary to associate lamps with vehicles. In the next chapter a technique for pairing identified lamps using symmetry analysis is presented, which also filters out false positives from the lamp identification stage.
Figure 2.10: (a), (c), (e), (g), Low exposure night video frames, showing the rear tail- and brake-lamps of vehicles at various distances and in various lighting conditions and (b), (d), (f), (h), resulting binary images from the corresponding HSV colour threshold and morphological closing. Tail- and brake-lamps are consistently successfully identified while non-red light sources are excluded.
Figure 2.11: (a), (c), (e), (g) Low exposure night video frames, showing headlamps of vehicles at various distances and in various lighting conditions and (b), (d), (f), (h) resulting binary images from the corresponding region growing threshold and morphological closing. Headlamps are consistently successfully identified while most other light sources are usually excluded.
Figure 2.12: (a), (b) False positive tail-lamp and (c), (d) false positive headlamp identification, caused by background and infrastructure light sources.
Chapter 3

Visual Vehicle Detection

3.1 Introduction

Once vehicle lamp candidates have been identified, this information can be used to determine the location of vehicles in video data. This is not a straightforward process, as false positive lamp candidates must be filtered out and matching lamps must be correctly paired together to determine vehicle location. In this chapter, vehicles are detected by pairing identified vehicle lamps, thus associating each lamp pair with a target vehicle. A review of the literature in the area of vehicle detection is presented, with emphasis on visual cameras and dark conditions. According to regulations, vehicle headlamps, tail-lamps and brake-lamps (excluding horizontal brake bars) must be placed in symmetrical pairs towards the extremities of the vehicle width. Therefore, the search for symmetrical pairs of lamps is a very useful tool in the filtering of false positive vehicle lamps and in the association of lamps with road vehicles. Vehicles are detected by pairing of detected lamps based on symmetry, assessed by means of cross-correlation. An automatic image transformation for potential lamp pairs is introduced to compensate for perspective distortion and achieve a more accurate measure of true symmetry. Detected target vehicles are tracked between frames using the Kalman filter and multiple target situations are addressed. A flowchart outlining
the general structure of the proposed vehicle detection system is presented in Figure 3.1.

Figure 3.1: Flow chart outlining the proposed structure of the vehicle detection system.

3.2 State of the Art

There are numerous techniques commonly used to detect vehicles in night-time video. Physical or heuristics based assumptions have been used to aid the detection of vehicles in imagery. The average target vehicle is assumed to be approximately 170cm wide in [36] and the width/height aspect ratio of a highway vehicle to be approximately 2.0. In [34] the assumption is made that most rear-lamps are 70 to 90cm above the ground. Close vertical position [36] and similar area [34, 42] of lamp regions have been used as pairing criteria. In [33] the aspect ratio constraints of resulting bounding boxes are considered, since it can be assumed that the resulting
box will generally be relatively flat and wide. In [24] comparisons between horizontal and vertical coordinates of candidates are examined, coupled with a search for a license plate between them. Chen et al pair bright regions by grouping areas that are adjacent and aligned horizontally [50]. The authors validate candidate vehicles if they satisfy rules based on size and shape of lamps and estimate distance to the detected vehicle using the average vehicle width and known camera parameters. Lane detection is commonly employed to focus attention and prioritise an area of the frame for visual vehicle detection [34, 36, 64].

Symmetry is commonly used to filter vehicle candidates and create lamp pairs for vehicle detection [65, 66], because the front and rear of a vehicle is typically symmetrical under all lighting conditions. At night-time, symmetry is a useful mechanism for detecting vehicles. As stated in Chapter 2, automotive regulations require vehicle lamps to be placed symmetrically in pairs, at the extremities of the vehicle.

One previous approach to vehicle detection attempted fit an axis of symmetry to targets [67], which has proved effective for daylight situations where the scene can be considerably more complex than in night scenes. In [68] symmetry is calculated within a candidate bounding box and considered with several other features in a weighted fusion process. In [32] the symmetry axis is the vertical axis of the tracking window.

Urban traffic monitoring with a stationary camera in night conditions is presented in [69], where headlamp candidates are generated using rule-based morphological analysis of bright regions. A headlamp template is scaled and compared to vehicle candidates and correlation between potential headlamp pairs is calculated to assess bilateral symmetry. Lights belonging to the same vehicle should have high correlation values as they will have very similar size, shape and luminance values. Correlation is frequently used in computer vision for matching detected objects with a template, or searching an entire image for a match to a template.

The assumption that the target vehicle will appear symmetrical is based on the
assumption that the vehicle’s axis is parallel to the camera’s optical axis (defining a vehicle’s axis as line running from the back to the front of the vehicle through the centre) so that foreshortening is negligible [35]. However, this assumption is not valid in situations such as overtaking manoeuvres, lane changes or sharp bends, where the target vehicle’s axis is not parallel to the optical axis. In these situations the image of the target vehicle is subject to perspective distortion and the vehicle lamp pair do not appear symmetrical. Perspective distortion also occurs when the vehicle axis is not directly aligned with the camera’s optical axis, for example, during overtaking, or when encountering vehicles in a neighbouring lane. This phenomenon is considered in [70], where Inverse Perspective Mapping (IPM) is applied to the image followed by motion based detection. Symmetry of vehicle lamp pairs can also be affected when turn-signal-lamps or single rear-fog-lamps are engaged, or when other light sources appear in close proximity, such as lamps from other vehicles or street lamps. No known prior work addresses the issue of these frequently occurring distortions.

Temporal continuity of video data can be used to improve detection rates and improve robustness of vehicle detection systems by predicting the future location of a target based on the known current position and past trajectory. The Kalman filter [71, 72] is a well known approach for target tracking which uses a Gaussian distribution density to represent the target. It estimates the state of a dynamic system from a series of incomplete or noisy measurements and can continue tracking through short occlusions [73]. It uses the previously estimated state and the current measurement (if available) to estimate the current state of the system. The Kalman filter has been used to track pedestrians in night vision applications [74] and for multi-vehicle tracking [75]. A vehicle lamp’s trajectory has been used to distinguish it from static lights such as street lamps and reflective road signs [76, 77]. In [78] a mean-shift estimator is used for tracking vehicles during day time. Bayesian templates in conjunction with a Kalman filter are used in [79] for tracking of vehicles during daylight conditions. A particle filter is used in [68] to merge multiple target cues, including tail-lamps and
to track multiple targets. A Kalman filter is used to track oncoming headlamps up to 50m in grey scale video for automatic headlamp dimming in [19]. The trajectory of a lamp has been used to distinguish it from static lights such as street lamps and reflective road signs [76].

3.3 Symmetry Analysis

Bilateral symmetry is a useful cue for vehicle detection in all light conditions, including darkness. While the lamp identification algorithms presented in Chapter 2 are useful tools for locating automotive light sources, not all detected regions will be vehicle lamps. Symmetry also provides a mechanism to filter false positives from lamp identification.

3.3.1 Spatial Features

All potential pairs of regions are firstly subjected to a comparison of area (3.1) and assessment of the angle adjoining their centroids (3.2). These preconditions reduce the total number of symmetry analysis operations required while ensuring valid lamp pairs are examined.

\[
\frac{\min(\text{area}(J), \text{area}(K))}{\max(\text{area}(J), \text{area}(K))} > a_{\text{min}} \tag{3.1}
\]

\[
\left| \tan^{-1}\left( \frac{y_J - y_K}{x_J - x_K} \right) \right| < \theta_{\text{max}} \tag{3.2}
\]

where \( \theta_{\text{max}} \) is the maximum adjoining angle and \( a_{\text{min}} \) is the minimum area ratio of lamp segments \( J \) and \( K \) which have centroids \((x_J, y_J)\) and \((x_K, y_K)\) respectively. \( \theta \) is the angle of a line joining the centroids of the two lamp candidates relative to the horizontal (0\(^\circ\)). A lamp pair progresses to the bilateral symmetry analysis stage if it satisfies these conditions. To derive threshold values for \( \theta_{\text{max}} \) and \( a_{\text{min}} \), 300
vehicle images captured at night were analysed. Distributions of adjoining angles and area ratios were calculated from these images and were observed to be approximately normal. Gaussian curves were fit to the distributions and the thresholds were derived from the 84% probability point \([0, \mu_\theta + \sigma_\theta]\) and \([\mu_a - \sigma_a, 1]\). The distributions and the Gaussian curve fits are displayed in Figure 3.2 and statistical analyses of them, from which the thresholds are derived, are presented in Table 3.1.

![Figure 3.2](image)

(a) Adjoining angles distribution
(b) Area ratios distribution

Table 3.1: Statistical analysis of Gaussian curve fit of heuristic features from vehicle lamp image database

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(\mu)</th>
<th>(\sigma)</th>
<th>RMSE</th>
<th>Threshold value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle ((\theta))</td>
<td>0.2502°</td>
<td>1.606°</td>
<td>0.7324</td>
<td>(\theta_{max} = \mu_\theta + \sigma_\theta = 1.8562°)</td>
</tr>
<tr>
<td>Area Ratio ((a))</td>
<td>0.9054</td>
<td>0.3057</td>
<td>0.5641</td>
<td>(a_{min} = \mu_a - \sigma_a = 0.5977)</td>
</tr>
</tbody>
</table>

3.3.2 Cross-Correlation Symmetry Analysis

Lamp pairs satisfying the area and angle constraints are then subjected to cross-correlation symmetry analysis. Although there are no shape constraints for auto-
motive lamps in the regulations, headlamp, tail-lamp and brake-lamp pairs must be symmetrical. Image cross-correlation has been used previously in automotive vision systems for lamp detection with a stationary camera [64], as well as template matching of vehicles [32] and pedestrians [80].

Fast normalised cross-correlation [81] is used to measure the bilateral symmetry between lamp candidates. One of the lamp regions is mirrored horizontally and used as the template \( T \), this is then compared to the image of the potential matching lamp \( I \). The cross-correlation matrix \( \gamma \) is calculated between the two lamp image segments by:

\[
\gamma = \sum_{x,y} \frac{(T(x, y) - \overline{T})(I(x, y) - \overline{I})}{\sigma_T \sigma_I}
\]

(3.3)

Where \( \overline{T} \) and \( \overline{I} \) are the mean values of \( T \) and \( I \) respectively and \( \sigma_T \) and \( \sigma_I \) are the standard deviations of \( T \) and \( I \) respectively. To utilise colour information, correlation matrices are calculated for the \( R \), \( G \) and \( B \) channels and the mean is calculated. A lamp pair is confirmed as a valid vehicle if the maximum value in the cross-correlation matrix \( \gamma \) is greater than a threshold \( \gamma_{\text{min}} \). Values for \( \gamma_{\text{min}} \) for rear-lamps and headlamps were derived from the distributions of the correlation coefficients from the database of vehicle images. Gaussian curves were fit to the histogram data and \( \gamma_{\text{min}} \) thresholds were established at the lower 95.4\% probability points \( ([\mu_{\gamma} - 2\sigma_{\gamma}, 1]) \). The statistical analysis of these curves and pairing correlation thresholds are presented in Table 3.2, while the distributions themselves and corresponding Gaussian curve fits are shown in Figure 3.3.

Table 3.2: Statistical analysis of Gaussian curve fit of correlation values from vehicle lamp image database

<table>
<thead>
<tr>
<th></th>
<th>( \mu )</th>
<th>( \sigma )</th>
<th>RMSE</th>
<th>Threshold value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rear-Lamp</td>
<td>0.9381</td>
<td>0.0421</td>
<td>2.192</td>
<td>( \gamma_{\text{min}} = \mu - 2\sigma = 0.8538 )</td>
</tr>
<tr>
<td>Headlamp</td>
<td>0.9149</td>
<td>0.0451</td>
<td>0.7043</td>
<td>( \gamma_{\text{min}} = \mu - 2\sigma = 0.8247 )</td>
</tr>
</tbody>
</table>
This approach is a size and shape independent method of pairing detected lamps. There are many benefits to using cross-correlation to pair light candidates. This method is not directly dependent on the result of the lamp identification stage. The binary result of that is used only to locate image segments in the source image for correlation. Colour data from the source image is utilised in the cross-correlation symmetry check and a single numerical parameter indicating how well the regions are matched is produced.

### 3.4 Perspective Distortion Correction

Previous research has assumed the target vehicle is not rolled or yawed relative to the host vehicle and that the target vehicle’s axis is aligned with the camera’s optical axis [35]. Therefore, the image plane containing the lights will be parallel to the image sensor and perspective foreshortening of the target will be negligible. This is the case in many traffic scenarios and the vehicle lamps will appear symmetrical and the foreshortening in the bounding box is negligible.

Perspective foreshortening refers to the visual effect or optical illusion where an
object appears shorter than it actually is because of its position or angle relative to the viewer. A difference in yaw angle between target and host vehicle commonly occurs during many driving scenarios such as road bends, turns, interchanges and during overtaking and lane changing manoeuvres. In these situations, vehicle lamps frequently do not appear entirely symmetrical and the foreshortening in the bounding box is not negligible, as symmetry measurements are adversely affected. This assumption also fails to consider perspective distortion of target vehicles, caused by the targets vehicle’s position relative to the camera. The effects of this are especially prevalent at shorter distances, so perspective correction is especially important for close range applications such as monitoring of overtaking vehicles. Correction for perspective distortion is also important for tracking of a target vehicle while it navigates tight urban turns and bends.

Perspective correction is commonly used in automatic number plate recognition [82] to correct for perspective warped number plate images before performing character recognition. In the system proposed in this thesis, correction for perspective distortion of vehicle lamps is performed by conducting a projective transformation on the portion of the image containing each pair of potential vehicle lamps. A transformation matrix is created that transforms the bounding quadrilateral enclosing the light pair to a rectangular box. This transformation matrix is applied to the image of the lamp pair so that cross-correlation can be used to obtain a true measure of bilateral symmetry between them. This projective transformation is performed through a unit square and is presented in detail in Appendix A.

The size of the rectangle used in this process in inconsequential to the task of symmetry analysis. To obtain a true representation (i.e. correct aspect ratio, true shape) of the lights, a rectangle representative of the true size of the lamp pair would have to be used. However to assess the bilateral symmetry between lights, the size of the rectangle used is not important. This is demonstrated in Figure 3.4, where a synthetic image of a pair of identical circles is artificially warped. The
Chapter 3: Visual Vehicle Detection

perspective correction algorithm calculates the appropriate transformation matrix and transforms the image in the quadrilateral to a rectangle. It can be observed that, unlike the original shapes, the resulting shapes are not perfectly circular, but they remain symmetrical. This is verified by a correlation value of $max(\gamma) = 0.9863$ (near to the maximum of 1.0000). This automatic perspective correction allows for robust detection of a target vehicle even when it is distorted by perspective and doesn’t appear symmetrical in an image.

![Figure 3.4](image)

Figure 3.4: (a) A synthetic test image of two identical circles ($max(\gamma) = 0.9934$), (b) the image after an artificial perspective warping ($max(\gamma) = 0.7428$) and (c) the resultant image after automatic perspective correction ($max(\gamma) = 0.9863$). While the original circular shapes are not reconstructed perfectly, there is near perfect symmetry between the resulting regions. This is verified by the near maximum correlation value, similar to the original image.

3.5 Kalman Filter Tracking

Inter-frame tracking of detected targets is conducted for a number of reasons: to smooth noise in the detection process and noise due to vehicle movement, to extrapolate target features (size and position) during short periods where detection may fail and to predict future position of targets.

Once vehicles are detected, they are tracked using the Kalman filter [71]. The Kalman filter is a least-squares estimator of linear movement, often applied as a target tracking algorithm. The tracking system is structured as follows: Firstly, a
prediction stage estimates the current position of tracked targets. Secondly, after the vehicle detection process, a correction stage associates the detections in the current frame with the tracked targets. These detections are used to update the trackers using the following Kalman filter update equations (the notation for these equations has been taken from [72]). Predictions of the state vector \( \hat{x}^- \) and state error covariance matrix \( P^- \) are generated for a target at time \( k \):

\[
\hat{x}^-_k = A\hat{x}_{k-1}
\]  

\[
P^-_k = AP_{k-1}A^T + Q
\]

where \( A \) is the state transition matrix and \( Q \) is the process noise covariance matrix. A state transition matrix \( A \) relates the state at the previous time step \( (k - 1) \), to the state at the current step \( (k) \), in the absence of either a driving force or process noise. For this system, \( A \) is:

\[
A = \begin{pmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]
The result of this is the updated state vector:

$$\hat{x}_k = \begin{pmatrix}
\hat{x}_{k-1} + \Delta \hat{x}_{k-1} \\
\hat{y}_{k-1} + \Delta \hat{y}_{k-1} \\
\hat{w}_{k-1} + \Delta \hat{w}_{k-1} \\
\hat{h}_{k-1} + \Delta \hat{h}_{k-1} \\
\Delta \hat{x}_{k-1} \\
\Delta \hat{y}_{k-1} \\
\Delta \hat{w}_{k-1} \\
\Delta \hat{h}_{k-1}
\end{pmatrix}^T$$

(3.7)

These predictions are then used to associate detections in the current frame \((z)\) with targets being tracked. These system measurements \((z)\) are used to correct and update the corresponding trackers. The Kalman Gain \((K)\) is computed by:

$$K_k = P^{-1}_k H^T (HP_k H^T + R)^{-1}$$

(3.8)

where \(H\) is the matrix which relates the true state space with the measurement space and \(R\) is the measurement noise covariance matrix. The Kalman gain is then used to correct the previous estimate of state and error covariance:

$$\hat{x}_k = \hat{x}_k^{-} + K_k (z_k - H \hat{x}_k^{-})$$

(3.9)

$$P_k = (I - K_k H) P^{-1}_k$$

(3.10)

Targets are tracked using the four parameters of a bounding box surrounding the lamp pair (x-position, y-position, width and height), these form the measurement vector \(z\). New trackers are created for each new target that enters the camera’s field of view.

The Kalman filter measurement covariance parameter \((R)\) determines the sensitivity of the trackers response to updates. Higher values of measurement covariance
will mean less weighting on the current measurements and smoother movement, while lower values will mean heavier weighting on the current measurements and a more responsive tracker. If this parameter is too low the tracker can become unstable during occlusions and detection failures, so there is a trade-off when choosing a value for $R$.

Bumps and inclines in the road surface are a frequent challenge in automotive image processing systems. This is tackled by the tracking system defining a higher measurement covariance value for tracking the vertical coordinate of the target so that the tracker is less responsive in the vertical plane and target motion due to camera shake is dampened. A lower measurement covariance is set for tracking the horizontal coordinate and width of the target. This ensures the tracker is responsive laterally, which is important to ensure that vehicles are tracked effectively while turning, overtaking and changing lanes.

To add extra robustness to the system a “tracking-based-detection” stage is introduced. When a tracker fails to update, its predicted bounding box is examined for identified vehicle lamps. The cross-correlation between lamp candidates found in the predicted area and the image of lamp regions from the previous detection is calculated. If there is a high correlation (greater than 85%) for both of the lamps the tracker is updated with the new location. This concept has previously proven successful at tracking preceding vehicles, when correlation is performed on a larger portion of the image [83].

### 3.6 Multiple Vehicle Distinction

Concurrent detection and tracking of multiple vehicles has been commonly cited as a challenge for vehicle detection systems [34]. Vehicle detection at night is especially challenging as there may be little or no difference in appearance between lamps of different vehicles. The main problem encountered is in distinguishing between several
Chapter 3: Visual Vehicle Detection

potential pairs of lamps at similar distances. This scenario is usually presented in multi-lane urban or motorway environments.

Potential erroneous detections in multi-vehicle situations are addressed by considering overlapping candidate pairs - i.e. lamps that are paired successfully with more than one partner lamp. A lamp cannot be paired more than once, as it cannot belong to more than one vehicle. In this situation priority is given to trusted targets, vehicles that have already been detected and tracked for a number of frames. If there are no established trackers before a multi-pair conflict, the common lamp between overlapping boxes is paired with its more exact symmetry match. Figure 3.5 shows the result of successful detection of three vehicles in the same frame in a typical complex urban road situation. A fourth vehicle present in this image is not detected as one of its lamps is partially occluded by one of the vehicles in the foreground.

![Figure 3.5: Three oncoming vehicles at similar distances, successfully distinguished from each other.](image-url)
3.7 Results

3.7.1 Data Capture

Forward facing and rear facing video data were captured using cameras mounted inside the host vehicle\(^1\). The forward facing camera was mounted behind the rear view mirror and the rear facing camera was mounted inside the rear window. Figure 3.6 shows an image of the forward facing camera mounted behind the rear view mirror.

![Forward facing video capture configuration, camera mounted behind rear view mirror.](image)

This colour video was captured using a regular consumer video camera with CMOS sensor and Bayer RGB Filter. The camera was set up with the configuration proposed in Section 2.3, including a polarising filter. It should be noted that a Bayer RGB sensor effectively has one third the spatial resolution of a grey scale sensor, such as the one used in [19] and two thirds of the spatial resolution of the clear/red sensor used in [44]. Video was processed at a resolution of 720×576 pixels at 25Hz. The video camera has a horizontal field of view of 48°. This ensures that the system is not tuned to only detect distant vehicles, but can also detect vehicle at close

---

\(^1\)This data has been made available online at [http://www.eee.nuigalway.ie/Research/car/projects/night_time/index.html](http://www.eee.nuigalway.ie/Research/car/projects/night_time/index.html)
range through complex turning manoeuvres. Much of the video contains overtaking manoeuvres, corners and road bends of various radii where perspective distortion of lamps is prevalent. The majority of video was captured late at night, well after dusk and includes scenes on motorways, in urban areas and on unlit rural roads. Some of this video data was captured in wet weather conditions where road reflections were intense and visible to the eye. Over eight hours of on-road video were captured comprising of speeds ranging from 0 to 100 km/h.

Due to the novel low-exposure configuration, there are no appropriate publicly datasets available for benchmarking. As the preprocessing configuration is an essential part of the performance this systems, unique video data must be captured for performance analysis.

A custom Graphical User Interface (GUI) was developed for manual classification of video processing results and accumulation of performance statistics. This is similar to a system created by Bertozzi et al [84] for evaluating the performance of vision based pedestrian detection. A screenshot of the GUI is shown in Figure 3.7.

Figure 3.7: GUI developed to collate vehicle detection results. Data on detection before and after tracking, false positives and estimated distance is recorded for each frame.
3.7.2 Cross-Correlation Results

The cross-correlation vehicle detection process not only ensures that detected vehicle lamps are not mistakenly paired with false positive lamps, but also ensures that lamps with similar appearance from different vehicles can be distinguished. Figures 3.8 and 3.9 show example images and cross-correlation matrices from typical vehicle detection examples.

![Example Images](image)

(a) Original low exposure image. A vehicle is clearly visible by its tail-lamps. (b) Median filtered image of the left tail-lamp (c) Median filtered and horizontally mirrored image of the right tail-lamp. (d) Resulting cross-correlation matrix ($\gamma$). The maximum value is located in the centre (marked by ×) and is 0.97, indicating a 97% correlation and a valid vehicle lamp pair

Examples of a valid lamp pair, a pair of similarly sized non-identical lamps and their respective correlation values are presented in Figure 3.10. This indicates the ability of the algorithm to distinguish between similar lamps.
Figure 3.9: Result of headlamp cross-correlation pairing process. (a) Original low exposure image. A vehicle is clearly visible by its headlamps. (b) Enlarged image of the left headlamp (c) Enlarged and horizontally mirrored image of the right headlamp. (d) Resulting cross-correlation matrix $\gamma$. The maximum value is located in the centre (marked by $\times$) and is 0.9821, indicating a correlation value $max(\gamma)$ of 98.21% and a valid vehicle lamp pair.

### 3.7.3 Perspective Correction Results

The perspective correction process ensures that the high vehicle detection rate is maintained in situations where perspective distortion is prevalent, such as bends and turns in the road and overtaking manoeuvres. Several ISO standards for ADAS testing [85] require systems to function on roads with curvature down to a radius of 125m. However the perspective correction process enables detection and tracking to continue through road curves as low as 15m in radius in the proposed system. Figure
Figure 3.10: (a) A valid vehicle tail-lamp pair, with correlation $\max(\gamma) = 0.8874$, above the rear-lamp matching threshold. (b) Similarly sized tail-lamps from different vehicles, with correlation $\max(\gamma) = 0.5771$, below the rear-lamp matching threshold. This demonstrates the ability of the cross-correlation method to distinguish between similar lamps.

Figure 3.11 shows headlamp images from various stages of a target vehicle engaging a 35m radius road bend. The corresponding perspective corrected images are also displayed with correlation values for both. This shows that the perspective correction technique results in a higher symmetry measurement, meaning vehicle detection doesn’t fail during the manoeuvre.

Figure 3.12 displays a graph showing a comparison of correlation values for vehicle headlamps before and after perspective correction, throughout a video segment of a similar 35m radius road bend. Overall, there is an 8% average improvement in correlation after perspective correction. Post-perspective correction correlation stays above the threshold for positive vehicle detection ($\gamma_{\text{min}} = 0.8247$) throughout the segment while pre-perspective correction correlation is usually below the threshold. Successful detection of tail-lamps and headlamps under severe perspective distortion is displayed in Figure 3.13.

The perspective correction stage of the algorithm resulted in a marginal increase in false positives. This was not found to be significant as lamp pairs must still satisfy area and adjoining angle checks and must appear for consecutive frames to progress to tracker based filtering.
Figure 3.11: (a),(c),(e),(g) Images of headlamps and the correlation between them ($\max(\gamma)$) from a target vehicle through various stages of a 35m radius road bend and (b),(d),(f),(h) the perspective corrected versions of these images and the correlation between them. Perspective correction ensures a higher measure of symmetry throughout so vehicle detection doesn’t fail.
Chapter 3: Visual Vehicle Detection

Figure 3.12: A graph of the correlation before and after perspective correction between headlamps of a target vehicle during a video segment of a road bend of radius 35m. There is an 8% average improvement in correlation after perspective correction. Post-perspective correction correlation stays above the threshold for detection throughout the segment while pre-perspective correction correlation is predominantly below the threshold.

![Graph of correlation before and after perspective correction.](image)

Figure 3.13: (a) Tail-lamp and (b) headlamp vehicle detection under extreme perspective distortion.

![Images of tail-lamp and headlamp detection.](image)

3.7.4 Video Processing Results

3.7.4.1 Rear-Lamp Detection Results

Video processing results are presented categorised into three environment sets: urban, rural and motorway, with each set containing 30 video segments of 20 seconds
in duration. The rear-lamps of at least one vehicle are visible for the duration of the video segments. Video segments categorised as urban were captured in inner city and suburban environments. Video segments in the rural category were captured on unlit country roads, while motorway environment video segments were recorded on lit and unlit multi-lane dual-carriageways and motorways.

Table 3.3 presents the detection rates before and after tracking and the false positive rates for each environment type.

Table 3.3: Summary of rear-lamp vehicle detection results - by environment type

<table>
<thead>
<tr>
<th>Environment</th>
<th>Pre-Tracking Detection Rate (%)</th>
<th>Post-Tracking Detection Rate (%)</th>
<th>False Positive Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>93.2880</td>
<td>97.4177</td>
<td>4.6315</td>
</tr>
<tr>
<td>Rural</td>
<td>93.1769</td>
<td>96.8653</td>
<td>1.3112</td>
</tr>
<tr>
<td>Motorway</td>
<td>92.4720</td>
<td>97.5105</td>
<td>1.8207</td>
</tr>
<tr>
<td>Mean</td>
<td>92.9790</td>
<td>97.2645</td>
<td>2.5878</td>
</tr>
</tbody>
</table>

Detection rate is defined here as the ratio of the number of frames with successful detection and pairing of valid vehicle lamp pairs, to the total number of frames where a valid vehicle lamp pair is present. Detection rate after tracking is the ratio of the number of frames where the tracker successfully updates with new detection measurements, or tracking based detection measurements, to the total number of frames where a valid vehicle lamp pair is present. Finally the false positive rate is the total number of false detections in proportion to the total number of frames. A false positive detection is when the system detects a vehicle where there is none, or where the system incorrectly pairs headlamps from different vehicles. Several frames from video results showing examples of successful rear-lamp detection are presented in Figure 3.14.

It can be deduced from these results that detection performance does not vary widely between environments. It was observed that slightly higher post-tracking detection rates in urban and motorway environments may be partly attributed to
the generally consistent road surface which reduces camera shake, simplifying the tracking process.

As the appearance of a target vehicle is highly dependent on the distance it is from the camera, the distance is obviously an important factor to consider in analysing the performance of the system. The regulations note that one of the functions of vehicle lamps is to indicate location and width of a vehicle to other drivers and therefore must be located towards the extremities of a vehicle, as far apart as practically possible. Thus the width of the target vehicle in an image (in pixels) can be obtained from the vehicle lamp detection and tracking results. Using this information, combined with camera parameters and an estimation of average real world vehicle width, the distances to detected vehicles are estimated using the perspective projection equation for a pin-hole camera model. This model has been used for similar applications in [19] and [32] and is described in detail in Appendix B. Table 3.4 presents the system detection rate after tracking, classified by distance. This system has a high detection rate for vehicles up to 50m away, while vehicles are sometimes detected between 50m and 80m. These results are comparable to the state of the art grey scale results [19]. The highest detection rate was achieved for vehicles in the middle distance (15-30m) as the fixed exposure ensures optimal appearance at these distances. Vehicles at shorter distances can sometimes experience saturation while vehicles at larger distances can sometimes appear too faint to be identified.

Table 3.4: Summary of rear-lamp vehicle detection results - by distance

<table>
<thead>
<tr>
<th>Distance Range</th>
<th>Detection Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-15m</td>
<td>96.5827</td>
</tr>
<tr>
<td>15-30m</td>
<td>98.5067</td>
</tr>
<tr>
<td>30-50m</td>
<td>94.9645</td>
</tr>
</tbody>
</table>

To evaluate the sensor independence of the system, static test images were captured at a range of distances using three different cameras\(^2\). Each camera was con-

\(^2\)Sony Handycam CX6EK, Fuji Finepix S5600 and Cannon PowerShot SD400
figured according to Section 2.3 and images were all scaled to the same resolution. While it is not appropriate to directly compare colour threshold results from different cameras, they can be compared in terms of detection results for the same test scenarios. Test images of the rear of a stationary vehicle were taken with each camera at 10m intervals of distances from 10m up to and including 80m. The results of these tests are presented in Table 3.5. The vehicle was successfully detected at each distance by each camera images up to 50m demonstrating comparable performance between different cameras. There were slight differences in results between cameras at 70m, at the limits of when the vehicle is first detected.

Table 3.5: Rear-lamp detection performance comparison of different camera’s

<table>
<thead>
<tr>
<th>Camera</th>
<th>10m</th>
<th>20m</th>
<th>30m</th>
<th>40m</th>
<th>50m</th>
<th>60m</th>
<th>70m</th>
<th>80m</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>2</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>3</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

●=detected, ○=not detected

Transitions from tail-lamp to brake-lamp and vice versa, are handled flawlessly and neither detection nor tracking are interrupted. Numerous target vehicles in the test video were recent models and had LED tail-lamps. Neither their high-frequency pulsed luminance nor contemporary shapes caused problems. They appear with consistent intensity in each frame and are detected with the same accuracy as regular incandescent bulb lamps.

3.7.4.2 Headlamp Detection Results

Over 2 hours of on-road video data was segmented into 44 video clips, each containing at least one vehicle for the duration of the clip. Many of the clips feature multiple vehicles in multi-lane scenarios. Table 3.6 outlines the results of the vehicle front detection algorithm from this video data. Detection rate was calculated as the number of frames in which a target vehicle was detected and tracked out of the
number of frames in which a vehicle was present. The false positive rate is the total number of false detections as a proportion of the total number of frames.

Table 3.6: Headlamp vehicle detection results summary

<table>
<thead>
<tr>
<th>Total Duration (Frames)</th>
<th>Detection Rate</th>
<th>Post-Tracking Detection Rate (%)</th>
<th>False Positive Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30876</td>
<td>85.3997</td>
<td>92.8631</td>
<td>1.0372</td>
</tr>
</tbody>
</table>

An average detection rate of 92.8631% is achieved over the data set. Headlamps were identified and paired effectively at distances of $3 \leq 80m$. The system displays immunity to lighting variations due to the static exposure configuration, so transition from lit to unlit areas and vice versa has no effect on the image and therefore detection performance.

Figure 3.15 displays a selection of frames from the results videos, showing headlamp vehicle detection at different distances and in different lighting conditions, while Figure 3.16 displays a sequence of images from a video segment showing the detection and tracking of an overtaking vehicle approaching from the rear.

### 3.7.5 Tracking and Tracking Based Detection Results

The tracking algorithm enables recovery from short failures in detection. This can be caused by the occlusion or distortion of the shape, or symmetry of the lamps by other objects or light sources.

The tracking algorithm increases the detection rate for two reasons. Firstly failed detections are frequently only for a small number of frames. In these situations the tracking algorithm prediction is used until the target is re-acquired. Secondly, tracking based detection has demonstrated robustness to distortions caused by other light sources. When non-headlamp or non-tail-lamp light source appears near to a vehicle lamp in an image, the shape and colour of the vehicle lamp can change, distorting the
regular appearance of the lamp. Figure 3.17 shows an example of successful tracking based detection, where a turn signal (indicator) lamp partly distorts one of the target vehicle’s rear-lamps. The normal vehicle detection process fails as the correlation between lamps is too low to detect the target as a vehicle. As the vehicle is being tracked and one of the lamps has maintained its appearance, tracking based detection successfully detects the vehicle.

The tracking based detection algorithm did not result in any measurable increase in false positives. This can be attributed to tracker based filtering, as a target must be detected and tracked for a time before being considered for tracking based detection.

### 3.7.6 False Positives and Detection Failures

The highest occurrence of false positives was from video data recorded in an urban environment. False positives were most commonly caused by background light sources such as street lamps and other road vehicle lamps. The majority of false positive detections appear for only single frames and are easily filtered. An example of a false positive detection is presented in Figure 3.18. The rural environment presented the fewest number of false positives which can be attributed to the relatively few background light sources compared to the other environments.

Vehicle detection results demonstrate low false positive rates of 2.5878% and 1.0372% for rear-lamps and headlamps respectively. During initial testing it was found that street lamps can sometimes be falsely identified as headlamps and if symmetrical in the image, can be mistaken for a vehicle. Masking the top portion of captured frames was effective at removing the majority of these street lamp false positives while not interfering with vehicle detection, as target vehicles never appear in the top portion of a frame. While previous research in the area masked the top two-thirds of the image [50], that was for a much shallower camera mounting angle. In this research masking the top one-third of the image was found to consistently remove most street lamps, while never masking vehicle lamps.
A control data set containing no vehicles was processed; this produced a control false positive rate of 0.3%. Failure to detect target vehicles was most commonly caused by failure to identify the vehicle lamps or occlusions of lamps. Detection also fails when the target vehicle does not meet legal requirements, e.g. they have a blown headlamp or rear-lamp bulb. In situations where one vehicle occludes another, the target nearest the host is detected, as its lamps tend to remain visible.

While this system was tested in dark night environments, it was not affected by ambient lighting conditions thanks to the low static exposure configuration. In accordance with this performance, it is envisaged that this system can perform effectively through less dark environments, as long as target vehicle lamps are turned on. In an embedded system, an algorithm such as the one presented in [86] could be used to handle the transition from day to night time, by monitoring the histograms of image intensities.

To assess if this system could function as part of a collision warning system in high speed situations, a worst case scenario was envisaged: two vehicles approaching each other at 120km/h. This 240km/h relative speed translates to 66.67m/s. At 80m detection range (headlamp system), that is 1.2 seconds between detection and impact. In that 1.2 seconds, the target vehicle will be detected in 36 frames. Once a vehicle is detected in more that two frames, the difference in size of the target can be used to establish an approximation of relative speed and hence a time to collision. As the 36 frames in this scenario is substantially more than the 2 frames required, it can be concluded that this system is more than capable of functioning in high speed situations.

3.8 Conclusions

In this chapter a video processing system for detection of vehicles at night has been presented. The task of associating identified vehicle lamps with a vehicle is not
a straightforward one, especially in environments where background light sources are abundant and in multiple-lane environments where vehicle lamps that are similar in appearance frequently appear in close proximity.

Identified vehicle lamps are paired by assessment of symmetry between each potential pair using cross-correlation. This filters out false positive lamps as well as providing useful information such as location and distance to vehicles. A projective image transformation corrects for perspective distortion of the target lamps, ensuring robust vehicle detection performance through tight road bends and turns. To increase detection performance, especially in multi-lane environments, detected vehicles are tracked between frames using the Kalman filter.

Vehicle rear-lamps are detected with more accuracy than headlamps but there are less false positives produced by headlamp identification than rear-lamp identification. These results confirm that this vehicle detection system satisfies each of the requirements of an automotive safety application, as set out in Section 1.1. Robust detection of targets is achieved, with a low occurrence of false positives. An approximation of distance to the detected vehicles is calculated and the maximum detection distance of the system is comparable to the state of the art.

In the next chapter, the focus moves from detection of other road vehicles to detection of the most common VRU: pedestrians. A system for detecting pedestrians in far-infrared video is presented. The system compensates for clothing-based distortion and detects pedestrian candidates using a region growing technique. Pedestrians are classified using machine learning and tracked using the Kalman filter.
Figure 3.14: Successful rear-lamp vehicle detection result frames, representative of results from the video data set. These vehicles are detected at distances of (a) 4m, (b) 15m, (c) 26m, (d) 30m, (e) 53m, (f) 83m. Detection is indicated by a rectangular bounding box around the lamps and an enlarged view of the detected vehicle is included in (b) - (f).
Figure 3.15: Successful headlamp vehicle detection result frames, representative of frames from the video data set. Detection of a target is highlighted by the bounding quadrilateral around the headlamps.
Figure 3.16: A sequence of images from a video segment showing the detection and tracking of an approaching, overtaking vehicle. Headlamp detection is highlighted by bounding quadrilateral around the headlamps (magenta) and status of the tracker is highlighted by the bounding box (cyan).
Chapter 3: Visual Vehicle Detection

Figure 3.17: Results of successful tracking based detection when a vehicle lamp is distorted because of a turn signal (indicator) lamp.

(a)  
(b)

Figure 3.18: False positive detection of vehicles, caused by (a) incorrect matching of background urban light sources and (b) incorrect matching between lamps from different vehicles.
Chapter 4

Infrared Pedestrian Detection

4.1 Introduction

This chapter moves from detection of other road vehicles to focus on detection of the most vulnerable road users: pedestrians [23]. At night pedestrians are particularly vulnerable as they are only visible by the light that they reflect. For the same reason, pedestrians can be extremely difficult to identify in night-time video, even for a human observer. While pedestrians do not emit light as road vehicles do at night, they do emit heat in the form thermal infrared radiation. In video captured with a Far-Infrared (FIR) camera, pedestrians typically appear bright and distinct from the background scene. This greatly improves the probability of detecting pedestrians automatically using image processing.

In this chapter a night-time pedestrian detection system based on automotive infrared video processing is presented. Far-infrared or thermal night vision is a technology well suited for automatic detection of pedestrians at night as they generally appear brighter than the background. A pre-processing step is presented, which uses morphological operations to compensate for clothing-based distortion of pedestrians. Pedestrian candidates are segmented using a region growing technique and classified using machine learning. Targets are tracked between video frames using a Kalman...
filter, adding robustness and increasing the performance of the system.

4.1.1 Far-Infrared Technology

Infrared radiation was discovered accidentally by Sir William Herschel in 1800. He found that when holding a thermometer outside the red band of light refracted by a prism, the temperature rose. This portion of the spectrum went on to be called the infrared wavelengths. Technology to form pictures from heat sources, called thermographs, was advanced around 1840, followed by the invention of the bolometer in 1880. This was composed of two blackened strips of platinum forming two branches of a Wheatstone Bridge. Such a device was said to have been able to detect the heat from a cow at 400m. The development of IR imaging technology accelerated rapidly during the military research of World War I and remained veiled in secrecy until the 1950’s. At this point it became feasible for thermal-imaging devices to be used in industry.

Infrared is the part of the electromagnetic spectrum from 0.7-300µm, between visible light and RADAR. Any object which has a temperature radiates heat in the infrared spectrum. Far-Infrared (FIR) or Long-Wavelength Infrared (LWIR) is the thermal imaging part of this spectrum and resides at wavelengths of 7-14µm. Microbolometers are modern uncooled solid-state thermal image sensors. A pixel matrix made of infrared absorbing material sits on a silicon substrate. Readout circuitry measures a change in resistance to form the pixel reading. As the fabrication process for this technology advances, thermal imaging systems are becoming more readily available and lower in cost, making them viable as automotive safety devices. Thermal night vision systems have been included in several high-end automobiles such as the BMW 7-series, Cadillac Deville, Audi A8 and the Honda Legend.

As thermal radiation from humans peaks in the far-infrared spectral band, FIR is an ideal technology for displaying pedestrians, particularly at night. It is also advantageous that the technology to sense and form an image is completely passive,
meaning no illumination is required, as is the case with near-IR and there is no danger of interference with other systems, as can be the case with RADAR. A comparison of a pedestrian captured at 35m with dipped headlamps by visible and IR cameras is shown in Figure 4.1. It can be observed that while the pedestrian is hardly visible in the visible spectrum, they are clearly visible and seen with ample contrast from the background, in the far-infrared spectrum. The visual image was captured in an unlit environment with automatic exposure and is representative of what a driver would see in this situation.

![Comparison of a pedestrian captured at 35m with dipped headlamps by visible and IR cameras.](image)

Figure 4.1: Examples images of a pedestrian at 35m with dipped headlights with minimal ambient lighting in (a) the visible spectrum and (b) far-infrared.

### 4.1.2 System Overview

A flowchart outlining the structure of the infrared pedestrian detection system described in this chapter is presented in Figure 4.2.
4.2 State of the Art

Previous work in the area of night-time pedestrian detection is described here, with a particular emphasis on night conditions and far-IR thermal night vision technology. Due to fundamental differences between visual and thermal imagery, many of the conventional techniques for pedestrian detection in visible spectrum images are not directly applicable [87, 88]. The first step in most systems is to determine Regions Of Interest (ROIs) in the image, which are examined more closely and classified as “pedestrian” or “non-pedestrian”. It has been established that pedestrians are usually warmer and hence appear brighter than the background [89]. Therefore, image thresholding is a common starting point for extracting pedestrian candidates.

The technique outlined in [90] defines a bright pixel threshold as the difference
between maximum image intensity and a set constant. In [91] a static threshold is derived by performing Bayes Classification on a set of templates known to contain pedestrians. A region growing style threshold is implemented in [92] using two static threshold values. The lower of the thresholds is restricted to areas spatially connected to seed regions resulting from the higher threshold. In [28] region growing is conducted until edge pixels are reached to extract vehicular targets from thermal night vision video. A threshold value is defined from the mean and maximum image intensity values in [74], while in [8] a threshold value is chosen as the last local minimum of the image histogram before the saturation point. These types of global image threshold that adapt to the properties of the current frame fail to address the potential differences in appearance between multiple pedestrians in the same frame. A thresholding technique in this area should extract each pedestrian candidate individually from the background.

While it has been observed that pedestrians appear brighter than the background, segmentation based on thresholding is not a trivial task. A threshold value that is too low may extract background regions as well as the candidate pedestrian, distorting the candidate’s shape and intensity characteristics. On the other hand, a threshold that is too high may fragment the candidate into numerous sub-regions, creating a new task of attempting to group these regions to form potential candidates, using such techniques as active contours [93, 94]. Xu et al. label such fragmented candidates “body-ground candidates”. In fact, this phenomenon is particularly common, when heavy clothing insulates body heat, especially the torso. The result of this is that it can be difficult to determine an optimum threshold to extract the pedestrian and in some cases it may be impossible to extract a pedestrian by thresholding as the image intensity of well insulated clothing nears the intensity of background objects.

The number of non-pedestrian ROIs generated from the thresholding step can be reduced by filtering the regions identified, based on the expected physical features of a pedestrian. Objects can be filtered according to aspect ratio [74] since pedestrians
are generally expected to be taller than they are wide. In [90] the concept of “inertial”
is introduced to identify pedestrians. Inertial is the total rotational momentum of a
candidate with respect to its centre. This value is generally static for different images
of pedestrian and varies for non-pedestrian objects.

Horizontal and vertical grey level intensity profiles have been analysed to form a
shape-independent approach composed of two steps [90]. The method first segments
the image into vertical strips by finding the local minima of a thresholded horizontal
image intensity profile. These strips are then further segmented vertically. As humans
generally display a profile with bilateral symmetry, grey level symmetry and edge
symmetry can also be used to segment ROIs as demonstrated in [95]. Edge density can
also be analysed as pedestrians frequently display a sharp change in image intensity
at their edges [95]. The gradient operator can also be used to aid detection [8, 96].

Stereo sensor configurations have also been proposed for infrared pedestrian detec-
tion [88, 92, 96, 97, 98, 99] and there are many advantages to utilising stereo sensors
for this task. Stereo systems have demonstrated robust detection, as disparity can be
used to effectively find ROIs. Multiple thermal imaging cameras however are not a
viable option for many automotive systems; as cost, power consumption and physical
space are significant factors.

Road detection can be used to create a focus of attention for filtering ROIs in an
image. The road surface can be found by convolution of a wide flat morphological
kernel in an edge image, as in [74]. This method is based on the assumption that
the road has a relatively constant temperature and so produces no edges in an edge
detected image. The authors in [74] exclude ROI candidates from outside the centre
half of the image, as these candidates are not within reasonable distance of the path
of the vehicle and use the detected road area to estimate the height of the candidates.
A similar approach is taken in [80].

Head detection is performed in [100] using the P-tile method. This method is based
on the assumption that the desired object occupies a set proportion of the image. The
threshold value is automatically varied until the desired fraction of image pixels are above this value. This is effective in the particular case examined in [100], namely a stationary camera. It has been stated that head detection can be a useful tool in the detection of groups [101]. Models and templates of pedestrians are frequently matched against images [80, 102] and a probabilistic template is created in [91]. This template matching is commonly performed using cross-correlation. Background subtraction is a very useful tool for segmenting ROIs from stationary thermal imagery [103], but it is not useful for a fast moving environment presented in automotive applications.

Night vision systems have become an important safety feature in some road vehicles. Commercial automotive IR pedestrian detection systems are starting to appear on the market, e.g. the system described in [89]. This system uses several features to distinguish relevant (pedestrian) objects from non-relevant hot or warm objects. A motion based cue is primarily used. A hot or warm object that changes shape at a rate above a certain threshold is considered relevant. This is combined with size and shape assumptions, template matching and image profiles. This approach relies on the assumption that pedestrians appear significantly warmer than the background. While it is acknowledged that clothing can distort the thermal signature of a pedestrian, no special consideration is given to this problem.

Classification is the process by which ROIs are placed into groups, in this case pedestrian or non-pedestrian. This decision is based on characteristics (features) determined from a training set of manually labelled images. This is conventionally achieved by means of a classification method such as Support Vector Machines (SVM) [74, 96], Artificial Neural Networks (ANN) [104] or Boosting [105]. It was found in [74] that grey scale classification was more successful than binary classification, as the binary candidates were too sensitive to shape. Experimental results showed that a single classifier for all types of pedestrians (along road, across road) performed better than the application of multiple classifiers. Other types of classifiers are explored in [101].
The choice of feature for classification of pedestrians is crucial to system performance. Histogram of Oriented Gradients (HOGs) features have been utilised for pedestrian classification, stereo hybrid visual-IR surveillance [106], automotive stereo thermal imagery [96, 98] and for road sign detection [107]. HOGs are used most effectively with a SVM classifier. Alternatively, the grey scale and binary images themselves are used as feature vectors in [74], whereas in [90], histogram, inertial and contrast features are explored. The wavelet transform is used to extract features from stationary camera thermal imagery in [108].

The appearances of pedestrians change with their distance from the camera. Multi-resolution image processing is used to enable the same algorithm to detect pedestrians at different distances [95]. The detection algorithm is initially tuned to detect pedestrians further away. The source image is then down sampled and the algorithm is reapplied to look for pedestrians closer to the camera. A technique tuned for the detection of distant pedestrians is presented in [80].

Tracking of pedestrians between frames is common and makes detection more robust as the future location of a target can be predicted, focusing the area for detection or extrapolating position when detection temporarily failure. The Kalman filter is a common tool for object tracking in pedestrian detection [74, 109] and computer vision in general [73]. Tracking by shape has been done for IR video with a stationary camera [110] but is difficult in the automotive domain as target objects move towards the camera and can change shape rapidly. An alpha-beta ($\alpha - \beta$) tracker is used to estimate object state parameters in [104], while a particle filter has also been used. In [74] a Kalman filter is supplemented with a mean-shift tracker to find the precise position of the pedestrian from the estimated position. The authors reported a dramatic improvement in performance after implementing tracking. The achieved detection rate went from 0.35 to 0.94 once tracking was introduced.
4.3 Region Of Interest (ROI) Generation

Region of Interest (ROI) generation (also known as hypothesis generation) is the process of traversing the image or video frame and highlighting areas that may contain the image of a pedestrian on which further processing should be focused. The ROI segmentation challenge is approached in many different ways in the literature, none of which specifically deal with the challenge of heavily insulating clothing. It is also apparent there is a need for a segmentation technique that is not just adaptive to the frame, but to different candidates within the frame.

The ROI generation process is a crucial step in this detection system. While the a classifier could be trained and tested using high numbers of arbitrary image segments, this is undesirable for two reasons. Greatly reducing the number of images that require classification has a huge computational performance advantage, as the classification step will be several factors faster. Secondly, by performing high level filtering and removing obvious non-pedestrians before classification, the burden on the classifier during training to find a separation between pedestrian and non-pedestrian classes is reduced, this should yield performance advantages.

4.3.1 Morphology Based Clothing Compensation

It is widely stated that pedestrians usually appear amongst the warmest and therefore brightest objects in thermal imagery [74, 87, 89, 90, 100, 111] and there is usually significant contrast between pedestrians and background. The typical range of human skin temperature values is relatively narrow; on average it varies little from 33°C because of thermoregulation. In most situations this is well above nighttime background temperatures. However, clothing interferes with a person’s thermal appearance and they may not always appear as a continuous bright region in an image. Bertozzi et al. have stated, with regards to pedestrian detection in thermal imagery, that clothing greatly affects the thermal footprint of a person, making the
human shape detection challenging [97]. Winter clothing in particular, is designed to be insulating and trap body heat. This causes inconsistent texture in the appearance of the pedestrian. The appearance of a pedestrian in thermal imagery is strongly dependent on the clothing, the duration of time the person has been outdoors [101, 112], as well as the ambient temperature. Clothed parts of the body will appear darker than exposed parts such as a person’s head and hands. A person’s head and hands and legs will appear brightest in the winter time [74], while the torso tends to be the darkest part of a person’s appearance in thermal imagery. Some pedestrians will appear bright in their entirety, especially if not wearing heavy clothing, or if they have just left a warm environment. So techniques that compensate for clothing distortion must not interfere with the appearance of these types of pedestrians.

It has been found in previous work that thresholding can fragment pedestrians into several hotspots, making the task of pedestrian detection more challenging. In cold weather situations, it is assumed in [80] that small, vertically aligned hotspots correspond to head and leg regions of distant pedestrians and so are merged. A similar approach is put forward in [74], where ROIs resulting from this procedure are labelled as “body-ground candidates”. While this approach may work for distant pedestrians, the assumption of vertical alignment does not always hold if the pedestrian is closer, as more detail is present and the person’s pose may affect calculation of a region’s characteristics. Heads, legs and torsos are found separately in [100] using three different thresholds and spatially adjacent hot spots are assumed to belong to the same ROI, while a morphological gradient is used to find heads in [113].

To compensate for this clothing-based distortion, a morphological operation is performed on the grey scale image to increase the intensity between the upper and lower parts of the body which may appear separated by a well insulated torso. The binary closing operation is commonly used to connect nearby large regions, while approximately preserving the original area. The grey scale equivalent, as used here for IR images, is a natural extension to the binary case, using the maximum and
minimum operators [62, 114]. Morphological closing attenuates dark artefacts or noise from an image, while leaving bright detail relatively undisturbed. As with the binary image operator, grey scale closing is a dilation followed by an erosion of an input image \( f \) by a structuring element \( b \):

\[
f \circ b = (f \oplus b) \ominus b
\]  
(4.1)

A structuring element is the kernel that convolves the image in morphological operations. The grey scale dilation at a point \((x_0, y_0)\), takes the maximum value of image under the structuring element:

\[
(f \oplus b)(x_0, y_0) = \max \{f(x_0 - x, y_0 - y) + b(x, y) | (x, y) \in D_b; (x_0 - x, y_0 - y) \in D_f\}
\]  
(4.2)

where \(D_f\) and \(D_b\) are the domains of \(f\) and \(b\) respectively. The dual of this operation, grey scale erosion, is then performed, where the value at a point is the minimum value under the structuring element:

\[
(f \ominus b)(x_0, y_0) = \min \{f(x_0 + x, y_0 + y) - b(x, y) | (x_0 + x, y_0 + y) \in D_b; (x, y) \in D_f\}
\]  
(4.3)

This closing operation does not affect pedestrians that appear entirely bright and their appearance is maintained. Figure 4.3 (a) displays an infrared image and 3D intensity plot of two pedestrians situated closely together. The intensity of torso region of both pedestrians is considerably lower than the head and legs and is hence much closer to the background intensity level. The closing operation increases the intensity between bright regions under the structuring element, so the size and shape of the structuring element is an important factor in this process. Figure 4.3 (b) shows the result of closing the image with a standard disc shaped structuring element. While the intensity level of the torso region has been increased, the separation between the two pedestrians has been lost due to their close proximity.

To overcome this problem the structuring element \(b\), is a flat vertical rectangular
shape. Closing the image with this structuring element also increases the intensity of torso regions, while maintaining the subtle distinction between the two pedestrians. The rectangular kernel ensures that torso intensity is increased, without joining the pedestrian to other horizontally adjacent high intensity objects, such as other pedestrians. The final result of the clothing compensation using this vertical rectangular grey scale morphological closing is presented in Figure 4.3 (c).

To ensure this process works effectively for pedestrians at different distances, two sizes of structuring element are used to create several closed images. A larger kernel (13 × 30 pixels) is effective on closer pedestrians while a smaller one (3 × 13 pixels) is more effective on targets further away (on a 320 × 240 pixel image). After this operation, dark pedestrian torso areas will be of higher intensity than the background, greatly aiding the segmentation of whole pedestrians.

### 4.3.2 Feature Based Region Growing

After applying morphological compensation for clothing distortion, there will be significant contrast between the entire shape of the pedestrian and the background. This enables effective segmentation using a thresholding based approach, while ensuring the pedestrian candidate remains as a single connected region. An appropriate threshold must still be determined. Numerous techniques for infrared pedestrian detection begin with some form of high intensity thresholding. Some approaches define a static threshold at the outset to process all frames, whereas others define a threshold value that varies dynamically over time to take into account the current scene. As outlined, the appearance of a pedestrian in a frame is not just dependent on global properties such as the ambient temperature, but on properties unique to the individual such as clothing characteristics and the amount of time spent outdoors. Therefore, different pedestrians will have different characteristics, even in the same frame. To extract ROIs a seeded region growing [61] technique that adapts to these differences is proposed.
Figure 4.3: IR images of two closely positioned pedestrians whose torsos are insulated by clothing and corresponding 3D intensity plots. (a) The original image and (b) corresponding 3D intensity plot, (c) image closed with a disc kernel and (d) corresponding 3D intensity plot, (e) image closed with vertical rectangle kernel and (f) corresponding 3D intensity plot. Closing with a vertical rectangle increases torso intensity, aiding segmentation, without joining the pedestrians, or distorting their shape.
A region growing technique is defined by the method used to choose seeds, the inclusion criteria for neighbouring pixels and the stop criteria, to determine when the region has finished growing [62]. Region growing has previously been used to segment pedestrian ROIs in thermal imagery [115], where suitable seeds were found by applying a high static threshold and the stop conditions in the region growth are determined by a second lower static threshold. This is very effective at finding seeds inside pedestrians, working in the majority of cases, but the lower static threshold to determine a stopping point for region growth is not adaptive. Stereo IR cameras were also used to focus detection on specific areas using a disparity map [115].

In the system presented here, seed points are located using a high threshold, close to the maximum intensity value of the image. Connected pixels in the resulting binary image are labelled and a single pixel is taken from each of these regions to use as a seed point. Of course, seeds are also found in warm background objects such as street lights, vehicle lamps and other hot vehicle parts such as exhaust pipes. An example of this seed selection method is displayed in Figure 4.4. This indicates that not only are seed regions derived from a pedestrian, but also from surrounding warm objects.

![Figure 4.4: IR image of (a) a pedestrian and (b) seed regions for region growing, produced by a high threshold, near the maximum image intensity.](image)

Incremental region growing is then applied from the seed points, lowering the intensity threshold for connected pixel inclusion in each iteration. As a seed region
grows it merges with lower intensity parts of the image. The criteria for stopping region growth come from the analysis of two shape-based features from the growing region. Aspect ratio is the ratio of the width to the height of the region:

\[ Aspect \ Ratio = \frac{w}{h} \quad (4.4) \]

where \( w \) is the width of the region and \( h \) is the region height. This is a distinguishing feature of pedestrians, who will predominantly appear taller than they are wide (assuming they are not sitting down, crouched or fallen). The second feature is extent, which is a measure of the proportion of the region’s enclosing bounding box it fills:

\[ Extent = \frac{a}{wh} \quad (4.5) \]

where \( a \) is the area of the region in pixels. This is also a distinguishing feature of pedestrians, as the shape of a person is usually relatively compact compared to many background objects and a person’s shape generally fill a high proportion of its bounding box. Such features have been used in the classification of vehicles in video [116]. The limits of each feature that describe pedestrian appearance, were derived from analysis of a database of pedestrian images, for which the seeds were manually grown. These limits are presented in Table 4.1. At the point where the growing pedestrian region begins to merge with the lower intensity background, it typically breaks the aspect ratio or extent threshold. When this transition occurs, the region before the cross-over point is recorded as a ROI.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect Ratio</td>
<td>0.20-0.49</td>
</tr>
<tr>
<td>Extent</td>
<td>0.52-0.93</td>
</tr>
</tbody>
</table>

An example of the various stages of this region growing process is presented in Figure 4.5. The original image is thresholded for seed regions, morphologically closed,
then grown until it breaks the aspect ratio or extent feature boundaries. The resulting ROIs from a complex urban scene are highlighted in Figure 4.6. This dynamic region growing is more adaptable to different scenes and different pedestrian appearances than the static region growing thresholds proposed in [92] and works independently for individual pedestrians in each frame. A check for duplicate ROIs is performed as pedestrians may be detected in more than one of the closed images from the morphological clothing compensation stage.

Figure 4.5: The various stages of region growing ROI generation. (a) The original and (b) morphologically closed IR images of a pedestrian. (c) The seed regions generated by a high intensity threshold, (d) the grown region (threshold=0.7137) within ROI feature boundaries (Aspect Ratio = 0.5, Extent = 0.5575) and (e) the grown region (threshold=0.6745) outside the ROI feature boundaries (Aspect Ratio = 0.9855, Extent = 0.4358).

Figure 4.6: Result of region growing ROI generation in a complex urban scene with three pedestrian ROIs and one non-pedestrian ROI. These ROIs are subsequently classified as pedestrian or non-pedestrian by the classification stage.
4.4 Pedestrian Classification

Robustness is a key characteristic of any active VRU protection systems. False positive results, could lead to the driver being unable to trust the system. More seriously, false negative results could cause VRUs to go undetected and could lead to serious accidents. In this work Histogram of Oriented Gradients (HOG) features are extracted from ROI and classified as pedestrian or non-pedestrian using a Support Vector Machine (SVM) [117].

4.4.1 Histogram of Oriented Gradients (HOG) Features

For training and validation purposes, a database of grey scale images was formed by processing a database of 630 image frames, extracted from on-road IR video data, with the ROI region growing algorithm. From this, a collection of 800 ROI image segments were extracted and manually labelled (400 pedestrian, 400 non-pedestrian). Examples of some of the database entries are shown in Figure 4.7.

![Examples of database entries](image)

Figure 4.7: Examples of (a) pedestrians and (b) non-pedestrians from the database training set. Non-pedestrians are generated by the ROI algorithm and are therefore regions that could potentially be mistaken for a pedestrian.

Histograms of Oriented Gradient (HOG) feature vectors were extracted from generated ROIs. HOG were introduced by Dalal and Briggs [118] and has been proven...
to be an effective feature selection technique for pedestrian detection in the visual spectrum, especially when combined with SVM. To calculate a HOG feature vector for a source image, such as the pedestrian ROI image segment in Figure 4.8 (a), it must first be scaled to a pre-determined size. All ROIs are scaled to 20x40 pixels for this process, as in [74]. The gradient of this source image is computed by performing a convolution with the horizontal and vertical gradient kernels

\[ G_x = [-1, 0, 1] \ast A \]  
\[ G_y = [-1, 0, 1]^T \ast A \]

and combining the results. An example of a gradient image of a pedestrian is presented in Figure 4.8 (b). This image is split into rectangular sub-regions or cells, as shown in Figure 4.8 (c). A histogram of edge orientation magnitudes is computed within each cell, where each pixel “votes” for an orientation and each bin represents the strength of an edge in certain directions from 0° - 180°. Each vote is weighted by the magnitude of the gradient at that pixel. As gradient strengths can vary widely between ROIs, contrast normalisation must be applied to ensure good performance. Cells are grouped into larger overlapping blocks and each is normalised separately using Lowe-style clipped \( L^2 \)-norm [119]. The \( L^2 \)-norm of an unnormalised vector containing histograms of a certain block (\( u \)) is:

\[ v = \frac{u}{\sqrt{\|u\|^2 + \epsilon^2}} \]  

where \( \epsilon \) is a small constant to prevent division by zero. The values of \( v \) are clipped at a maximum of 0.2 as suggested by Lowe [119]. This normalisation scheme is referred to by Dalal and Briggs as \( L^2 \)-Hys. The final normalised vector forms a feature vector that is representative of the source ROI image segment and is subsequently used for classification using a SVM. A resulting HOG descriptor is shown in Figure 4.8 (d),
where the gradients orientation histograms within each cell can be observed.

![Images](a) (b) (c) (d)

**Figure 4.8**: Images from the various stages of generating a Histogram of Oriented Gradients (HOG) feature vector. (a) Original pedestrian image, scaled to 20×40 pixels, (b) gradient image, (c) image divided into cells of 5×5 pixels, resulting in 4 cells × 8 cells (d) resulting HOG descriptor for the image showing the gradient orientation histograms in each cell.

The parameters used to extract the HOG features were based on procedures described in [118] and were empirically refined. The parameters are presented in Table 4.2.

**Table 4.2**: Parameters used for Histogram of Oriented Gradient feature extraction

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Width</td>
<td>5 pixels</td>
</tr>
<tr>
<td>Cell Height</td>
<td>5 Pixels</td>
</tr>
<tr>
<td>No. of Orientations</td>
<td>9 (0-180°)</td>
</tr>
<tr>
<td>Overlap</td>
<td>0.5</td>
</tr>
<tr>
<td>Normalisation Method</td>
<td>L₂-Hys</td>
</tr>
</tbody>
</table>

### 4.4.2 Support Vector Machine Classifier

A Support Vector Machine (SVM) is a data classification tool which calculates the optimal separating hyperplane between classes in high dimensional space. In this work, it is used in the classification stage to determine if new ROIs are pedestrian...
or non-pedestrian. In the training phase, the SVM analyses a set of features vectors derived from the image database described above, which are labelled as pedestrian and non-pedestrian. The SVM was trained on the HOG feature vectors from the image database described in Section 4.4.1. Several practical steps were taken to achieve optimal results from the classifier, many suggested in [120]. A Radial Basis Function (RBF) is used as a kernel ($K$) for SVM classification:

$$K(x, y) = e^{-\gamma \|x-y\|^2} \tag{4.9}$$

An RBF kernel was used in accordance with the technical report from Hsu et al. [120]. The RBF kernel has less parameters to set than the polynomial kernel and is well suited to training with a low number of features. To avoid over-fitting and to ensure accurate performance metrics, training was performed using K-fold cross-validation, with $K=10$. This entailed breaking the training set into 10 subsets, then sequentially testing one subset, using the classifier trained on the remaining 9 subsets. In addition, a grid search was used to determine optimal SVM parameters, cost ($C$) and $\gamma$.

### 4.5 Target Tracking

Tracking between frames plays an important role in adding robustness to the system, by estimating the future positions of objects that have been detected based on previous measurements. This data can be used to interpolate between failed detections when the algorithm fails to detect the object or the object is temporarily occluded.

As with the vehicle tracking system presented in Section 3.5, a Kalman filter [71, 72] is used for tracking pedestrians between frames. The filter estimates current pedestrian position based on incomplete or noisy measurements. It is ideally suited to automotive object recognition systems because it can continue tracking when objects
are occluded and can cope with vehicle movement. Once a ROI has been classified as a pedestrian, it is tracked using four parameters of an enclosing bounding box (x-position, y-position, width and height); these form a measurement vector $z$. The set of equations that define the Kalman filter were introduced in Section 3.5.

To add extra robustness to the system, when detection fails and the associated tracker fails to update, its predicted bounding box is examined, in a manner similar to the one presented in the vehicle tracking system in Section 3.5. The cross-correlation between the image in the predicted bounding box and the image of the corresponding pedestrian from the previous detection is calculated. If there is a high correlation between them, the tracker is updated with the new location. This is a similar concept to template matching and works on the assumption that pedestrians will not change dramatically between frames. Image cross-correlation has been used previously for template matching of pedestrians [80]. Fast normalised cross-correlation [81] is used to determine the correlation ($\gamma$) between the two ROI image segments ($J$ and $K$):

$$\gamma = \frac{\sum_{x,y} (K(x,y) - \bar{K})(J(x,y) - \bar{J})}{\sigma_K \sigma_J}$$

(4.10)

Where $J$ is the part of the current frame under the predicted tracking bounding box, $K$ is the image of the pedestrian from the previous detection and is used as a template, $\bar{K}$ is the mean value of $K$ and $\bar{J}_{u,v}$ is the mean of $J(x, y)$, the region under the template. This is the same correlation equation used to measure the symmetry between vehicle lamps in Chapter 3 (Equation (3.3)).

### 4.6 Results

#### 4.6.1 Data Capture

On-road video data were captured using a FLIR PathFindIR® automotive specification thermal infrared microbolometer (Figure 4.9 (a)). The sensor was mounted
on the bonnet of a host vehicle as shown in Figure 4.9 (b) and video data were captured at a resolution of 320×240 and a frame rate of 25Hz. The majority of video was captured in low temperature (−3°C to 8°C) winter environments, as pedestrians would be likely to be wearing well insulating clothing, testing the morphological clothing compensation algorithm. On-road video was captured in multiple types of environment (urban, rural, suburban), encompassing pedestrians at a range of distances and in many different poses and situations. Data was captured late at night, well after dusk, in wet and dry weather conditions. Video was mainly captured in urban areas as pedestrians were more likely to be present, and include host vehicle speeds of 0-100 km/h.

![Figure 4.9: (a) FLIR PathFindIR® automotive far-infrared microbolometer sensor and (b) the sensor mounted on a host vehicle during data capture.](image)

4.6.2 Performance Metrics

As described in Section 4.4 a database of 800 ROIs was extracted from this data to train the SVM classifier and analyse its performance. To analyse the performance of the video processing system, 20 video clips were extracted from the video data,

---

3Image used with the permission of FLIR Commercial Systems B.V.
each containing at least one pedestrian for the duration of the clip. There was no overlap between video data used to form the training image database and data used for video testing results. Results were compiled using a GUI similar to the one used for vehicle detection results described in Section 3.7.

A number of different approaches have been taken towards the quantification of detection results. Bertozzi et al. presented a review on performance evaluation of pedestrian detection systems [84]. In this chapter, similar to [74], results are presented in the form of a “Detection Rate” and a “Detection with Tracking Rate”. The detection rate is the proportion of frames in which a pedestrian is successfully detected \(d\), out of the total number of frames they appear in \(n\):

\[
Detection Rate = \frac{d}{n} \tag{4.11}
\]

The detection with tracking rate is the proportion of frames where a pedestrian is successfully detected or tracked \(t\), out of the total number of frames they appear in \(n\):

\[
Detection With Tracking Rate = \frac{t}{n} \tag{4.12}
\]

The incidence of false alarms is also an important factor to consider. The false positive rate is the proportion of positive detections that are not pedestrians \(f\) out of the total number of frames \(n\):

\[
False Positive Rate = \frac{f}{n} \tag{4.13}
\]

Results of the quantitative performance of the infrared pedestrian detection system are presented in two parts: firstly a statistical analysis of the performance of the SVM classifier is presented. This is followed by an analysis of results from the full video processing system, including ROI detection, classification and tracking using Kalman filtering, to demonstrate the system performance when tested with IR road video.
It should be noted that it is quite difficult to directly compare results with other systems for several reasons: firstly, previous work in this area has generally used proprietary video data and there is no known standardised automotive IR test set (a point echoed in [95] and [23]). Secondly, few previous studies have given focus to the challenge of clothing distortion. Finally, numerous algorithm parameters vary widely in previous work, such as field of view and focal length of the IR camera and the smallest size of ROI considered for classification [80, 90, 102].

By way of comparison, a system which focused on the detection of distant pedestrians (less than 30 pixels in height, equating to greater than 40m away) in FIR data reported a detection rate of 0.692 with 0.036 false positive rate [80]. They noted that distant pedestrians can be difficult to label, even for human operators attempting to determine the ground truth. Another approach using multi resolution image processing and human model fitting [95], reported a detection rate of 0.7 with a false positive rate of 0.2 from video captured in different seasons. Separate detection systems for summer and winter are presented by Fang et al. [90]. They report a detection rate of 0.95 with a false positive rate of 0.03 for a winter video sequence of 240 frames, taken in a “low-complexity” environment. In [74], summer and winter video sequences containing 39 pedestrians are analysed. This results in a detection rate of 0.35 before tracking and 0.94 after tracking with a false positive rate of 0.025.

4.6.3 Classifier Performance

The classifier was tested using K-fold cross validation with 10 groups. Each group was used to test the classifier after it was trained on the other nine, then the results were averaged. This gives a more accurate representation of the performance of the classifier as well as preventing over-fitting of the separating hyperplane in the SVM. The ROC curve for the classifier is presented in Figure 4.10, where the different points were generated using a search for optimal SVM parameters.

For testing with video the operating point was chosen at a detection rate of 0.96
Figure 4.10: ROC curve for SVM pedestrian classifier. Each point represents a detection and false positive rate for a certain parameter set. The dotted diagonal line represents the performance of a random classifier.

with a false positive rate of 0.01. This point was judged to be the best trade-off between detection rate and false positive rate. The SVM parameters that gave this performance were: cost parameter ($C$) of 2048 and RBF kernel parameter ($\gamma$) of 8.

The classifier confusion matrix for this operating point is presented in Table 4.3.

Table 4.3: Confusion matrix for SVM pedestrian classifier at chosen operating point, displaying true positive (TP), false positive (FP), false negative (FN) and true negative (TN) rates

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
<th>Pedestrian</th>
<th>Non-Pedestrian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedestrian</td>
<td>0.96 (TP)</td>
<td>0.01 (FP)</td>
<td></td>
</tr>
<tr>
<td>Non-Pedestrian</td>
<td>0.04 (FN)</td>
<td>0.99 (TN)</td>
<td></td>
</tr>
</tbody>
</table>

These performance results are an improvement on the HOG-SVM approach presented in [98] (0.87-0.91 detection rate) and are comparable to the stereo-IR HOG-SVM system results presented in [96] (0.967 detection rate, 0.025 false positive rate). These figures are also an improvement on previous winter focused results [90] (0.95 detection-rate, 0.03 false positive rate).
4.6.4 Video Processing Results

By utilising temporal information provided by video data, it is possible to create a more robust pedestrian detection system. Firstly, failed detections usually occur for only a small number of frames. In these cases the tracking algorithm prediction is used until the target is re-acquired. Secondly, as false positives usually only appear for a short number of frames, they can often be readily filtered out. Pedestrian detection results from video data are presented in Table 4.4.

Table 4.4: Pedestrian detection video processing results summary

<table>
<thead>
<tr>
<th>Length (frames)</th>
<th>Detection Rate</th>
<th>Detection After Tracking Rate</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>834</td>
<td>82.374%</td>
<td>96.643%</td>
<td>1.799%</td>
</tr>
</tbody>
</table>

A selection of images representative of the video processing system output is presented in Figure 4.11. These include examples of pedestrian detection at a range of distances and examples of detection of warm (or bright) pedestrians who are wearing poorly insulating clothing, or who have been outdoors for a short period of time. Also included are examples of detection of multiple pedestrians of varying appearance present in the same frame, sometimes in close proximity. Finally examples of detection failure and false positive detection are included for completeness.

The majority of pedestrians at close range are detected and classified correctly (Figure 4.11 (a)-(c)). It is a more challenging task to accurately classify pedestrians at greater distances (Figure 4.11 (d)-(f)), as IR sensors are generally low-resolution and they will be represented by a very small number of pixels. Pedestrian candidates of this size are also particularly challenging to classify as they will have little textural information. It should be noted that while this approach works well for detecting pedestrians distorted by insulating clothing, it is as effective in the detection of pedestrians appearing to have a uniform thermal spectrum, such as those present
in a low temperature environment for a short period of time, or those not wearing well insulating clothing (Figure 4.11 (g)-(i)). This system is also adept at detecting multiple pedestrians present in the same frame. Pedestrians with different body shapes and thermal spectra located in close proximity are detected effectively (Figure 4.11 (j)-(l)). False positives most commonly occurred in video data recorded in an urban environment. This might be expected because of the large number of other heat sources that may be confused with pedestrians. Even so, many false positive detections appear for only single frames and can be filtered by the tracking algorithm. The rural environment presented the smallest quantity of false positives which can be attributed to the relatively low number background heat sources compared to the other environments. The main causes of false positives were vehicle lamps, traffic lights and other warm vehicle parts such as tyres and exhaust pipes (Figure 4.11 (m)-(n)). Failure to detect pedestrians was most commonly caused by occlusion or proximity to another bright object (e.g. Figure 4.11 (o)).

4.7 Conclusions

An automotive infrared video processing algorithm for detection of pedestrians has been presented in this chapter. A morphology based clothing distortion compensation process adjusts for low intensity pedestrian pixels caused by this clothing, greatly assisting pedestrian segmentation. Regions Of Interest (ROIs) are segmented using region growing technique with feature-based stopping criteria. These regions are automatically labelled as pedestrian or non-pedestrian using HOG features and SVM classification. Testing using on-road automotive video data has been conducted and results indicate that this is an effective algorithm for detecting pedestrians in automotive thermal infrared video with a low rate of false positives.

When false positives do occur, they are predominantly caused by warm or hot non-pedestrian objects. These are commonly parts of other road vehicles. In the
next chapter, data from the visual vehicle detection system (presented in Chapter 3) is mapped to the IR camera coordinates and used to mask potential false positives caused by hot vehicle parts.

The pedestrian detection system presented in this chapter could be used in several ways to prevent or mitigate collisions with pedestrians. Highlighting pedestrians in an IR video stream to warn the driver is not ideal, as research has shown that IR imagery is not entirely intuitive for a human observer [21]. Therefore, in the next chapter, IR pixels from detected pedestrians are overlaid on a night-time visible spectrum video stream. This is beneficial for human consumer of the video as the environment is familiar and intuitive and pedestrians are highlighted to warn the driver of potential collisions.
Figure 4.11: A montage of results frames displaying successful pedestrian detection, representative of video processing results. These images display pedestrians detected at (a) - (c) close range and (d) - (f) at greater distances. (g) - (i) Pedestrians outdoors for a short amount of time or wearing poorly insulating clothing are detected by the same process. (j) - (l) Multiple pedestrians of varying appearance in the same frame are also detected, even in close proximity. (m) - (n) False positive detections (outlined in red) and (o) a false negative detection are also displayed.
Chapter 5

Fusion of IR and Visible Systems

5.1 Introduction

This chapter will present several benefits that can be attained through the fusion of data from the visible vehicle detection system presented in Chapter 3 and the IR pedestrian detection system presented in Chapter 4. Data from each system will be combined and utilised to gain advantages that would not be possible otherwise.

It was concluded at the end of Chapter 4, as well as in related research, that false positives in IR pedestrian detection are frequently caused by warm vehicle parts. By using vehicle location data from the visual domain to mask the IR frame during pedestrian detection, the number of hot-spots in the image and hence the probability of false positives can be reduced.

Night-time on-road visual spectrum video data is of limited use for human consumption, especially with regards to a person’s ability to recognise pedestrians. While IR video is well suited for automatic detection of pedestrians, it is not considered entirely intuitive to human viewers [21]. By transforming pixels from detected pedestrians in the IR domain, to the visual domain and replacing the pixels at these points, pedestrians are highlighted in the visual imagery. The result is an image that is intuitive to a driver, while making pedestrians far more visible. A flowchart outlining
the structure of the complete fusion system is presented in Figure 5.1.

Figure 5.1: A flow chart outlining the structure of the fusion of the visual and IR systems. Knowledge of vehicle location is used to improve the pedestrian detection system and detected pedestrians are used to highlight visual images.

The data capture equipment used to simultaneously capture IR and visual video data is shown in Figure 5.2. The optical axes of the two cameras are aligned as closely as possible to prevent parallax distortion after registration.

Visual and IR video data have been used together in previous work for driver assistance applications. Feature level fusion is conducted in [121], where IR pixels from detected oncoming vehicles are overlaid on a corresponding visual image in place of a glared headlamp appearance. It is common to use cameras in stereo configurations for this process. A “tetra-vision” approach was proposed in [97], consisting of 4 cameras (stereo-visual and stereo-IR), while a 3 camera solution (stereo-visual and a single IR camera) was presented by Krotosky et al. [122]. Disparity analysis is usually performed from stereo pairs, effectively segmenting foreground objects.
Figure 5.2: Data capture set up with far-IR microbolometer (left) and visible camera (right).

### 5.2 Image Registration

Registration of the visual and IR video data was conducted using a rectangular calibration target and a perspective transformation. Images from this registration procedure are displayed in Figure 5.3. The transformation matrix was created by mapping the four corner points of the rectangular target in the IR image to the four corner points in the visual image. The transformation, a quadrilateral to quadrilateral projective transformation, is similar to the one presented for perspective correction of vehicle lamps in Section 3.4. The matrix algebra behind this transformation is included in Appendix A. Once the videos were captured they were manually time aligned using features in the video.

### 5.3 Infrared Vehicle Masking

As described in Chapter 4, while false positives in the proposed IR pedestrian detection are not a common occurrence, when they do occur, they are frequently caused by vehicles in the frame. This is due to the fact that infrared pedestrian
detection techniques focus, in part, on the warmest and hence brightest regions of a scene. Vehicles commonly have many hot parts, such as brake discs, exhaust pipes, headlamps, rear-facing-lamps and engine covers. These items usually appear bright in an IR image, especially if the vehicle has been on the road for more than a few minutes. This link between false positives and vehicle parts is specifically acknowledged by Bertozzi et al. in [80]. They specify that the biggest issue in their system is that warm parts of vehicles get confused with parts of the body. They point out that most false positives are caused by incorrect classification of vehicles headlights and warm tyres and that the problem is especially prevalent at larger distances. An example of a false positive pedestrian detection caused by a vehicle tyre and headlamp is presented in Figure 5.4.

This issue is addressed in this thesis through data fusion with the visual vehicle detection system from Chapter 3. Bounding boxes of detected vehicle lamps are expanded vertically to encompass the rest of the vehicle. This is approximated by adding a proportion of the width of the detected vehicle to the vertical dimension of the bounding box, this is shown in Figure 5.5.

The vehicular bounding box is then transformed into the IR domain using the inverse of the transformation created in registration. All pixels inside the resulting bounding quadrilateral are set to zero. This process is conducted before any pedes-
Figure 5.4: A false positive pedestrian detection caused by warm vehicle parts such as tyres and lamps.

Figure 5.5: Expansion of vehicle lamp bounding box in visual domain to encompass the rest of the vehicle.

Triangular detection takes place, so no warm vehicle areas can be considered as seeds in the region growing ROI generation process and no vehicle pixels will be included in region growth. The result of this is a more robust pedestrian detection system with less probability of false positives. Results from the various stages of this process are presented in Figure 5.6.

5.4 Visual Pedestrian Highlighting

In low light conditions, a visible spectrum camera is severely limited, in terms of viewing on-road pedestrians, by the range of the headlamps and the presence of
Figure 5.6: Vehicle masking in IR pedestrian detection system through fusion with the visual vehicle detection system. (a) Vehicle detected in visual video frame and bounding box expanded to encompass entire vehicle, (b) IR image with vehicle and pedestrians present, (c) bounding box transformed to IR image, vehicle pixels masked to zero.

street lighting. This is evident in the visual image of a person at 35m captured under dimmed headlamps which was displayed in Figure 4.1 (a). This affects the practicality of using a visual camera for automatic detection of pedestrians or as a visual aid for the driver.

As described in Chapter 4, far-infrared technology is well suited to the task of
automatic pedestrian detection. However, from a human consumption point of view, far-infrared imagery is considered unnatural and difficult to understand \[21\]. In this system, pixels from pedestrians detected in the IR domain are transformed and overlaid on the visual image. This results in video that is natural to the driver, which they can easily interpret, but with the advantage of highly visible pedestrians. Sample images from this pedestrian highlighting system are presented in Figure 5.7.

![Figure 5.7: Pedestrian highlighting in visual image through fusion with IR pedestrian detection system. (a) Pedestrians detected in an IR video frame. (b) Corresponding visual image. (c) Pixels from detected pedestrians in IR image transformed to visual space and overlaid on image.](image)

### 5.5 Conclusions

In this chapter a fusion of the vehicle and pedestrian detection systems was described.

Images from a visible and an IR camera are registered using a rectangular calibration target and a quadrilateral to quadrilateral projective transformation. Vehicle detection image coordinates determined from the visual vehicle detection system are transformed into the domain of the IR camera. This is used to mask the area of the IR frame containing vehicles during the pedestrian detection process. This has the effect of reducing the number of non-pedestrian hotspots in the IR frame and hence reduces the probability of false positives.
Pixels from detected pedestrian in the IR domain, are transformed to the visual domain, replacing the pixels at these points. This results in highlighted pedestrians in the visual imagery, which are useful for a human consumer of these images.

The result is a system that detects the two main classes of other road users at night, working collaboratively and fusing data in both directions for mutual benefit. In the next chapter final conclusions will be drawn, the key contributions of this thesis will be restated and possible directions for future work will be discussed.
Chapter 6

Conclusions and Future Work

6.1 Project Summary and Conclusions

This thesis has described a night-time road user detection system using visual and IR automotive video streams.

A video processing system that can reliably detect and track multiple vehicles in low light conditions has been proposed. Road vehicles were detected using a regular low cost CMOS colour camera. A camera configuration process has been presented which addresses the issues of reproducing and verifying results, portability between different camera hardware, as well as ensuring lamp colour information is not lost due to pixel saturation. A low static exposure camera configuration was implemented to eliminate lamp “blooming” and ensure consistency of target appearance.

Red colour thresholds for identifying rear tail- and brake lamps have been derived from automotive regulations and adapted to real world conditions utilising the HSV colour space. A region growing technique has been described to identify headlamp candidates. A shape and size independent cross-correlation bilateral symmetry analysis approach to pairing detected lamps has been presented. An automatic projective transformation has been employed to correct for perspective distortion of lamp pairs, ensuring robust performance of vehicle detection, especially during overtaking and
sharp bends. A Kalman filter tracking algorithm has been implemented to improve robustness by ensuring continuity of operation through small detection failures and predicting future location of targets. Results from on-road testing have been presented, demonstrating the system’s high detection rates, low false positive rates and robustness in different environments.

An automotive infrared video processing algorithm for detection of the most vulnerable road users: pedestrians, has been described. A morphology based process to compensate for clothing distortion adjusted for low intensity pedestrian pixels caused by well insulating clothing, greatly assisting pedestrian segmentation. ROIs were segmented using region growing technique with feature-based stopping criteria. High intensity seeds were grown until a viable pedestrian candidate no longer possessed the typical features of a pedestrian. HOG features were extracted from a database of images to train a support vector machine classifier, labelling ROIs as pedestrian or non-pedestrian. Detected pedestrians were tracked using Kalman filtering to smooth noise, to extrapolate position during short periods where detection has failed and to predict the future position of targets. Testing using on-road automotive video data indicated that this is an effective algorithm for detecting pedestrians in automotive thermal infrared video.

Finally, a cooperative fusion of the visual vehicle detection and IR pedestrian detection systems has been described. The two domains were registered using a rectangular calibration target and a four point quadrilateral to quadrilateral projective transformation. Visual vehicle detections were used to mask vehicles in the IR frame, reducing the chance of false positives by eliminating those caused by warm vehicle parts such as lamps and exhaust pipes. Concurrently pixels from detected pedestrians in IR video were used to highlight the visual image, making pedestrians much more visible for human video consumers.
6.2 Primary Contributions

A system to detect other road vehicles and pedestrians in low light conditions has been presented in this thesis. The primary contributions of this thesis can be defined as:

1. A configuration process for a colour camera has been put forward. The configuration optimises the appearance of automotive lamps for identification, while diminishing the appearance of lower intensity non-lamp objects. This process ensures that the appearance of vehicle lamps in video data is not affected by ambient lighting conditions. A polarizing filter is used to reduce the intensity of reflections.

2. An algorithm for identifying rear-facing vehicle lamps from low exposure images has been presented. To derive parameters for a colour threshold, limits from regulations were transformed to the RGB colour space. These limits were adapted to real world conditions in the HSV colour space.

3. A process for identifying vehicle headlamps from low exposure images has been proposed. A high intensity threshold finds seed regions inside lamps which are systematically grown until reaching an image edge.

4. A vehicle detection system has been put forward, based on the pairing of identified lamps using symmetry analysis. A perspective correction process ensures robust detection during road bends and overtaking manoeuvres, a problem not considered in related work. Multiple vehicle situations are considered and targets are tracked using the Kalman Filter. Results have been presented verifying high detection rates and low false positive rates for vehicle detection at night.

5. An infrared pedestrian detection system has been presented. A morphological pre-processing stage compensates for image distortion caused by well insulating
clothing worn by pedestrians. High intensity seeds are grown to form ROIs which are then classified using HOG features and an SVM classifier. Detected pedestrians are tracked using the Kalman Filter.

6. A night-time road user detection ADAS has been put forward, which intelligently combines the visual vehicle detection and IR pedestrian detection systems. Knowledge of the location of detected vehicles is utilised to mask the corresponding IR image and lower the probability of false positives in pedestrian detection. The IR image of detected pedestrians is fused with the corresponding visual image to create an enhanced image, where pedestrians are highlighted.

6.3 Suggestions for Future Work

There are several potential areas that could be focused upon in future work:

1. To encompass detection of all types of other road vehicles, single light vehicles such as motorcycles should be considered. While bilateral symmetry may not be an optimal feature for detection of these vehicles, a temporal method such as optical flow may be effective.

2. The accuracy of the pin-hole camera distance estimation could be increased by using a reference object with a known real world size such as a number plate.

3. To ensure robust detection of all types of VRU, many more categories must be examined. Sitting or prone pedestrians, children, cyclists, animals, groups of people occluding each other and other possibilities should be considered. Adapting the proposed system, using customised parameters for ROI segmentation and dedicated classifiers may be adequate for some types.

4. The main challenge to be overcome in pedestrian detection is occlusion caused by other pedestrians when people are in close groups. Detection of groups
could be attempted by utilising head detection, as people’s heads frequently remain unoccluded when closely positioned. However, detection of groups is a thoroughly complex and difficult problem.

5. Finally, development of embedded implementation of the algorithms presented in this thesis, with the real time performance, would be desirable.
Bibliography


Appendix A

Projective Transformation

The projective transform utilised for the perspective correction of rear lamps in Chapter 3 and the image registration between IR and visual cameras in Chapter 5 is performed through a unit square, a technique presented and well documented by Wolberg [123].

The projective transformation matrix is formed by combining the transformations of the bounding quadrilateral to a unit square \((A_x)\) and the rectangle to a unit square \((A_u)\) (Figure A.1), as presented in [123].

This simplifies and speeds up the calculation of the transformation matrix. To construct the transformation matrix \((A_x)\) which will transform an arbitrary quadrilateral \((x_k, y_k)\) to a unit square, the following terms are defined:

\[
\Delta x_1 = x_1 - x_2, \quad \Delta x_2 = x_3 - x_2, \quad \Delta x_3 = x_0 - x_1 + x_2 - x_3 \\
\Delta y_1 = y_1 - y_2, \quad \Delta y_2 = y_3 - y_2, \quad \Delta y_3 = y_0 - x_1 + y_2 - y_3
\]

The two elements of the transformation matrix that determine the perspective
Figure A.1: A projective transformation is formed by combining the transformations of the bounding quadrilateral to a unit square \( (A_x) \) and an arbitrary rectangle to a unit square \( (A_u) \).

The quadrilateral to unit square projective transformation matrix can then be defined as (A.3):

\[
A_x = \begin{pmatrix}
  x_1 - x_0 + a_{13} x_1 & y_1 - y_0 + a_{13} y_1 & a_{13} \\
  x_3 - x_0 + a_{23} x_3 & y_3 - y_0 + a_{23} y_3 & a_{23} \\
  x_0 & y_0 & 1
\end{pmatrix}
\] (A.3)

The transformations for bounding quadrilateral and rectangle are then combined...
to give a single transformation matrix (A.4).

\[ A = \frac{A_u}{A_x} \]  \hspace{1cm} (A.4)
Appendix B

Distance Estimation

Rear-lamps are present on vehicles to indicate location and width of a vehicle to other drivers and therefore must be located towards the extremities of a vehicle, as far apart as practically possible. Thus the width of the target vehicle in an image (in pixels) can be obtained from the detection and tracking results. Combining this information with the characteristics of the camera, one can form a geometrical approximation of the distance to the target vehicle from a pin-hole camera perspective projection model. This approach relies on the assumption that the road is a flat plane, as documented in [32] and displayed in Figure B.1.

The equation used to estimate the distance to the target vehicle is:

\[ Z = cf \frac{W}{w} \tag{B.1} \]

Where \( Z \) is the estimated distance to the target vehicle in metres, \( c \) is the conversion from pixels to mm for the image sensor, \( f \) is the focal length of the lens in mm, \( W \) is the width of an average vehicle in metres and \( w \) is the width of the target vehicle in pixels. In vehicle detection system presented in Chapter 3, the average width of a vehicle \((W)\) was taken as 1.7m, as in [36]. The result of this generalisation is that error in estimated distance is directly proportional to the difference between this average width and the actual vehicle width.
Figure B.1: Model of camera and parameters used to estimate distance.
Appendix C

Publications

The publications related to the work presented in this thesis are presented in this section. Copies of published journal publications are included for completeness, while other publications can be accessed at http://car.nuigalway.ie/students/romalley/index.html.

C.1 Journals

C.1.1 Published


- Ronan O’Malley, Martin Glavin and Edward Jones, “Vision Based Detection

C.1.2 In Preparation

- Diarmaid O’Cualain, Ronan O’Malley, Ciarán Hughes, Martin Glavin and Edward Jones, “Forward facing automotive visual and IR camera fusion system with aerial view for night road environments”.

C.2 Book Chapter


C.3 Conferences


- Ronan O’Malley, Martin Glavin and Edward Jones, “Vehicle Detection at Night Based on Tail-Light Detection”, *1st International Symposium on Vehicular*