THALES

Analysis of infrared polarisation signatures for vehicle detection

Christopher N. Dickson, MChemPhys (Hons)

A thesis submitted in fulfilment of the degree of

Engineering Doctorate

Engineering & Physical Sciences
Heriot Watt University

Land and Air Systems
Thales UK

June 2015

Abstract

Thermal radiation emitted from objects within a scene tends to be partially polarised in a direction parallel to the surface normal, to an extent governed by properties of the surface material. This thesis investigates whether vehicle detection algorithms can be improved by the additional measurement of polarisation state as well as intensity in the long wave infrared.

Knowledge about the polarimetric properties of scenes guides the development of histogram based and cluster based descriptors which are used in a traditional classification framework. The best performing histogram based method, the Polarimetric Histogram, which forms a descriptor based on the polarimetric vehicle signature is shown to outperform the standard Histogram of Oriented Gradients descriptor which uses intensity imagery alone. These descriptors then lead to a novel clustering algorithm which, at a false positive rate of 10^{-2} is shown to improve upon the Polarimetric Histogram descriptor, increasing the true positive rate from 0.19 to 0.63.

In addition, a multi-modal detection framework which combines thermal intensity hotspot and polarimetric hotspot detections with a local motion detector is presented. Through the combination of these detectors, the false positive rate is shown to be reduced when compared to the result of individual detectors in isolation.

Acknowledgements

I would like to acknowledge my academic and industrial supervisors: Andy Wallace, Matt Kitchin, and Barry Connor. Thank you to Andy for questioning everything with a critical and constructive eye, you have helped propel my work forward while ensuring everything stayed on track. Thank you to Matt for always being willing to find time to talk and for your extensive help in reading multiple draft versions of papers, presentations, and thesis chapters. Thank you to Barry for making me aware of the larger applications of this project within Thales, and for encouraging and supporting me to showcase this work at a number of internal and external events.

Additional thanks go to Thales and the Engineering and Physical Sciences Research Council for funding this work and allowing me to present at a number of international forums. These events have not only helped to share my work but have also introduced me to a number of people I may not otherwise have met. Also, thanks to the large number of well-informed people within Visionlab at Heriot Watt, and the Algorithms group at Thales, who were always available to talk and provide help where necessary.

Finally I would like to thank my fiancée Catherine, and all of my family for supporting me throughout my study.

ACADEMIC REGISTRY
Research Thesis Submission

Name*:*	Christopher Nigel Dickson		
School/PGI:	Institute of Signals, Sensors and Systems		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought (Award **and** Subject area)	Engineering Doctorate

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission *(Handed in to SSC; posted through internal/external mail):*			
*E-thesis Submitted (**mandatory for final theses)***			
Signature:		Date:	

Please note this form should be bound into the submitted thesis.

Updated February 2008, November 2008, February 2009, January 2011

Contents

List of Figures

List of Tables

List of Algorithms

Nomenclature

Roman Symbols

$B_\lambda(T)$	Blackbody radiance from an object at temperature T
c	Speed of light in vacuum, $c \approx 3 \times 10^8\,\mathrm{m\,s^{-1}}$
e	Ellipticity parameter of polarisation
fn	False Negative Detection
fp	False Positive Detection
h	Planck constant, $h \approx 6.626 \times 10^{-34}\,\mathrm{J\,s}$
$i(\theta, \xi)$	Intensity of light vibrations in the direction making an angle θ with OX, when the y-component is subjected to a retardation ξ with respect to the x-component.
k_B	Boltzmann constant, $k_\mathrm{B} \approx 1.381 \times 10^{-23}\,\mathrm{J\,K^{-1}}$
M	Mueller matrix
P	Degree of polarisation
P_linear	Degree of linear polarisation
tn	True Negative Detection
tp	True Positive Detection
$\vec{E}(z,t)$	Electric field with $\vec{z}$ the direction of propagation and t time
fpR	False positive rate
$\vec{S}$	Stokes vector

I	Stokes I component, equivalent to S_0
Q	Stokes Q component, equivalent to S_1
U	Stokes U component, equivalent to S_2
V	Stokes V component, equivalent to S_3
S_0	Stokes S_0 component, equivalent to I
S_1	Stokes S_1 component, equivalent to Q
S_2	Stokes S_2 component, equivalent to U
S_3	Stokes S_3 component, equivalent to V
tpR	True positive rate

Greek Symbols

$\alpha(T, \lambda, \theta)$	Absorption with T temperature, λ wavelength and θ emission angle
ϵ	Polarisation ellipticity angle
$\epsilon(T, \lambda, \theta)$	Emissivity with T temperature, λ wavelength and θ emission angle
λ	Wavelength of light
ϕ	Angle of linear polarisation
ρ	Reflectivity
τ	Transmissivity

Acronyms

DS1	Data Set 1
DS2	Data Set 2
FAST	Features from Accelerated Segment Test
HC	Hypothesis Combination Method
HG	Hypothesis Generator
HOG	Histogram of Oriented Gradients
HPA	Histogram of Polarisation Angle

HPAD	Histogram of Polarisation Angle Differences
LiDAR	Light Detection And Ranging
LWIR	Longwave Infrared
MWIR	Midwave Infrared
NIR	Near Infrared
PH	Polarimetric Histogram
QWIP	Quantum Well Infrared Photodetector
RANSAC	RANdom SAmple Consensus
ROC	Receiver Operating Characteristic
VATIC	Video Annotation Tool from Irvine, California
VJ	Viola and Jones algorithm [55]

Chapter 1
Introduction

Object detection in still and video imagery is a key component in a number of commercial applications including road traffic management, driver assistance technology, and general situational awareness systems. Typically detection algorithms learn the difference between features extracted from a training set of object and non-object examples.

Different sensors can reveal different distinguishing features from a scene. Sensors which detect visible light (light visible to humans with wavelengths in the range 390 nm to 750 nm) are the most common with ubiquitous colour and greyscale sensors spanning a multitude of consumer and professional applications. Visible imagery is immediately interpretable by humans, as it represents the human view of the world, but can present problems for computer vision systems. The illumination of a scene can change significantly across an image due to the presence of bright and shadowy areas or over time as day changes to night. To capture a useful image at all times of day, sensors operating in visible wavelengths either require the ability to a record a high dynamic range of brightness levels or the ability to control the lighting in some way.

At slightly longer wavelengths, just beyond what is visible to the human eye, near infrared imagery (NIR, with wavelengths of 750 nm to 1200 nm) has been shown to help in assessing the health of vegetation [1] allowing the improved segmentation of vegetation from a scene [2]. It can also be used to see through atmospheric haze [3], and due to the different absorption spectrum from visible wavelengths, NIR dyes can also be used for forgery prevention [4].

Imagery in the mid- and long-wavelength infrared (MWIR at 3 µm to 8 µm and LWIR at 8 µm to 15 µm respectively) is commonly used in military applications. LWIR imagery, often known as thermal imagery, measures the heat emitted from objects within a scene. This means that similar imagery can be recorded day or night, and thermal imagery can reveal objects which may be camouflaged in visible imagery. A range of other imaging sensors operate in further regions of the optical

spectrum, each capable of capturing different phenomena in a scene.

Optical sensors, such as those mentioned above, will typically measure the intensity of light within their wavelength range. However, reflection and emission from surfaces gives rise to partial linear polarisation, the parameters of which depend on the properties of the surface and its geometry. By capturing the polarisation state in addition to intensity, it is possible to distinguish different types of surface [5], and also to obtain information about the three dimensional structure of a scene [6, 7].

In subsequent chapters the theory of polarisation is explained in more detail. The main contribution of this thesis is the development of a novel vehicle detection algorithm derived from polarisation imagery captured in the longwave infrared.

1.1 Motivation

An Engineering Doctorate is a doctorate level research project which is guided by the research requirements of industry. This work was performed between the Industrial Doctorate Centre in Optics and Photonics Technologies at Heriot Watt University and Thales UK in Glasgow. The project was co-sponsored by both Thales and the Engineering and Physical Sciences Research Council.

Thales is a large multinational company that designs and builds electronic systems for the aerospace, space, defence, security, and transportation industries. At the Glasgow site, the principal focus is on the design and manufacture of a variety of specialist optical systems including high quality thermal cameras, and a recent polarimetric infrared imager. In addition to designing the hardware, Thales also develop algorithms designed to improve the image quality and to provide additional information to the end user.

This EngD project addresses the following challenges:

1. Can a successful vehicle detection algorithms be developed using the infrared polarimetric content of a scene, and does this show improved performance above algorithms which process infrared intensity imagery alone?

2. Can the polarimetric content of a scene be combined with other information, such as infrared intensity, or visible imagery, to further reduce the false alarm rate of detections?

1.2 Thesis Outline

Chapter 2 introduces the theory of electromagnetic radiation and explains the phenomenon of polarisation. This chapter provides definitions for a number of key con-

cepts used throughout the thesis, such as the set of four Stokes parameters which form the Stokes vector, and the degree and angle of polarisation.

Chapter 3 introduces the Catherine MP LWIR Polarimeter, a polarimetric variant of a long-wave infrared imaging sensor manufactured by Thales. The layout of the sensor is explained, and the method of obtaining polarimetric imagery from its output is outlined. Following this, the results from some calibration experiments, and example images captured with the sensor are shown. Finally, two datasets captured for the design and testing of polarimetric vehicle detection algorithms are presented.

Chapter 4 forms a literature review and an introduction to the methods of object detection in imagery. This chapter grounds the contributions of the thesis in the context of some of the peer-reviewed literature in feature detection, classification, and object detection, and highlights work where polarimetric imagery has previously been used to enhance these tasks.

Chapters 5 and 6 present the major novel content of this thesis, the development of a polarimetric descriptor for vehicle detection. The first set of descriptors, discussed in Chapter 5, are based on computing histogram representations of the data; while the second approach, discussed in Chapter 6, instead forms descriptors from the properties of clusters detected in the polarimetric imagery. The cluster-based descriptors are shown to outperform previously presented methods for vehicle detection.

Chapter 7 presents an initial case study into the formation of a multi-sensor vehicle detection system. Here, the idea of a series of detection pipelines is introduced, where different imagery (visible, thermal and polarimetric) feed into different detection algorithms, the output of which is combined to form a final detection map. As this is a preliminary study, the detectors discussed in this chapter are much simpler than those previously presented, with the main focus being the combination of different sensors and detectors to reduce the overall false alarm rate of individual detectors.

Chapter 8 concludes the thesis by combining the different areas of the thesis and suggests some future development paths for the work.

Chapter 2
Polarimetry

This chapter introduces the theory of electromagnetic radiation as relied on in later chapters of this thesis. Section 2.1 briefly covers the representation of light as an electromagnetic wave. Section 2.2 expands on how the polarisation properties of light can be represented and measured, introducing the Stokes Vector which is used throughout this thesis to represent the polarisation state. Section 2.3 then presents an overview of the theory of blackbody emission from objects at different temperatures. This leads naturally on to Section 2.4, an examination of the mechanisms by which radiation emitted from or reflected off a surface can become partially polarised, and the subsequent interplay between emission and reflection.

2.1 Electromagnetic Waves

Light has been regarded as electromagnetic in nature since the results of James Clark Maxwell in the late nineteenth century [8, 9]. The modern view of Quantum Electrodynamics describes light in terms of massless elementary 'particles' known as photons. While photons are an important concept, especially for phenomena such as blackbody radiation, in many areas of Optics the quantum nature of light is not readily observed. For the purposes of this thesis light is more conveniently characterised as an electromagnetic wave in terms of its wavelength, amplitude, and polarisation state.

Electromagnetic waves are fully described by four vectors, related through Maxwell's equations [9, 10]: the electric-field strength, $\vec{E}$, the magnetic-field strength, $\vec{H}$, the electric-displacement density, $\vec{D}$, and the magnetic-flux density, $\vec{B}$. The dominant force between light waves and matter is the electric interaction between the electric-field of the light and the electrons in the matter. As a result the principal focus shall be the electric-field vector, considered to be a transverse wave of the form

$$\vec{E}\left(z,t\right) = \vec{A}\cos\left(\omega t + \vec{k}\cdot\vec{z} + \delta\right), \tag{2.1}$$

where $\vec{A}$ describes the orientation and magnitude of the electric field, ω is the angular frequency, t is time, $\vec{k}$ is the wave vector, $\vec{z}$ is the direction of propagation, and δ is a constant phase shift.

The angular frequency is related to the wave vector and the wavelength, λ, through the following relation involving the speed of light, c,

$$|k| = \frac{\omega}{c} = \frac{2\pi}{\lambda}. \tag{2.2}$$

Maxwell was able to theoretically predict the speed of all electromagnetic waves in a vacuum as a function of the permittivity, ϵ_0, and permeability, μ_0, of free space

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \approx 3 \times 10^8 \,\mathrm{m\,s^{-1}}. \tag{2.3}$$

2.2 Polarisation

The majority of light sources produce incoherent and unpolarised light as they produce a random combination of waves with different spatial characteristics, frequencies, phases and polarisation states. As an initial simplification, consider a coherent plane wave that is travelling in a single direction, with defined frequency, phase, and polarisation state. If this wave is travelling perpendicular to the plane of the page, coming towards the reader along the positive $z-$axis, the electric vector will in general trace an elliptical pattern in the plane of the page, the $x-y$ plane, in a regular repetitive manner such as that shown in Figure 2.1 [11]. This ellipse, known as the polarisation ellipse as it uniquely defines the polarisation state of the wave, may be described by the path of the $x-$ and $y-$components of (2.1)

$$E_i = A_i \cos\left(\omega_0 t + \delta_i\right) \qquad i = x, y, \tag{2.4}$$

where A_i is the amplitude of the $i-$component, ω_0 the frequency of the wave, and δ_i the phase of the $i-$component. The polarisation ellipse can be more conveniently defined using the polarisation parameters described below.

2.2.1 Polarisation Parameters

The *azimuth*, or *angle of polarisation*, ϕ, measures the angle between the major axis of the ellipse and the positive direction of the x-axis and defines the orientation of the ellipse in its plane. It is limited in value, for all physically distinguishable

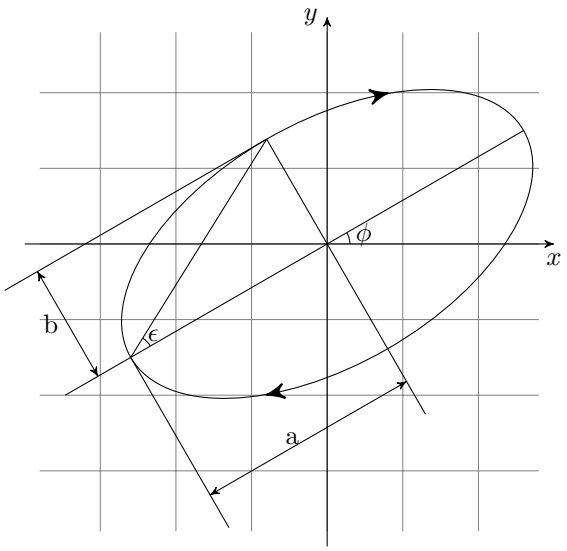


Figure 2.1: The four parameters which define the polarisation ellipse are the azimuth, or angle of polarisation, ϕ, the ellipticity $e = \pm\frac{b}{a} = \tan\epsilon$ where a and b are the lengths of the semi-major and semi-minor axes of the ellipse respectively and ϵ the ellipticity angle, the total amplitude $A = \sqrt{a^2 + b^2}$, and the absolute phase δ as defined by the field at $t = 0$. The pictured polarisation is right-handed.

ellipses, to

$$\frac{-\pi}{2} \leq \phi < \frac{\pi}{2}. \tag{2.5}$$

The *ellipticity*, e, is defined by the ratio of the length of the semi-minor axis, b, to the length of the semi-major axis, a, of the ellipse,

$$e = \frac{b}{a}. \tag{2.6}$$

The *handedness* of the ellipse defines the direction in which the electric vector rotates over time, such that an ellipse is termed right-handed if it rotates in a clockwise direction when observed against the direction of travel. The handedness is encoded into the ellipticity by allowing e to span positive and negative values for right-handed and left-handed polarisations respectively thus all distinguishable ellipses are defined with e in the range

$$-1 \leq e \leq 1. \tag{2.7}$$

Further, the ellipticity is more conveniently represented through the equivalent *ellipticity angle*, ϵ, such that

$$e = \tan\epsilon. \tag{2.8}$$

The ellipticity angle is limited to the range

$$-\frac{\pi}{4} \leq \epsilon < \frac{\pi}{4}. \tag{2.9}$$

Linear and circular polarisation states are special cases of polarisation and occur when the ellipticity takes on the value of zero and ± 1 respectively. For circular polarisation $e = -1$ corresponds to left-handed circular polarisation and $e = +1$ corresponds to right-handed circular polarisation.

Monochromatic light, with defined propagation direction, frequency, phase and polarisation state, can always be described by the polarisation ellipse around which its electric vector processes and as such is known as *completely polarised* light. In general, most sources of light produce incoherent light which is thus *unpolarised*. However, if one particular polarisation state contains more power than another, the light may be described as *partially polarised*. Partially polarised light may statistically be described as the superposition of unpolarised and completely polarised components with the *degree of polarisation*, P, representing the fraction of the total power carried by the polarised component. Unpolarised light is characterised by $P = 0$, completely polarised light by $P = 1$ and partially polarised light by $0 < P < 1$.

By viewing partially polarised light over discrete time intervals, the two field components can be represented by

$$E_i = A_i(t) \cos(\omega_i t + \delta_i(t)) \tag{2.10}$$

where the amplitudes $A_i(t)$ and phases $\delta_i(t)$ are functions of time. Note that (2.4) is simply a special case of (2.10) where $A_i(t)$ and $\delta_i(t)$ are constant over time.

2.2.2 Modifying the Polarisation State

The polarisation state can be modified by passing the light though optical filters. Two of these, the linear polariser and the quarter wave plate, are briefly introduced. A diagram highlighting the operation of these filters, and the passed polarisation states is shown in Figure 2.2.

Linear Polariser

A linear polariser is an optical filter which passes light with a specific angle of polarisation and blocks other polarisation states. There are a number of different methods of constructing a linear polariser, the simplest of which being a wire-grid polariser. A wire-grid polariser consists of an array of fine parallel metallic wires

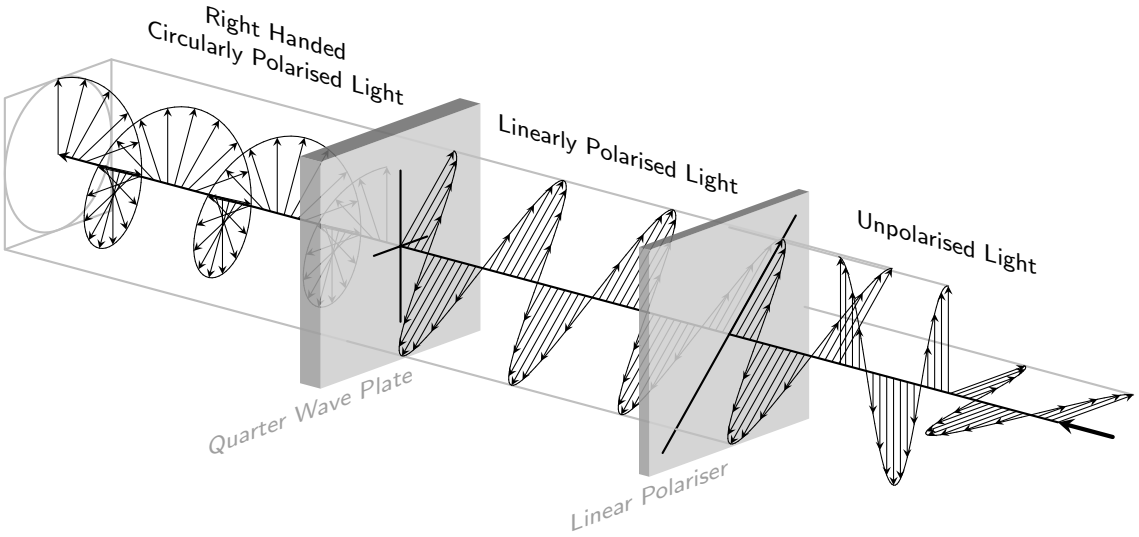


Figure 2.2: An example showing initially unpolarised light pass through a linear polariser to become linearly polarised, and then through a further quarter wave plate to become circularly polarised.

placed in a plane orthogonal to the direction of travel of the incident beam. Incident waves with their electric field aligned with the orientation of the wires induce motion of the electrons in the wires in a similar manner as if they were incident on a metal sheet, and so are reflected. Incident waves with their electric field aligned perpendicular to the orientation of the wires are able to pass directly through the grid, as the electrons in the wire can only move as far as the wires are wide. For this type of polariser to work, the separation between the wires must be less than the wavelength of light, and the width of wires a small fraction of this distance.

Quarter Wave Plate

A wave plate is a type of optical retarder which alters the phase difference between the orthogonal components of the incident light. When the phase difference introduced is $\frac{\pi}{2}$, or a quarter of the wavelength, the device is called a quarter wave plate, and it can convert incident light between linear and circular polarisation.

A wave plate can be made from a birefringent crystal cut such that the optical axis is parallel to the surfaces of the plate. This results in two axes in the plane of the cut with different refractive indices, thus orthogonal components of the incident light will travel at different speeds through the crystal, introducing a phase difference. The amount of relative phase introduced can be determined by the refractive indices of each axis, the thickness of the crystal, and the wavelength of the incident light.

2.2.3 Stokes Vector

There are a number of equivalent representations of the polarisation state of light including the Poincaré sphere, the Jones vector, and the Stokes vector, of which the Stokes Vector will principally be used. The Stokes vector, $\vec{S}$, was formulated by George Gabriel Stokes in 1852 [10] and consists of four components known interchangeably as S_0, S_1, S_2, and S_3 or I, Q, U, and V.

$$\vec{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} \tag{2.11}$$

These components, each of which have units of intensity, are defined in terms of the Cartesian components of the electric field vector, $E_x(t)$ and $E_y(t)$ (2.10), as

$$I = \left\langle E_x^2(t) \right\rangle + \left\langle E_y^2(t) \right\rangle \tag{2.12a}$$

$$Q = \left\langle E_x^2(t) \right\rangle - \left\langle E_y^2(t) \right\rangle \tag{2.12b}$$

$$U = 2\left\langle E_x(t) E_y(t) \cos\delta \right\rangle \tag{2.12c}$$

$$V = 2\left\langle E_x(t) E_y(t) \sin\delta \right\rangle \tag{2.12d}$$

where $\langle\cdot\rangle$ denotes the time average, and $\delta = \delta_y - \delta_x$ is the phase difference between the components with δ_x and δ_y introduced in (2.10). It can be seen that the first parameter, I, measures the total intensity incident onto an imaging system, while the second parameter, Q, describes the difference in intensity of the horizontal and vertical components of linear polarisation. With some derivation (see [11] for details), the third parameter, U, can be shown to describe the preference for $\pm\frac{\pi}{4}$ linear polarisation and the fourth parameter, V, can be shown to describe the preference for right-handed or left-handed circular polarisation.

This suggests that the Stokes parameters can be experimentally measured. Indeed, suppose the $y-$component is subject to a retardation ξ with respect to the $x-$component such that the phase difference between these components is increased by ξ radians, and the intensity $a(\theta, \xi)$ of the light vibrations which make an angle θ with the positive $x-$direction is measured. The Stokes parameters can be experimentally obtained by measuring the intensity obtained when using different values of θ and ξ

$$I = a(0^\circ, 0) + a(90^\circ, 0), \tag{2.13a}$$

$$Q = a(0^\circ, 0) - a(90^\circ, 0), \tag{2.13b}$$

$$U = a\left(45°, 0\right) - a\left(135°, 0\right), \tag{2.13c}$$

$$V = a\left(45°, \frac{\pi}{2}\right) - a\left(135°, \frac{\pi}{2}\right). \tag{2.13d}$$

Degree of Polarisation

As discussed above, waves can be considered to consist of both polarised and unpolarised components. While the intensity of the light, I, represents the combined intensity of both polarised and unpolarised parts, the Q, U, and V components represent only the polarised part. As such the inequality

$$I^2 \geq Q^2 + U^2 + V^2 \tag{2.14}$$

will only reduce to an equality for completely polarised light. It follows that the Stokes vector, $\vec{S}$, may be separated into unpolarised, and totally polarised components, $\vec{S}_{\text{un}}$ and $\vec{S}_{\text{pol}}$ respectively,

$$\vec{S} = \vec{S}_{\text{un}} + \vec{S}_{\text{pol}} \tag{2.15}$$

where

$$\vec{S} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix}, \quad \vec{S}_{\text{un}} = \begin{pmatrix} I - (Q^2 + U^2 + V^2)^{\frac{1}{2}} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{S}_{\text{pol}} = \begin{pmatrix} (Q^2 + U^2 + V^2)^{\frac{1}{2}} \\ Q \\ U \\ V \end{pmatrix}. \tag{2.16}$$

From here it is straightforward to define the degree of polarisation, P, which measures the ratio of the intensity of the totally polarised component to the intensity of the complete wave

$$P = \frac{(Q^2 + U^2 + V^2)^{\frac{1}{2}}}{I} \tag{2.17}$$

such that completely polarised light is described by $P = 1$, and unpolarised light by $P = 0$.

Degree of Linear Polarisation

A similar expression exists for the degree of linear polarisation, P_{linear}, which measures the ratio of the intensity of the linearly polarised component to the intensity

of the wave

$$P_{\text{linear}} = \frac{(Q^2 + U^2)^{\frac{1}{2}}}{I}. \tag{2.18}$$

In natural environments the circular polarisation component of light is negligible and is commonly not measured [12–16]. In such cases, where the assumption that $V \approx 0$ is valid, the degree of polarisation and the degree of linear polarisation are equivalent and so these terms are often interchanged.

Polarisation Angle and Ellipticity Angle

The azimuth angle of polarisation, ϕ, and ellipticity angle, ϵ, can also be obtained from the Stokes parameters

$$\phi = \frac{1}{2}\tan^{-1}\left(\frac{U}{Q}\right), \tag{2.19}$$

$$\epsilon = \frac{1}{2}\sin^{-1}\left(\frac{V}{(Q^2 + U^2 + V^2)^{\frac{1}{2}}}\right). \tag{2.20}$$

2.2.4 Mueller Matrices

When a light wave of a particular polarisation state interacts with an optical system, the polarisation state may change. For an input wave characterised by Stokes vector $\vec{S}_{\text{in}}$, and an output wave characterised by Stokes vector $\vec{S}_{\text{out}}$ the transformation can be represented in the form of a Mueller matrix, $M_{4\times4}$,

$$\vec{S}_{\text{out}} = M\vec{S}_{\text{in}}. \tag{2.21}$$

Tables of Mueller matrices for different retarding and polarising filters are listed in textbooks such as [11, Appendix A]. As an example, the Mueller matrix representing an ideal linear polariser oriented at an angle θ to the horizontal is

$$M_{\text{linear}}(\theta) = \frac{1}{2}\begin{pmatrix} 1 & \cos(2\theta) & \sin(2\theta) & 0 \\ \cos(2\theta) & \cos^2(2\theta) & \sin(2\theta)\cos(2\theta) & 0 \\ \sin(2\theta) & \sin(2\theta)\cos(2\theta) & \sin^2(2\theta) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.22}$$

which, for the special cases of $\theta = \{0°, 45°, 90°, 135°\}$ leads to the following Mueller

matrices

$$M_{\text{linear}}\left(0^\circ\right) = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2.23a}$$

$$M_{\text{linear}}\left(45^\circ\right) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2.23b}$$

$$M_{\text{linear}}\left(90^\circ\right) = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2.23c}$$

$$M_{\text{linear}}\left(135^\circ\right) = \frac{1}{2}\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.23d}$$

Similarly, a quarter wave plate with fast axis horizontal (0°) can be described with the following Mueller matrix

$$M_{\text{QWP}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{2.24}$$

Consider again the scenario in Figure 2.2 where initially unpolarised light is transmitted through a linear polariser, resulting in linearly polarised light, and then through a quarter wave plate, resulting in circularly polarised light. If the input unpolarised light can be described by the Stokes Vector $\vec{S}_{\text{in}} = (1, 0, 0, 0)^T$, and the and the linear polariser is oriented at 45° as described by (2.23b), the Stokes Vector of light output from the linear polariser can be calculated as

$$\vec{S}_{\text{out}} = M_{\text{linear}}\left(45^\circ\right)\vec{S}_{\text{in}} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}. \tag{2.25}$$

From (2.17) it is straightforward to calculate that the output degree of polarisation

$P = 1$, and from (2.19) the angle of polarisation is $\phi = 45°$ as expected.

If this linearly polarised light is then passed through a quarter wave plate, described by (2.24), the Stokes Vector of light output can again be calculated as

$$\vec{S}_{\text{out}} = M_{\text{QWP}}\vec{S}_{\text{in}} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}. \tag{2.26}$$

The light is still completely polarised with $P = 1$, but now the polarisation is completely contained within the Stokes V parameter, confirming that the output is indeed circularly polarised.

Mueller matrices can also be combined to represent complete optical systems. In order to describe the combination of linear polariser followed by quarter wave plate as represented in Figure 2.2, the Mueller matrices can be combined as

$$\begin{aligned} M_{\text{system}} &= M_{\text{QWP}}M_{\text{linear}}\,(45°) \\ &= \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 \end{pmatrix}. \end{aligned} \tag{2.27}$$

such that when applied to the same incident light, $\vec{S}_{\text{in}} = (1, 0, 0, 0)^T$, the same output as calculated in (2.26) is obtained.

Ellipsometry [11] is an experimental optical technique typically performed in order to gain information about an optical system which modifies the state of polarisation and experimentally determine the system's Mueller matrix. In the general scheme, a beam of light with known polarisation state is allowed to interact with an optical system, and the state of polarisation of the emergent beam is measured. By repeating this process with different initial polarisation states, the law of transformation of polarisation, and thus the Mueller matrix, for a particular optical system can be determined.

2.3 Blackbody Emission

All normal matter at temperatures above absolute zero will emit electromagnetic radiation due to the conversion of thermal energy to electromagnetic energy [9]. This conversion from thermal energy to electromagnetic energy is termed emission, while the reverse process where electromagnetic energy is converted to thermal energy

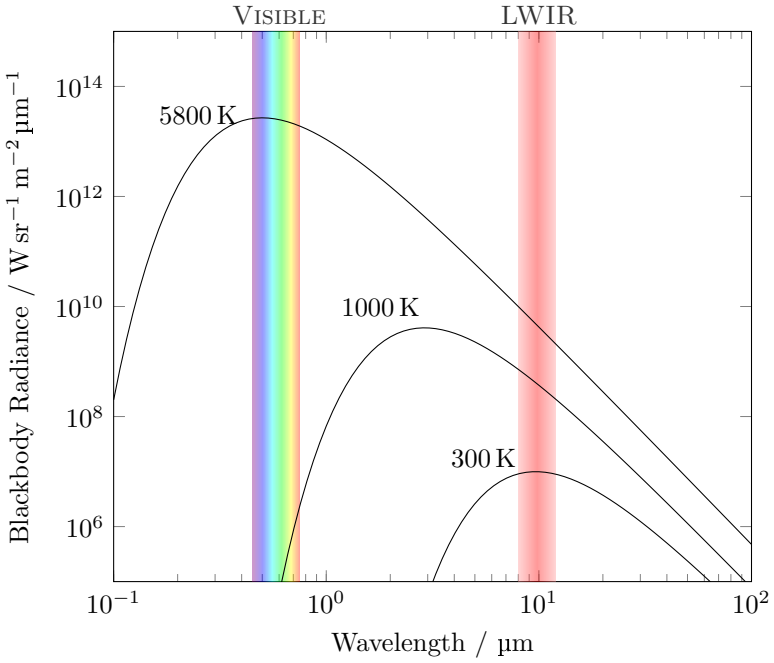


Figure 2.3: Blackbody radiance curves as a function of wavelength.

is termed absorption. A body which entirely absorbs all electromagnetic energy incident on its surface such that no energy is reflected is termed a blackbody.

The electromagnetic radiation emitted by a blackbody in thermal equilibrium with a known temperature, T, can be described by Planck's law, named after Max Karl Ernst Ludwig Planck who proposed it in 1900 [17, 18]. For a wavelength, λ, radiance from a blackbody at temperature T takes the form

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} \tag{2.28}$$

where B is a measure of spectral radiance, typically expressed in units of $\mathrm{W\,sr^{-1}\,m^{-2}\,\mu m^{-1}}$, $h \approx 6.626 \times 10^{-34}\,\mathrm{J\,s}$ the Planck constant, c the speed of light in the medium, and $k_B \approx 1.381 \times 10^{-23}\,\mathrm{J\,K^{-1}}$ the Boltzmann constant. Spectral radiance can be considered to be the power emitted from the surface per unit projected area of emitting surface, per unit solid angle, per unit wavelength.

Figure 2.3 shows blackbody radiance curves for blackbody objects at three different temperatures as a function of wavelength. The spectral radiance of the sun is often approximated by a blackbody at temperature 5800 K, notice the strong spectral radiance for an object at this temperature in the visible light area of the spectrum. The peak emission for blackbody objects at 300 K is located in the LWIR region of the spectrum, 8 µm to 12 µm, suggesting that imagery in this region will capture a combination of strong emission characteristics from objects in the scene and reflection of the emissions from other objects in the scene.

2.4 Polarisation on Reflection and Emission

It is fairly well known that the polarisation state of light may change upon reflection as a function of the angle of incidence [10], but the field of emission polarisation, where a surface emits light with a polarisation state related to the direction of observation was not explored until much later. Early qualitative experiments on emission polarisation were performed by Millikan in 1895 [19, 20] when he tabulated material classification against their emission polarisation strength. He found that smooth metals produced the strongest polarisation while non-metallic materials produced much weaker polarisation. The theory of emission polarisation can be derived from the four Maxwell equations, and in this section a brief overview will be given. For more extensive coverage of this topic, reference is given to the review on emission polarisation compiled by Sandus [21] and the papers of Bertilone [22] and Wolff *et al.* [23].

Materials can be characterised in terms of their emissivity, $\epsilon(T, \lambda, \theta)$, and absorption, $\alpha(T, \lambda, \theta)$ [23], where $\epsilon(T, \lambda, \theta)$ is the fraction of energy emitted compared with a perfect absorber and is in general a function of temperature, T, wavelength, λ, and emission angle, θ, and where $\alpha(T, \lambda, \theta)$ is the fraction of incident radiation absorbed and can be a function of the same parameters [9]. Kirchoff's law, which essentially asserts the law of conservation of energy, states that, for a surface in thermal equilibrium with a constant temperature, the energy absorbed at a particular wavelength at a point on the surface is equal to the emission. In general, Kirchoff's law holds for both orthogonal polarisation states independently [24] such that

$$\epsilon_s(T, \lambda, \theta) = \alpha_s(T, \lambda, \theta), \tag{2.29a}$$

$$\epsilon_p(T, \lambda, \theta) = \alpha_p(T, \lambda, \theta), \tag{2.29b}$$

where s and p denote two orthogonal polarisation states.[1]

2.4.1 Fresnel Equations

Figure 2.4 depicts a plane wave incident on an interface between two media 0 and 1 with complex indices of refraction $\tilde{n}_0$ and $\tilde{n}_1$ respectively. The indices of refraction

[1] $s-$ and $p-$polarisation states are two orthogonal linear polarisation states defined by a coordinate system relating to the plane made by the propagation direction of the light, and the surface normal vector of a reflective surface. The component of polarisation parallel to this plane is termed *p-like* (parallel), while the component of polarisation perpendicular to this plane is termed *s-like* (senkrecht, German for perpendicular).

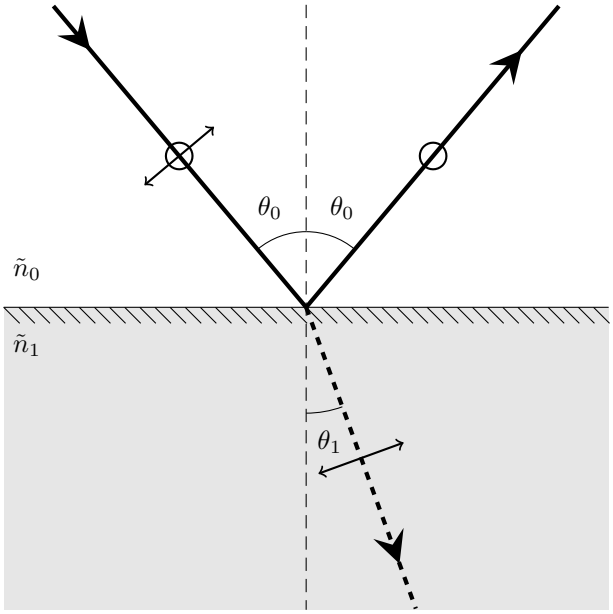


Figure 2.4: Reflection and transmission of a plane wave at the interface between two media 0 and 1. Angles θ_0 are the equivalent angles of incidence and reflection, and angle θ_1 is the angle of refraction. The symbols $\leftrightarrow$ and $\bigcirc$ indicate polarisation parallel and perpendicular to the incidence plane respectively.

are related to the angles of incidence θ_0 and refraction θ_1 through Snell's law

$$\tilde{n}_0 \sin\theta_0 = \tilde{n}_1 \sin\theta_1. \tag{2.30}$$

For a given amplitude and polarisation state of the incident wave, the amplitude of the reflected and transmitted waves can be determined by matching the tangential components of the waves across the interface. The ratio between the reflected and incident amplitudes leads to the Fresnel reflection coefficients (r) and the ratio between the transmitted and incident amplitudes leads to the Fresnel transmission coefficients (t) for each of the $s-$ and $p-$polarisations

$$r_p = \frac{\tilde{n}_1 \cos\theta_0 - \tilde{n}_0 \cos\theta_1}{\tilde{n}_1 \cos\theta_0 + \tilde{n}_0 \cos\theta_1}, \tag{2.31a}$$

$$r_s = \frac{\tilde{n}_0 \cos\theta_0 - \tilde{n}_1 \cos\theta_1}{\tilde{n}_0 \cos\theta_0 + \tilde{n}_1 \cos\theta_1}, \tag{2.31b}$$

$$t_p = \frac{2\tilde{n}_0 \cos\theta_0}{\tilde{n}_1 \cos\theta_0 + \tilde{n}_0 \cos\theta_1}, \tag{2.31c}$$

$$t_s = \frac{2\tilde{n}_0 \cos\theta_0}{\tilde{n}_0 \cos\theta_0 + \tilde{n}_1 \cos\theta_1}. \tag{2.31d}$$

The above equations relate the amplitudes of reflection and transmission for each polarisation, however, it is often more convenient to express the Fresnel equations

in terms of the reflectivity ρ and transmissivity τ

$$\rho_p = |r_p|^2 , \tag{2.32a}$$

$$\rho_s = |r_s|^2 , \tag{2.32b}$$

$$\tau_p = |t_p|^2 , \tag{2.32c}$$

$$\tau_s = |t_s|^2 , \tag{2.32d}$$

which describe the intensity of the waves, where intensity is proportional to the square of the amplitude.

By substituting (2.31) into (2.32), making the assumption that medium 0 is air ($\tilde{n}_0 \approx 1$), expanding the refractive index of medium 1 into its real and imaginary components ($\tilde{n}_1 = \tilde{n} = n + i\kappa$), and using Snell's law to replace θ_1, a more common form of the Fresnel equations can be derived for the reflectivities of both polarisations

$$\rho_s (\tilde{n}, \theta_0) = \frac{(a - \cos\theta_0)^2 + b^2}{(a + \cos\theta_0)^2 + b^2} \tag{2.33a}$$

$$\rho_p (\tilde{n}, \theta_0) = \frac{(a - \sin\theta_0 \tan\theta_0)^2 + b^2}{(a + \sin\theta_0 \tan\theta_0)^2 + b^2} \rho_s (\tilde{n}, \theta_0) \tag{2.33b}$$

where

$$2a^2 = \sqrt{(n^2 - \kappa^2 - \sin^2\theta_0)^2 + 4n^2\kappa^2} + \left(n^2 - \kappa^2 - \sin^2\theta_0\right) , \tag{2.34}$$

and

$$2b^2 = \sqrt{(n^2 - \kappa^2 - \sin^2\theta_0)^2 + 4n^2\kappa^2} - \left(n^2 - \kappa^2 - \sin^2\theta_0\right) . \tag{2.35}$$

These terms define the ratio of light energy incident on a smooth surface which is reflected in each polarisation state.

By conservation of energy, and assuming that the media is opaque such that any energy not reflected is absorbed into (and not transmitted through) the material, the ratio of energy absorbed is defined by $1 - \rho$. Consequently, for a body in thermal equilibrium, the emissivity in each polarisation state must be

$$\epsilon_s (\tilde{n}, \theta_i) = 1 - \rho_s (\tilde{n}, \theta_i) , \tag{2.36a}$$

$$\epsilon_p (\tilde{n}, \theta_i) = 1 - \rho_p (\tilde{n}, \theta_i) . \tag{2.36b}$$

The refractive indices of some common materials are presented in Table 2.1 and plots of (2.33a) and (2.33b) are presented for some of these materials in Figure 2.5. Notice that the $\rho_s > \rho_p$, and that the angle at which ρ_p is at a minimum is known as Brewster's angle.

The index of refraction is a complex number often split into its real and imaginary

Table 2.1: Table of the refractive indices of common materials

Material	$\tilde{n}$ at $\lambda = 589\,\text{nm}$	Reference
Hydrogen (0 °C)	1.0001	[25]
Oxygen	1.0003	[26, 27]
Water (25 °C)	1.3325	[28]
Sodium Chloride	1.5441	[29]
Diamond	2.4173	[30]
Aluminium	$1.0972 + 6.7943i$	[31]
Nickel	$1.9111 + 3.4725i$	[31]

parts as $\tilde{n} = n + i\kappa$. The real component of the index, n, determines how much light is bent, or refracted, on entering a material while the imaginary component, κ, is known as the extinction coefficient and determines to what extent light is absorbed when passing through a material. Notice from (2.33) that a larger value for κ leads to a smaller value for ρ, intuitively this confirms that if a material is able to absorb more light, less will be reflected, and by (2.36) more emission will be observed.

2.4.2 Observed Polarisation State

Consider the scene depicted in Figure 2.6 containing an object at temperature T_o located in a background with unpolarised emissivity and temperature T_b [32]. The total radiation, $L_\lambda(\theta)$, leaving the object in direction θ can be approximated as

$$L_\lambda^s(\theta) = B_\lambda(T_0)\,\epsilon_s(\theta) + B_\lambda(T_b)\,\rho_s(\theta), \tag{2.37a}$$

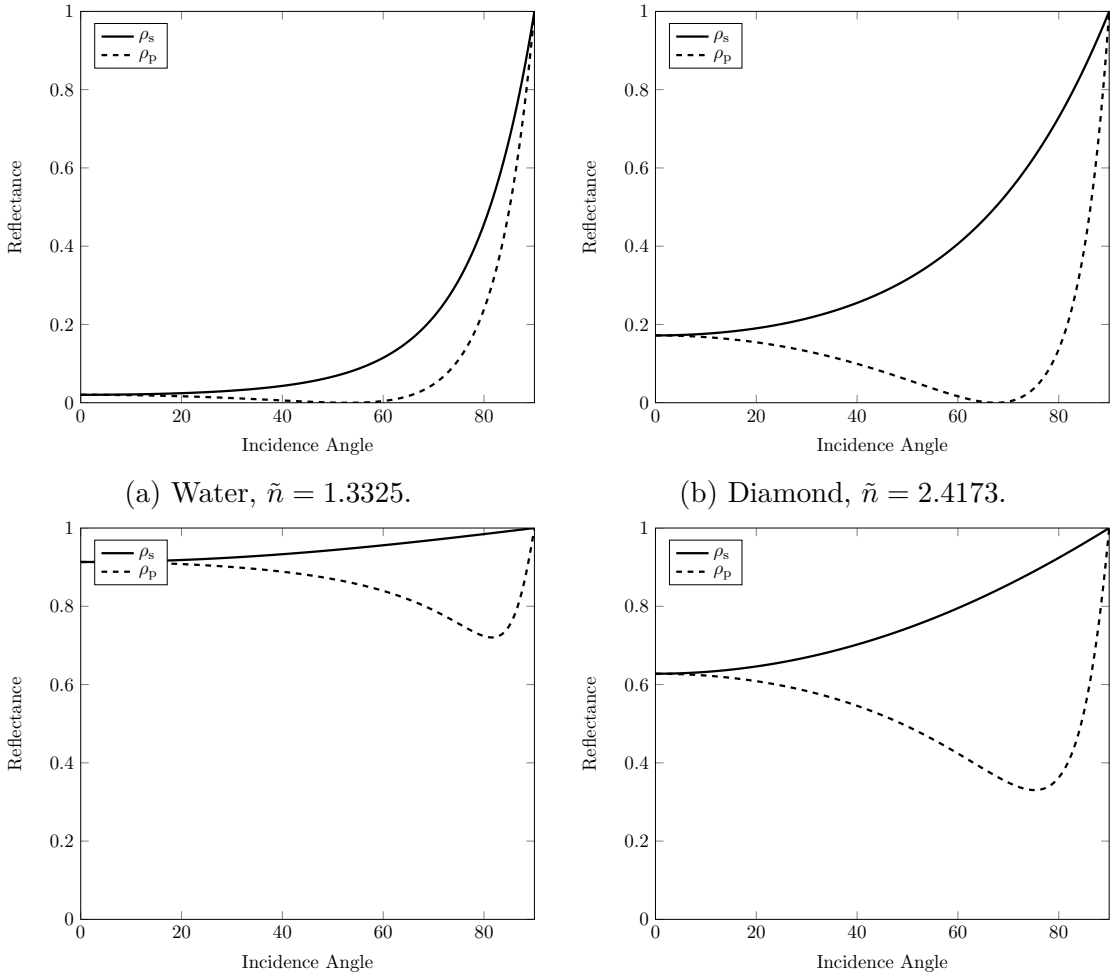
$$L_\lambda^p(\theta) = B_\lambda(T_0)\,\epsilon_p(\theta) + B_\lambda(T_b)\,\rho_p(\theta), \tag{2.37b}$$

where $B_\lambda(T)$ is the blackbody spectral radiance curve at temperature T as defined in (2.28). By substituting (2.36) into (2.37), this can be expressed in terms of the object and background temperatures, and the reflectivity as

$$\begin{aligned} L_\lambda^s(\theta) &= B_\lambda(T_0)\Big(1 - \rho_s(\theta)\Big) + B_\lambda(T_b)\,\rho_s(\theta) \\ &= B_\lambda(T_0) + \rho_s(\theta)\Big(B_\lambda(T_b) - B_\lambda(T_0)\Big), \end{aligned} \tag{2.38a}$$

$$\begin{aligned} L_\lambda^p(\theta) &= B_\lambda(T_0)\Big(1 - \rho_p(\theta)\Big) + B_\lambda(T_b)\,\rho_p(\theta) \\ &= B_\lambda(T_0) + \rho_p(\theta)\Big(B_\lambda(T_b) - B_\lambda(T_0)\Big). \end{aligned} \tag{2.38b}$$

Recalling from Figure 2.5 that $\rho_s > \rho_p$, from here it can be observed that three different regimes exist depending on the relative temperatures of the object and the optical background.



(a) Water, $\tilde{n} = 1.3325$.

(b) Diamond, $\tilde{n} = 2.4173$.

(c) Aluminium, $\tilde{n} = 1.0972 + 6.7943i$.

(d) Nickel, $\tilde{n} = 1.9111 + 3.4725i$.

Figure 2.5: Reflectances of difference materials. The $x-$axis plots the angle of observation from the surface normal (θ_0 in Figure 2.4), the $y-$axis plots the reflectance of both $p-$ and $s-$ polarisation components as calculated using (2.33).

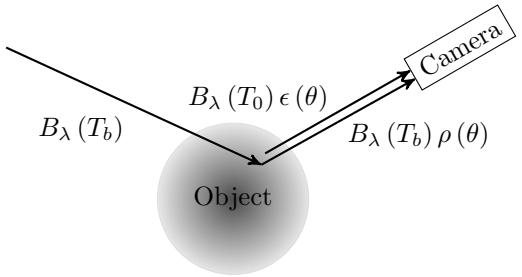


Figure 2.6: Image of a scene with a camera observing an object at temperature T_0 in a background characterised by a temperature T_b. Note that both emission and reflection are observed where reference has been made to the relevant terms of (2.37).

In the first case, when the temperature of the object is much greater than the temperature of the background ($T_o \gg T_b$), the dominant polarisation effects arise from thermally emitted radiation from the object and it is observed that $L_\lambda^s(\theta) \ll L_\lambda^p(\theta)$, and surfaces in this regime are expected to appear $p-$polarised. This is the dominant regime for the LWIR region when in general the cold sky is the optical background.

In the second case, when the temperature of the optical background is much greater than the temperature of the object ($T_o \ll T_b$), the measured polarisation state is largely due to reflections from the surface of the object and it is observed that $L_\lambda^s(\theta) \gg L_\lambda^p(\theta)$, and so surfaces in this regime are expected to appear $s-$polarised. This is the dominant regime for most applications in the visible and SWIR regions (0.4 µm to 2.5 µm).

For the third case, when the temperature of the object and the optical background are similar ($T_o \approx T_b$) it is observed that $L_\lambda^s(\theta) \approx L_\lambda^p(\theta)$, and thus no net polarisation effects will be observed.

The reasons why opposing effects are observed is relatively straightforward, and is simply because partial polarisation of the reflective component, and partial polarisation of the emissive component, are caused by complimentary processes (namely reflection and emission). Thus, the extent to which either polarisation state dominates (perpendicular to the incident plane for reflection, or parallel to the incident plane for emission) depends entirely on which is the dominant process. As both object and background are being approximated as blackbody sources, the relative temperature between object and background is a crucial parameter in determining the degree and state of polarisation.

2.4.3 Real-world Reflection and Emission

The above discussion makes the simplifying assumption that reflecting surfaces are perfectly flat. While this is a reasonable assumption for many surfaces such as glass or still water, surface roughness does impact on the Fresnel models. There are a number of mechanisms and models which take surface roughness into account, such as Torrance-Sparrow [33], Beard-Maxwell [34], Priest-Germer [35], as well as more modern extensions to specifically model how this impacts on the observed polarisation state [12, 13, 36, 37]. These will not be explored in detail here, but it is generally found that rougher surfaces are less polarising than smooth surfaces [36] due to effects such as secondary reflections [12].

2.5 Summary

This chapter started by introducing the wave theory of light. This included details about how the polarisation state is characterised in terms of the polarisation ellipse, and the Stokes vector – the experimental measurement of which, using a linear polarising element, and a quarter-wave plate, was described. Additionally, Mueller matrices were introduced and an example of their use in modelling transformations between two polarisation states was presented.

Following on from this, the concept of a blackbody was introduced and the Planck blackbody radiance equation, which models blackbody emission, was described. The Fresnel model of reflection and emission was then outlined, providing an explanation for why opposite polarisation states are observed for emitted and reflected light.

Chapter 3
Catherine MP LWIR Polarimeter

In this chapter, the Catherine MP LWIR polarimetric imager used in this work is discussed. Section 3.1 introduces the camera and explains the nature of the polarimetric measurements. Section 3.2 outlines the method of obtaining Stokes imagery from the output of the camera. Section 3.3 presents an experiment designed to test the linearity of the detector response to incident irradiance levels. Section 3.4 presents some results investigating the polarimetric sensitivity of the polarising elements which make up the detector. Results measuring the polarimetric response of different materials to different ratios of reflected and emissive radiation are presented in Section 3.5. Polarimetric imagery of outdoor scenes are presented in Section 3.6 along with a discussion on how best to display this multidimensional data. Finally, two datasets collected with the Catherine MP LWIR Polarimeter for vehicle detection are presented in Section 3.7.

Many of these investigations have previously been performed. As such this chapter does not claim to present novel analysis, but rather aims to introduce the sensor and explain the theory of what polarimetric imagery represents.

3.1 Thales QWIP Detector

Thales manufacture a number of cameras based on the Thales Research and Technology (TRT) 3rd generation gallium arsenide long wave Quantum Well Infrared Photodetector (QWIP) array [38] including the Catherine MP LWIR.

The Catherine MP LWIR, shown in Figure 3.1, uses an integrated detector cooler assembly manufactured by Softradir which comprises a 640×512, 20 µm pitch QWIP array, sensitive to long wave infrared radiation at wavelengths of 8 µm to 12 µm at a frame rate of 100 Hz. A QWIP detector will not detect radiation incident along the normal direction, and so two conventional QWIP detector consists of a 2-dimensional diffraction grating to deviate the radiation into angles where the quantum wells are sensitive [39]. The 2-dimensional grating, shown in Figure 3.2, is

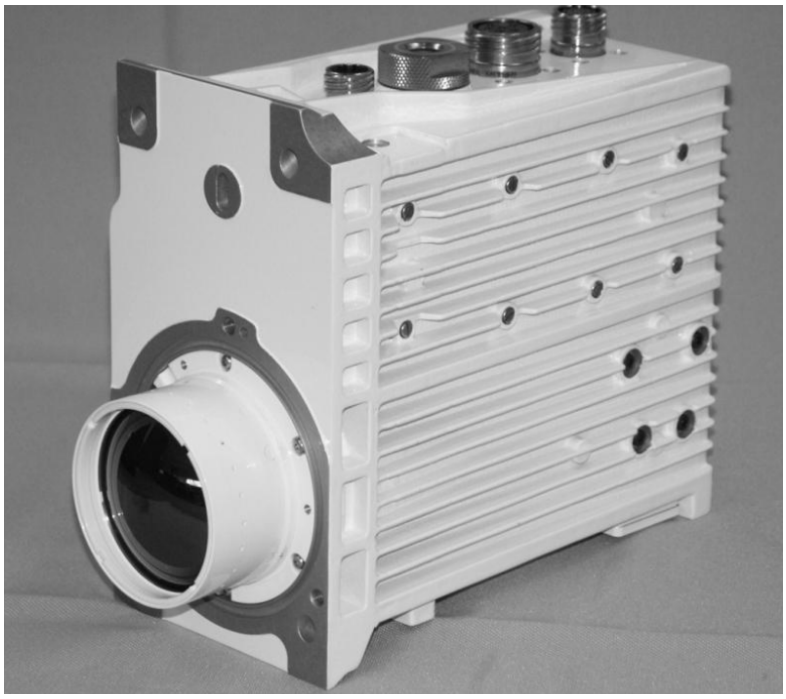


Figure 3.1: Image of the Catherine MP LWIR camera.

designed to refract light of any polarisation state in order to eliminate polarisation sensitivity. The Catherine MP LWIR Polarimeter, developed as part of the Electro-Magnetic Remote Sensing Defence Technology Centre by Thales in 2008 [40], uses an array of 1-dimensional diffraction gratings, shown in Figure 3.3, to re-introduce the polarisation sensitivity.

3.2 Stokes Images

The first three parameters of the Stokes Vector can be calculated, given the output of four linear polarisers oriented at 0°, 45°, 90° and 135°, using (2.13). The Catherine MP Polarimeter provides output in the format of a single 640×512 pixel image, with each set of 2×2 pixels containing the output of four linear polarising filters whose ordering is shown in Figure 3.4, where the direction of the lines indicate the direction of polarisation sensitivity.

The four images formed from each of the four filter orientations are termed i_a for the image formed when taking the pixels from the linear filters which pass polarisation at $a°$. This set of four images are slightly offset from each other, and so must be interpolated such that they are coincident. From these images, the Stokes images can then be calculated using equation (2.13). This process is described in [41] and summarised in Algorithm 3.1.

To measure the fourth Stokes component, V, which describes the degree (and direction) of circular polarisation, an additional quarter-wave plate is required as

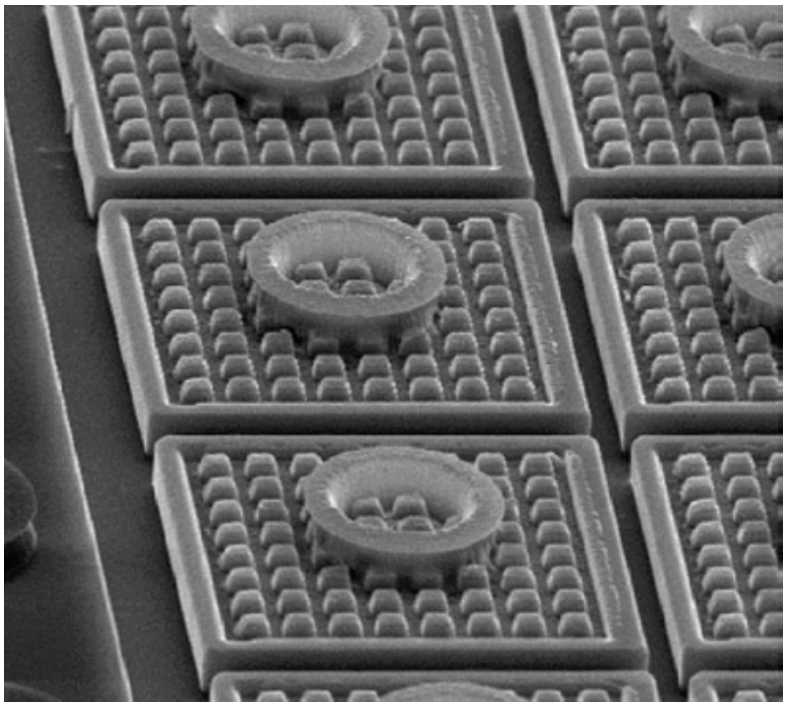


Figure 3.2: Electron micrograph image of a conventional 2-dimensional QWIP array. (Reproduced from [40, Figure 1].)

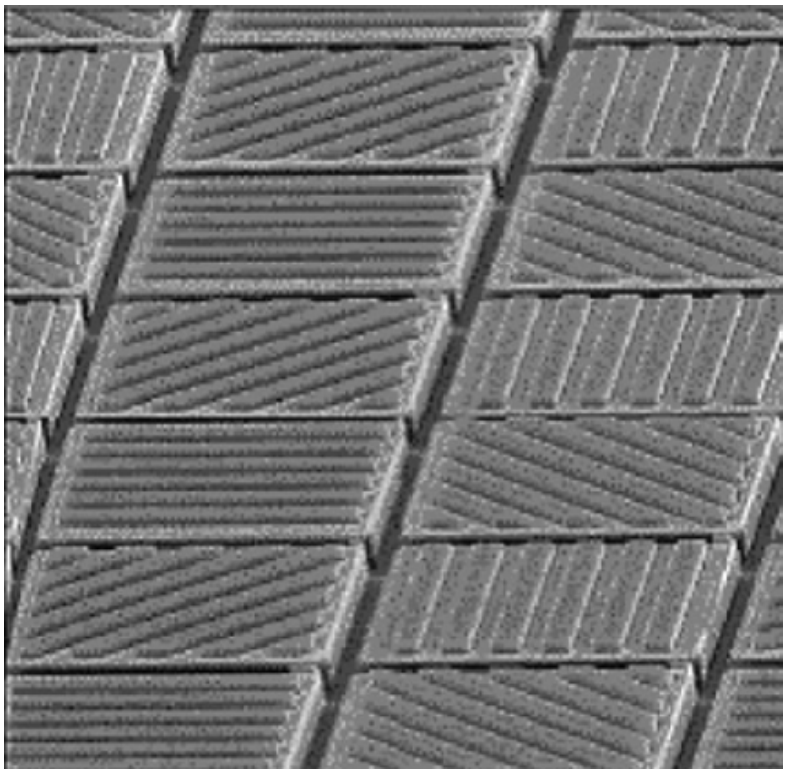


Figure 3.3: Electron micrograph image of a polarisation sensitive 1-dimensional QWIP array. (Reproduced from [40, Figure 4].)

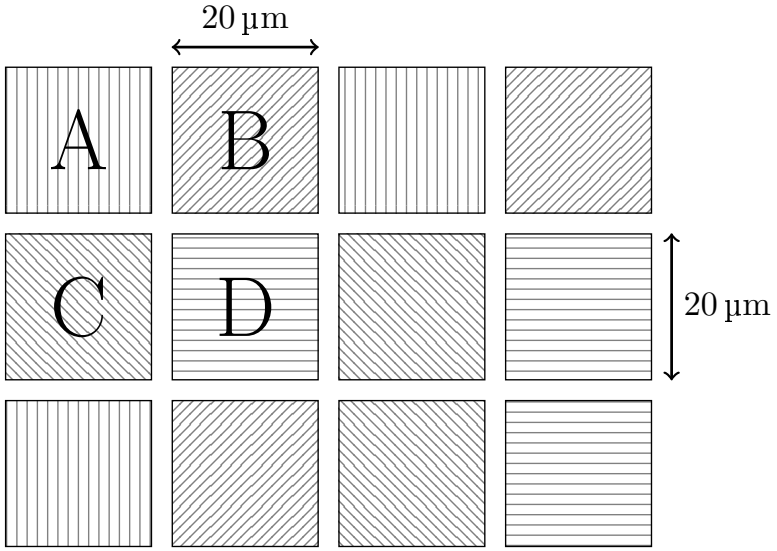


Figure 3.4: Layout of the polarisation-sensitive elements. Each element is 20 µm × 20 µm in extent, and the detector is made up of 640 × 512 elements. The direction of polarisation sensitivity is indicated by the direction of the grating in the diagram.

Algorithm 3.1 Obtain Stokes components from the Catherine MP Polarimeter

1: Acquire Frame (640 × 512) from camera.
2: Extract i_0, i_{45}, i_{90}, i_{135} (each 320 × 256).
3: Translate sub-images by $\frac{1}{4}$ pixel such that i_a are coincident.
4: Use registered i_a images to calculate Stokes images with (2.13).

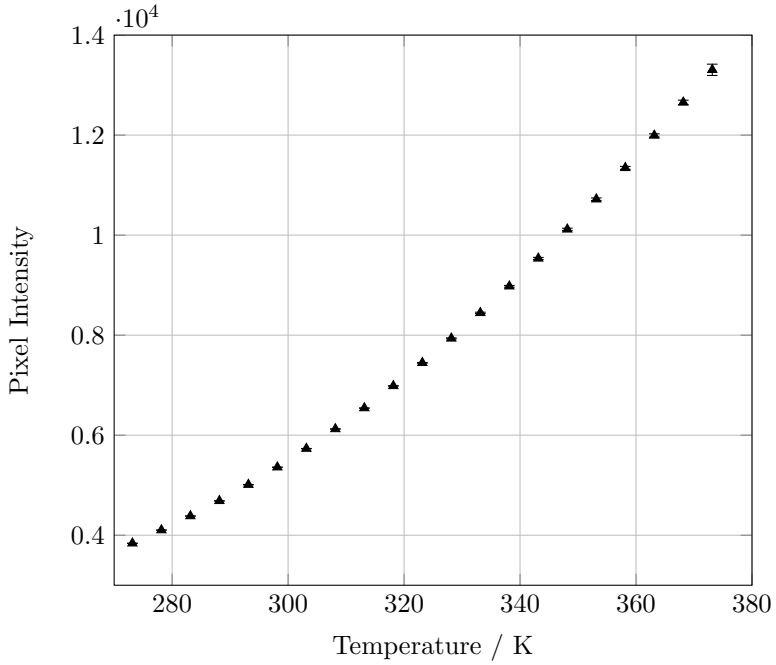


Figure 3.5: Plot of mean pixel intensity against blackbody temperature. The pixel intensity is the mean of all pixels in 100 frames, and the error bars are 1σ standard deviation.

per (2.13d). However, without a quarter wave plate, the Catherine MP can only measure the azimuthal angle of polarisation, ϕ (2.19), and not the ellipticity angle, ϵ (2.20).

3.3 Temperature-Intensity Relation

An experiment was designed to measure the raw detector response to a range of incident irradiance values. For reliable Stokes parameters to be obtained, the detector response should be linear within the range of expected incident irradiance.

A blackbody source (Electro Optical Ind.) was used to provide a reliable source of incident irradiance. The source consists of a black surface which is heated to a user-defined temperature and held constant such that the radiance emitted approximates that of a blackbody source at the set temperature. The Catherine MP camera was positioned such that the surface of the blackbody source filled the aperture of the camera.

The mean intensity value over all pixels and all polarisation directions over 100 frames was recorded for a range of temperatures (0 °C to 100 °C in 5 °C intervals) and is plotted in Figure 3.5.

The radiance emitted by the blackbody source at a wavelength λ, $B_{\lambda,\text{emitted}}$, is assumed to follow the Planckian distribution of (2.28). The spectral response of

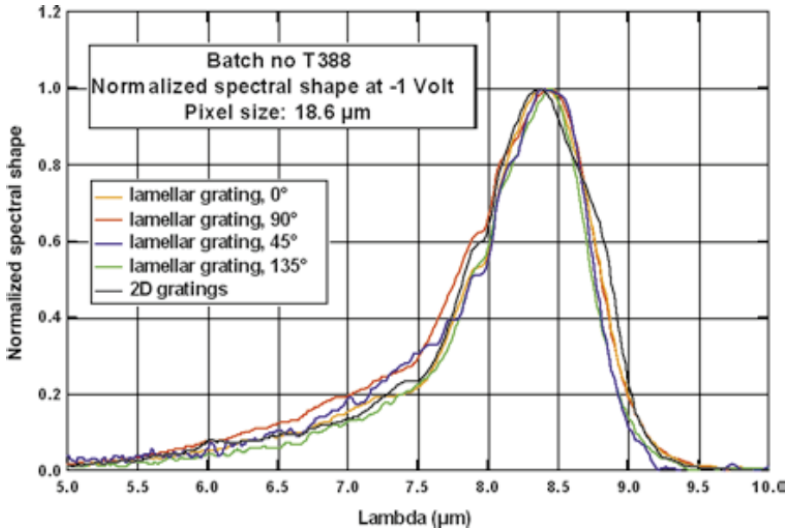


Figure 3.6: Spectral response, $R_{\text{detector}}(\lambda)$, of 1-D polarimetric QWIP detectors as a function of wavelength, λ. (Reproduced from [40, Figure 6].)

the detector, $R_{\text{detector}}(\lambda)$, has previously been measured [40] and is reproduced in Figure 3.6. Thus the radiance entering the detector can be calculated by integrating these two quantities

$$B_{\text{recorded}}(T) = \int B_{\lambda,\text{emitted}}(T)\, R_{\text{detector}}(\lambda)\, \mathrm{d}\lambda. \tag{3.1}$$

Using these relations the mean pixel intensity recorded is plotted against the theoretical irradiance recorded, $B_{\text{recorded}}(T)$, in Figure 3.7. The result confirms that the measured intensity is a linear function of the incident irradiance, with a coefficient of determination $R^2 = 0.9996$.

3.4 Polarisation Contrast

The responsivity of an element is a measure of the change in output recorded for a certain change in input. If the input can be characterised as a blackbody at two temperatures, T_1 and T_2, the responsivity, R, can be calculated from the difference in mean intensity, $I(T_i)$ recorded for each of these input temperatures

$$R = \frac{I(T_2) - I(T_1)}{B(T_2) - B(T_1)} \tag{3.2}$$

where $B(T)$ is the radiance observed with this detector for a blackbody source at temperature T and is calculated from (3.1).

Consider fixing an external polarising filter in order to completely polarise the input light to the sensor. The responsivity for a particular polarising element will be at a maximum, $R_{\max}$, when the incident light is linearly polarised with the same

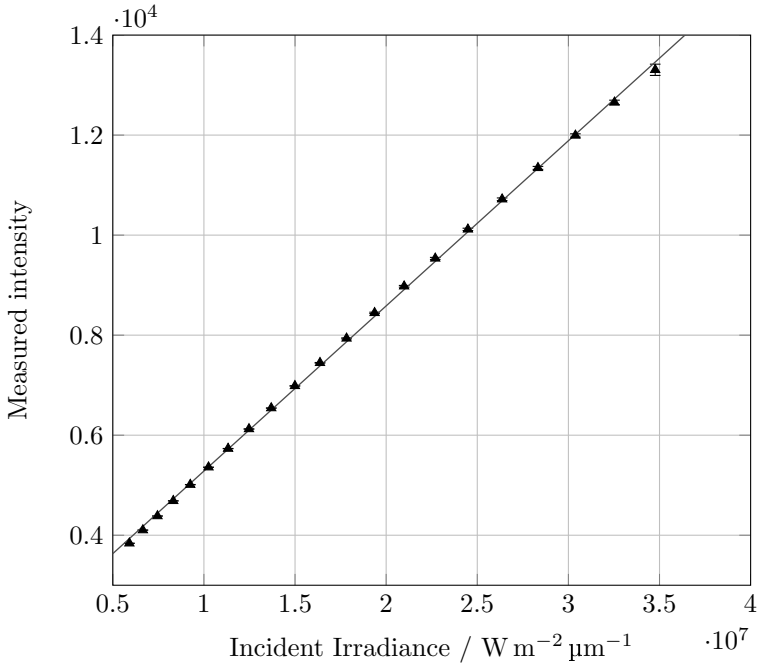


Figure 3.7: Plot of mean pixel intensity against incident irradiance. The pixel intensity is the mean of all pixels in 100 frames, and the error bars are 1σ standard deviation. The incident irradiance is calculated as the integral over wavelength of the product of Planck's blackbody distribution and the spectral response of the detector.

angle as the polarising element, and a minimum, $R_{\min}$, when incident light has a perpendicular angle of polarisation. The contrast, C, is then defined as

$$C = \frac{R_{\max} - R_{\min}}{R_{\max}} = \frac{2A}{\mu_R + A} \tag{3.3}$$

where A is the responsivity amplitude, and μ_R is the mean responsivity. The contrast can be considered to be a form of signal to noise ratio for polarimetry.

As shown in Figure 3.3 the detector array consists sets of four interleaved sets of linear polarising elements with different orientations. The horizontal and vertical (hv) polarising elements are expected to have similar contrast to each other, as they are essentially identical elements in two different orientations, and the diagonal (d) polarising elements were also expected to have similar contrast to each other for the same reason. However, the hv elements are expected to have a different contrast to the d elements due to the different grating pattern. With reference to Figure 3.3 note that the hv elements consist of 6 ridges of uniform length, while the d elements consist of 9 ridges of varying length.

To determine the contrast of the different polarising elements a Medway Optics ZnSe polariser (considered to be a perfect linear polarising filter) was mounted in a rotary stage in front of the sensor. A blackbody source was positioned in front of the

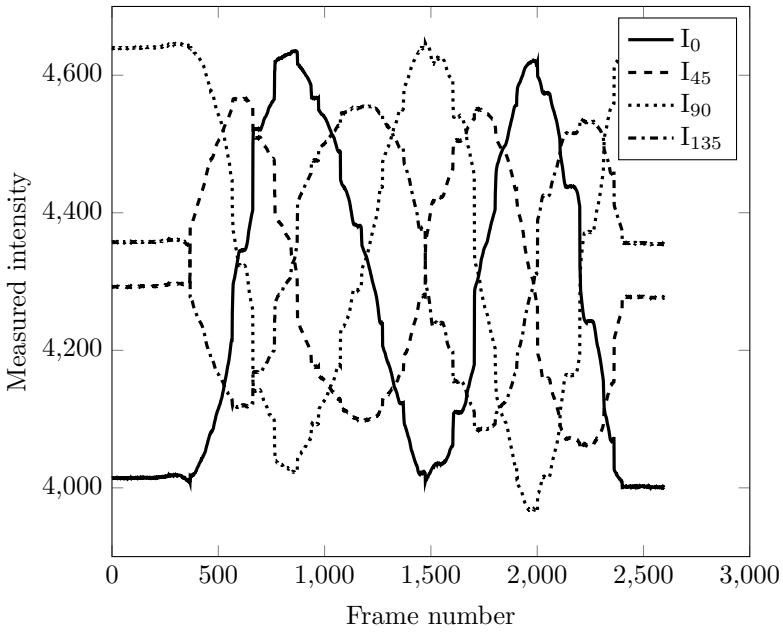


Figure 3.8: Plot of mean intensity recorded on each of the four polarising directions against frame number as a linear polariser is rotated in front of the camera.

Table 3.1: Polarisation contrast values for each linear polarising element as measured in [40]. Labels A, B, C, D are as in Figure 3.4.

Orientation	Contrast
A	0.52
B	0.41
C	0.41
D	0.53

sensor such that it filled the aperture. As an initial test the blackbody source was set to a constant temperature of 35 °C and the polariser was freely rotated. A plot of the measured intensity for each of the four polarising filter directions, averaged over similar pixels within that frame, against frame number (time) is shown in Figure 3.8. This shows a set of four sinusoids, one for each polarising element, each offset by 45° in phase as expected.

To determine the contrast, this procedure should be repeated for a second blackbody temperature. However, it may initially be observed that the I_0 and I_{90} plots show a larger amplitude than the I_{45} and I_{135} elements. A similar result was noted by Parsons and Craig [40] who determined the polarisation contrasts for each element to be as shown in Table 3.1. Due to damage to the external polarising filter this result could not be reproduced in the current work.

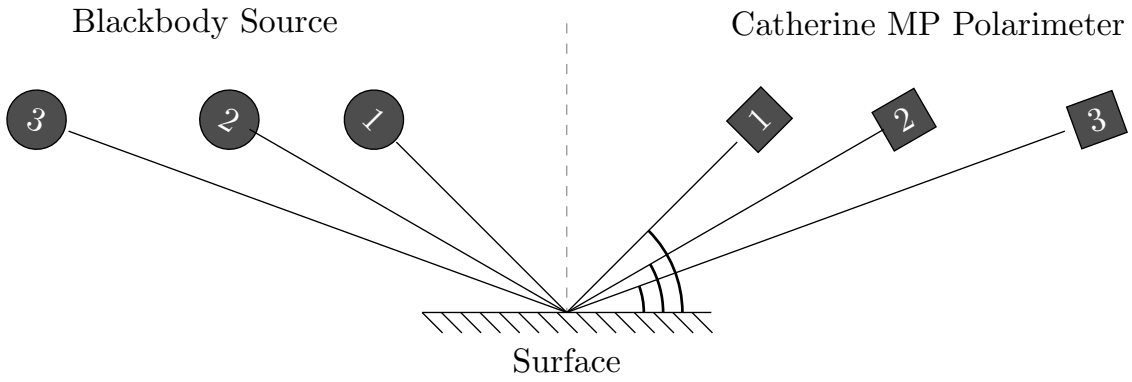


Figure 3.9: A series of different materials were imaged from three different viewing angles with a blackbody source in the optical background. The position of the materials were fixed as indicated by the Surface marked in the figure, while the position of the polarimeter and blackbody were moved to three positions (indicated 1, 2 and 3) such that polarisation images at three angles of incidence could be measured.

3.5 Response of Different Materials

As discussed previously in Section 2.4 the degree of polarisation is expected to vary as a function of viewing angle, with low degree of polarisation for observations close to the surface normal and increasing degree of polarisation for angles approaching the grazing angle. Also, the difference in temperature between the object and the optical background is expected to affect the polarisation state. Specifically opposite polarisation states are expected for emission dominant and reflection dominant regimes, and additionally different degrees of polarisation are expected for different materials (with different complex refractive indices). An experiment was designed to analyse some of these effects by recording the Stokes parameters observed for different material types, and for different angles of observation.

Four samples were chosen for this investigation: a wooden particle board, a kitchen tile, and a metal plate with two different paint coatings. As shown in Figure 3.9 each of these materials was mounted on a stationary table and observed at three different viewing angles qualitatively termed close (1), medium (2) and far (3) where the numbers in parenthesis correspond to the labels in Figure 3.9. A blackbody source was positioned in the optical background.

The ambient temperature of the materials was measured to be in the range 16 °C to 17 °C and data was collected for each of the samples with the blackbody source set at 0 °C and 50 °C, allowing both emissive and reflective effects to dominate respectively. The vertical axis of the camera was approximately aligned with the plane of incidence in order to allow the i_0 and i_{90} images to align with the expected $s-$ and $p-$polarisation components. This allows the sign of the Q parameter to

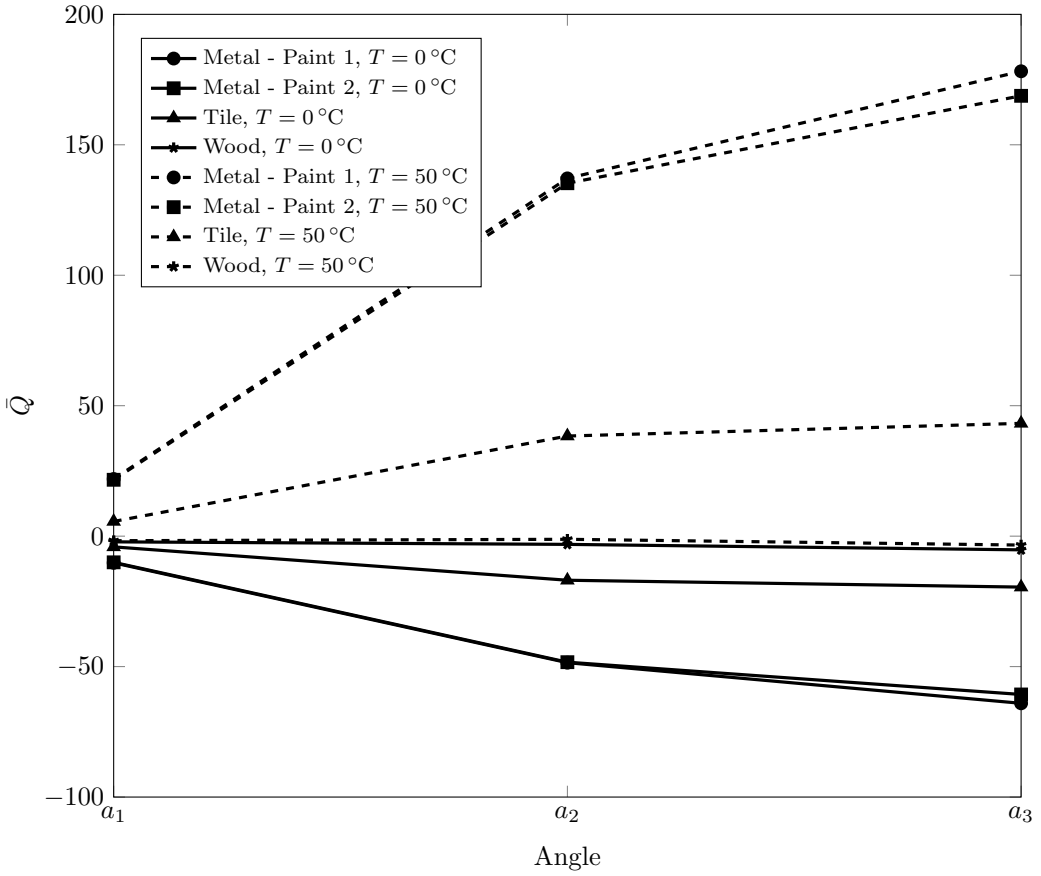


Figure 3.10: Plot of the mean Stokes Q component in specular region of each material at a range of observation angles. The arbitrary x-axis is relative to the angle of incidence, from a_1 which approaches the surface normal, to a_3 which approaches the grazing angle.

indicate whether a reflective (positive) or emissive (negative) characteristic is being observed. For each material, the mean Stokes Q values for pixels within a region of interest containing the specular reflection from the blackbody source were calculated over 100 frames and are plotted for each of the three observation angles in Figure 3.10.

There was significant specular reflection from the metal, allowing the specular region to be easily defined, but the regions on the tile and wood were less distinct, making these more difficult to define and so potentially less accurate.

3.5.1 Emission and Reflection

For all cases it is clearly apparent that with the blackbody source set to 0 °C (the lower solid curves in Figure 3.10), the Stokes Q parameter decreases with angle while when the blackbody source is set to 50 °C (the upper dashed curves in Figure 3.10), the Stokes Q parameter increases. When the blackbody source is at 0 °C, the emission from the object ($T \approx 16\,°C$) dominates the observation, while when

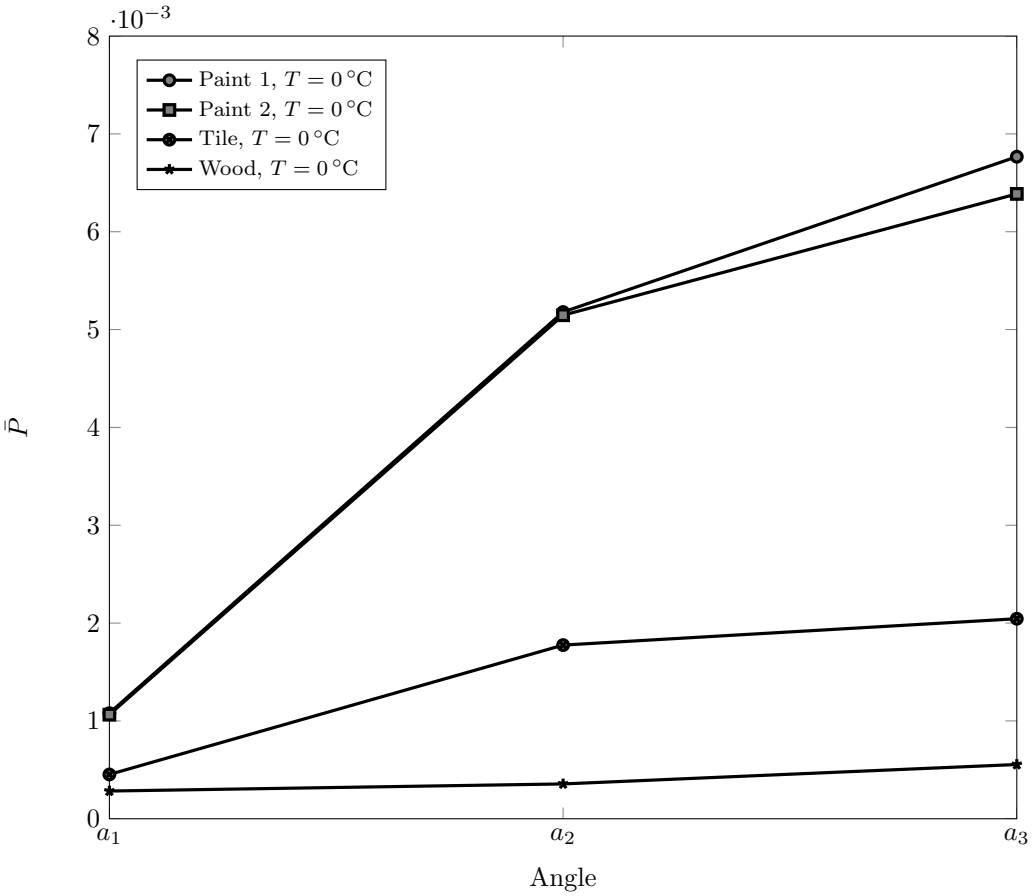


Figure 3.11: Plot of the Degree of Polarisation component, P, in specular region of each material at a range of observation angles. The arbitrary x-axis is relative to the angle of incidence, from a_1 which approaches the surface normal, to a_3 which approaches the grazing angle.

the blackbody source is at 50 °C, the reflective effect dominates.

Equation (2.38) suggests that when emission dominates, the material will exhibit a $p-$polarisation state. For a surface on the ground plane this is termed a vertical polarisation and is defined as 90° thus explaining the negative $Q = i_0 - i_{90}$ observed. Similarly, when the reflective effects dominate, an $s-$polarisation state is expected, or for a surface on the ground plane a horizontal polarisation (0°), explaining the positive Q component observed.

3.5.2 Degree of Polarisation

The degree of polarisation (2.18) for the observations with the blackbody source set to 0 °C is shown in Figure 3.11. The figure shows that the degree of polarisation increases with angle of observation, which is to be expected.

3.5.3 Material Differences

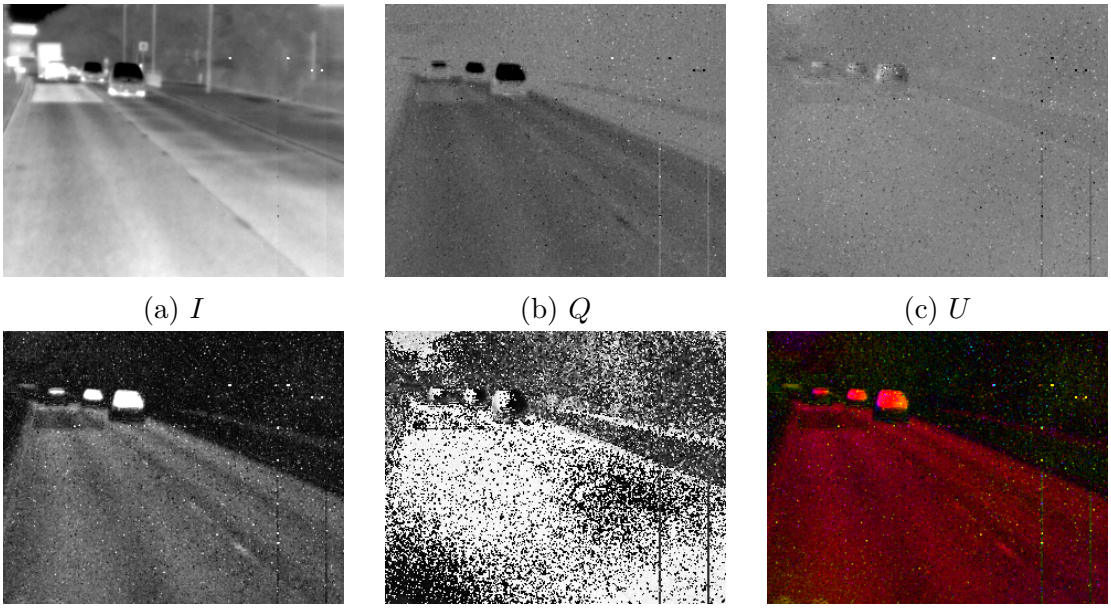
Finally, it is clear that the degree of polarisation observed at different angles and at different temperatures is also dependent on the nature of the material being observed. For example, the metal sheet with both types of paint coating produces the strongest polarisation, while the wooden board produces a negligible polarisation effect under the same conditions, with the tile producing an intermediate effect. However, note that the two different paint coatings on the same metal sheet have similar results, suggesting that these coatings have little, or similar effect on the properties of the metal sheet.

Predicting and modelling the polarimetric appearance of different materials is beyond the scope of this work. However, with reference to the Fresnel equations (2.33) the principal parameters which affect the polarisation state include the refractive index of the material, the observation angle, and the difference between emissive and reflective radiance. In addition, surface roughness is also an important parameter which can be taken into account by, for example, making the assumption that the majority of surface roughness occurs on a scale larger than the wavelength of light scattered from it, and that a rough surface may be approximated as the superposition of small "mirror-like" surfaces (micro-facets) oriented at a distribution of angles around the surface normal. This type of model has been used for example by Gartley [37] to consider the effect of surface roughness on the polarisation state observed in the LWIR. Typically it is found that the greater the surface roughness, the lower the degree of polarisation observed, as the polarisation effects from different micro-facets begin to cancel out [36].

3.6 Polarisation images

To this point the output from the polarimeter has been discussed in terms of the values of the Stokes components, however the array of detectors allows the visualisation of these components as an image. Figure 3.12 shows an example of a scene viewed with the polarimeter. The false colour representation shown in Figure 3.12f is formed in the HSV colour space with the angle of polarisation mapped to the hue, the saturation fixed at maximum, and the degree of polarisation mapped to value; this leads to areas of higher polarisation having a higher brightness in the image, and the specific colour relating to the angle of polarisation. Note that the incident intensity (I) is not encoded in this representation. Figure 3.13 shows a colour wheel which outlines the range of colours used for different degree and angles of polarisation. Further example scenes are shown in Figures 3.14 and 3.15.

The attractiveness of using the polarimetric information in addition to the in-



(a) I (b) Q (c) U

(d) P (e) ϕ (f) Polarimetric

Figure 3.12: Example scene showing images of (a) Stokes I, (b) Stokes Q, (c) Stokes U, (d) degree of polarisation P, (e) angle of polarisation ϕ, (f) false colour representation.

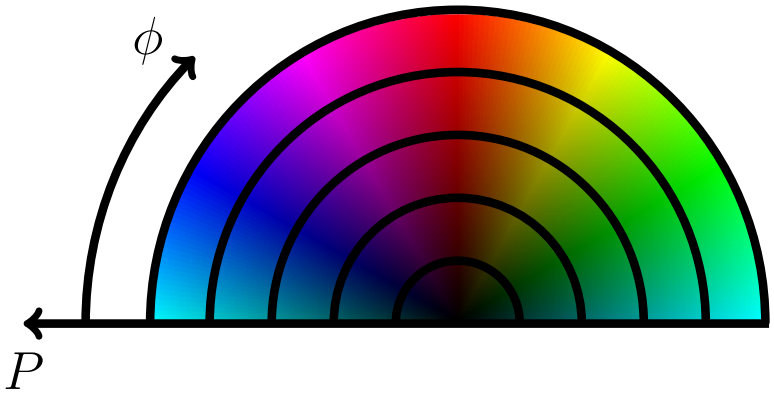
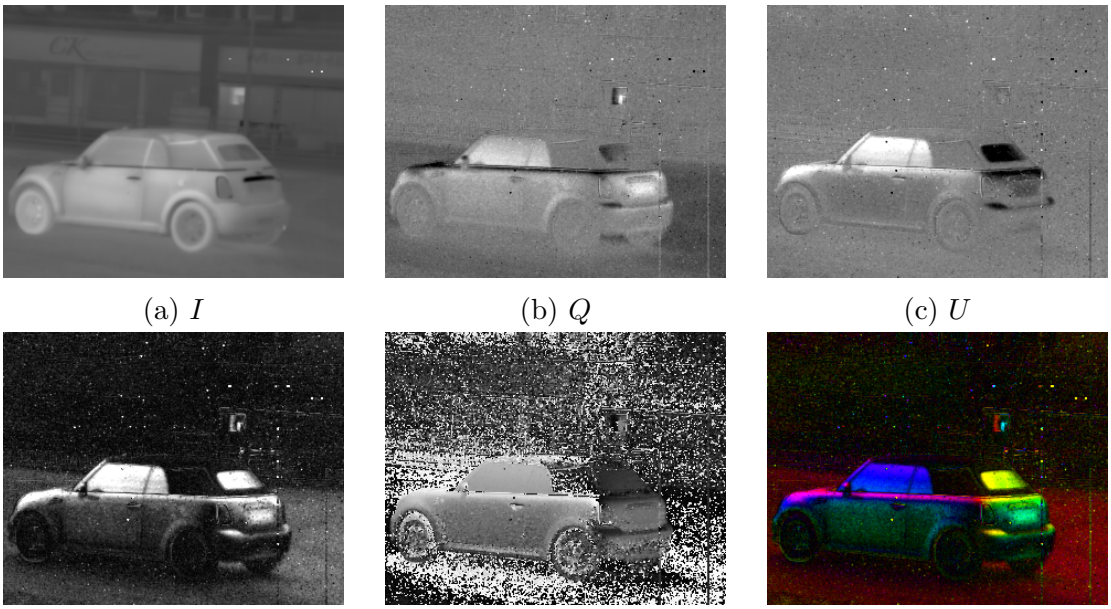


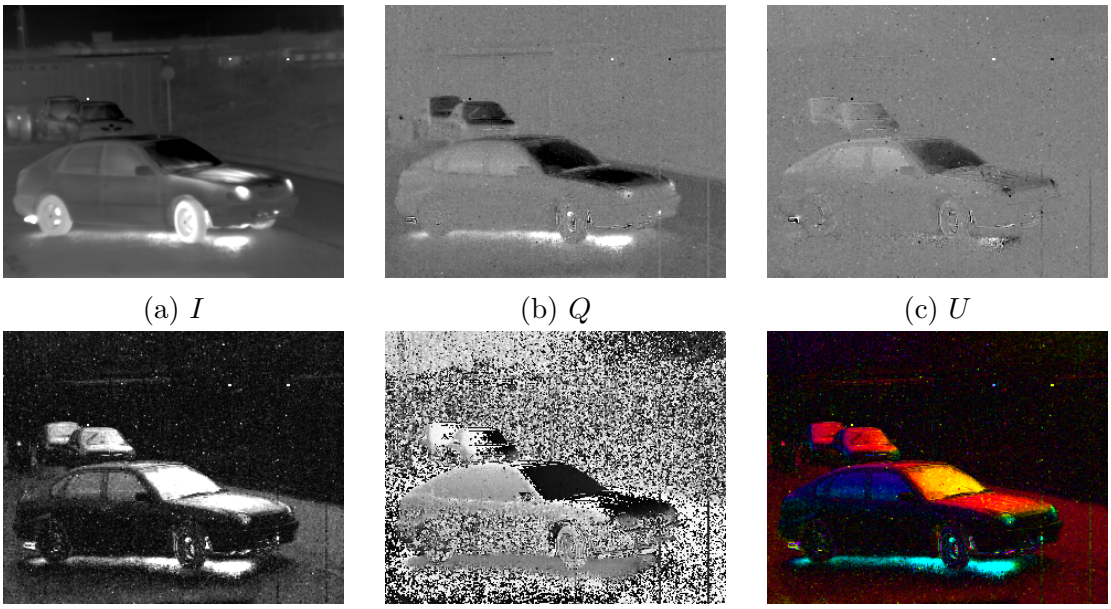
Figure 3.13: Colour wheel detailing the false colour representation of the polarimetric images. The degree of polarisation determines the HSV value (with high degree of polarisation pixels having a more intense colour), while the angle of polarisation is mapped to the hue. For example, strong vertical polarisation would appear red, while strong horizontal polarisation would appear light blue.



(a) I (b) Q (c) U

(d) P (e) ϕ (f) Polarimetric

Figure 3.14: Example scene showing images of (a) Stokes I, (b) Stokes Q, (c) Stokes U, (d) degree of polarisation P, (e) angle of polarisation ϕ, (f) and false colour representation.



(a) I (b) Q (c) U

(d) P (e) ϕ (f) Polarimetric

Figure 3.15: Example scene showing images of (a) Stokes I, (b) Stokes Q, (c) Stokes U, (d) degree of polarisation P, (e) angle of polarisation ϕ, (f) and false colour representation.

tensity is immediately obvious. From Figure 3.12 observe the low but consistent vertical polarisation from the tarmac surface of the road, additionally the strong signature from the vehicles, particularly in the Q and P images. In Figure 3.12 the vehicle surfaces being observed are upper surfaces (bonnet, windscreen, roof, or upper boot depending on orientation), and have a strong vertical polarimetric signature. In Figures 3.14 and 3.15 vehicles are present at a closer range, revealing more detailed structure of the vehicles being observed. Observe that the different materials polarise to differing degrees, with the metal sheets of the vehicle providing a large degree of polarisation. Additionally the orientation of the surfaces can be inferred from the angle of polarisation. Therefore, measurement of the polarisation state can also allow one to obtain three dimensional shape [7, 42, 43].

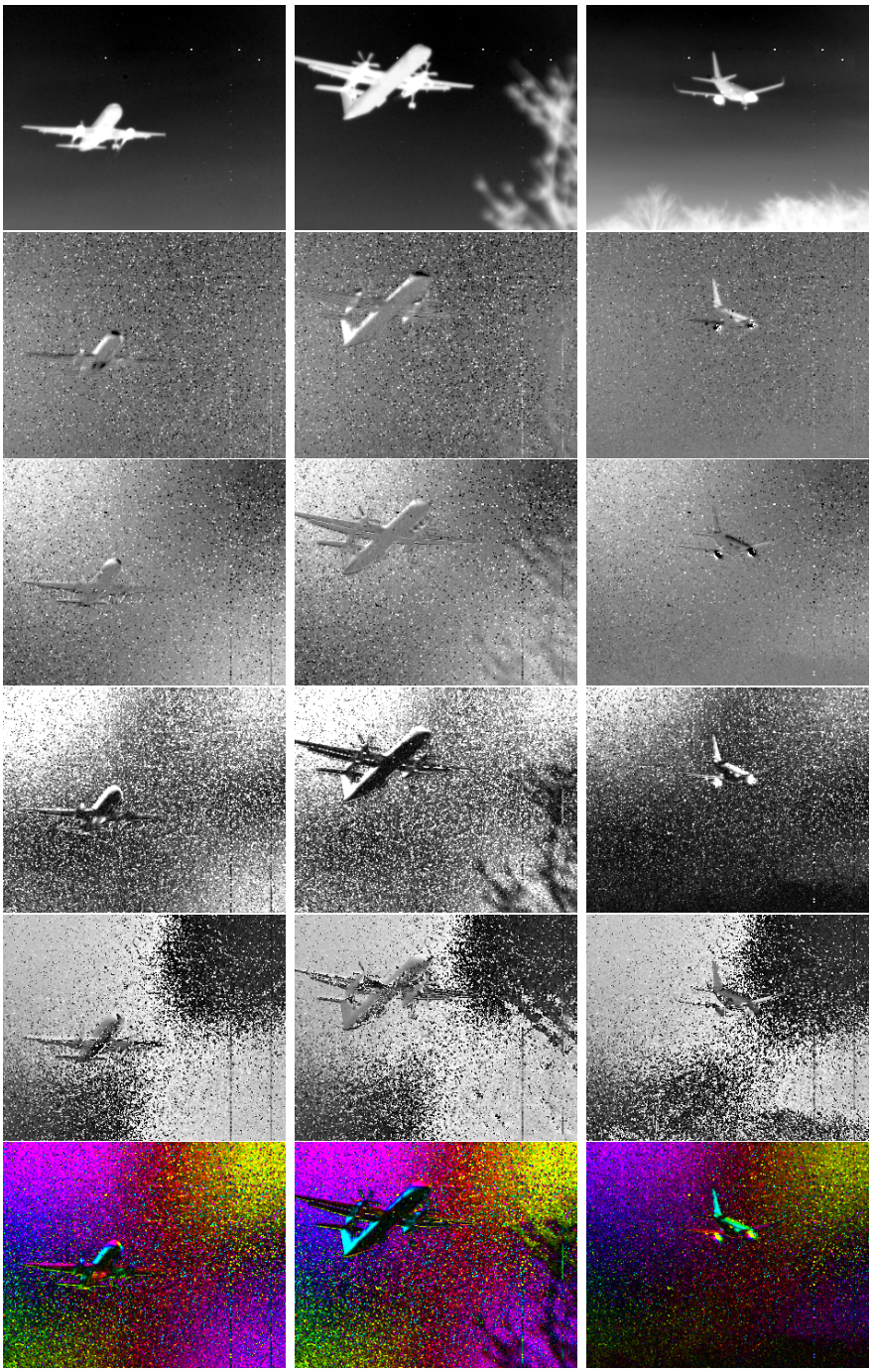
As a reminder, when the surface being observed is principally emissive (that is to say, if blackbody approximations are made, the temperature of the observed surface is larger than that of the optical background), then a preference for $p-$polarisation will be observed, that is parallel to the surface normal vector. For the majority of the strong polarisation observed, the optical background is the sky, which is often termed cold sky, and so most observations are emissive.

However, Figure 3.15 shows a strong horizontal polarisation in the region of road directly below the vehicle. In this case, the temperature of the vehicle is higher than the temperature of the road and so in this region reflection dominates the polarimetric effects. This leads to a preference for $s-$polarisation, which explains the horizontal angle of polarisation for this region of road.

As an additional example of imagery captured with the Catherine MP LWIR Polarimeter, Figure 3.16 shows planes taking off from or landing at Glasgow airport. It can be seen that the upper surfaces of the planes captured in these scenes follow a similar pattern to ground based vehicles in that the angle of polarisation is related to the surface normal. Additionally, the angles observed confirm that the upper surfaces are exhibiting $p-$polarisation, indicating an emissive characteristic. The underside of the planes shows very low degree of polarisation, likely due to the warmer optical background (the ground) for the lower section of the plane meaning the opposite effects of polarisation from emission and reflection start to cancel out. There also seems to be a partial polarimetric signature from the background sky.

3.7 Vehicle Detection Data Sets

The Catherine MP Polarimeter was used to collect two polarimetric LWIR datasets for vehicle detection in order to train and test detection algorithms introduced later. The first dataset, DS1, was collected from a static location outside a car park, which



(a) Scene 1 (b) Scene 2 (c) Scene 3

Figure 3.16: Example images of aircraft captured with the Catherine MP LWIR Polarimeter. From top to bottom, the I, Q, U, P, ϕ and coloured Polarimetric image are shown. The Polarimetric image contains (P, ϕ) information as described in Fig. 3.13.

captured vehicles leaving and arriving from a number of camera orientations. The second dataset, DS2, was captured with the sensor mounted to the roof of the trials van whilst driving through Glasgow. The sequences in DS2 feature a greater range of vehicle poses, distances, and backgrounds and is therefore considered a harder dataset.

The ground truth locations of all vehicles in a subset of each dataset were manually annotated using the Video Annotation Tool from Irvine, California (VATIC) [44]. This tool splits a video into small sections and allows a number of workers to annotate ground truth bounding boxes to each section independently. For the annotation of DS1 and DS2, the author was the only worker and so the multi-user functionality was not utilised.

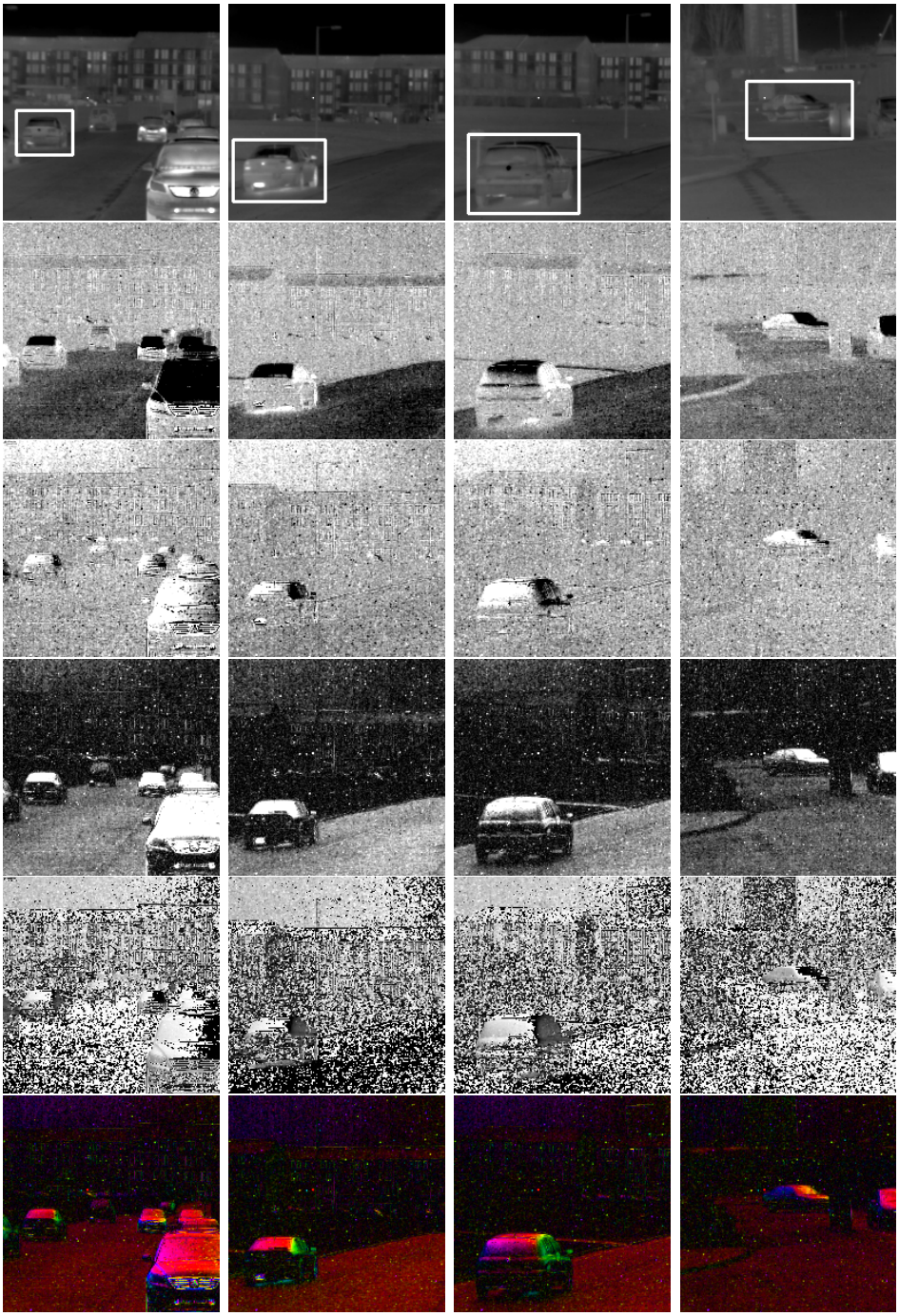
More details on the collection of DS1 and DS2, and some example imagery with ground truth annotation from both, are presented below.

3.7.1 Data Set 1

Data set 1, DS1, was collected with the Catherine MP LWIR polarimeter over a two day period. Vehicles were recorded leaving the car park of Thales UK, Glasgow between 4pm and 5pm on Thursday 14th March 2013 (day one), and arriving at the car park between 7am and 8.30am on Friday 15th March 2013 (day two). The weather on day one was cloudy with some light rain, while the weather on day two was heavy rain initially, drying as the morning progressed.

The two times of day were chosen to maximise the number of vehicles (morning/afternoon commute) and to capture changes in vehicle appearance due to the temperature difference - those leaving the car park had recently been started while those arriving had been running for a longer period of time resulting in a subtly different appearance with hotter tyres, and engine. The camera position was also varied using a pan and tilt mechanism while situated in a position on the side of the road to allow the collection of vehicles at a number of different orientations.

Data was collected at 100 Hz, producing a total of more than 180,000 frames. From this 691 frames collected on day one containing either a unique vehicle or a vehicle in a significantly different position were extracted and manually annotated with a bounding box around each vehicle present. This provided a total of 1,427 ground truth vehicle bounding boxes which were randomly partitioned into 1,295/132 positive training/testing examples respectively. Some example frames from the different viewpoints, along with their ground truth annotations, are shown in Figure 3.17.



(a) Frame 1 (b) Frame 2 (c) Frame 3 (d) Frame 4

Figure 3.17: Example frames from DS1 with one ground truth bounding box marked on the I image, note that all vehicles in each frame are annotated in the dataset. From top to bottom, the I, Q, U, P, ϕ and coloured Polarimetric image are shown. The Polarimetric image contains (P, ϕ) information as described in Fig. 3.13.

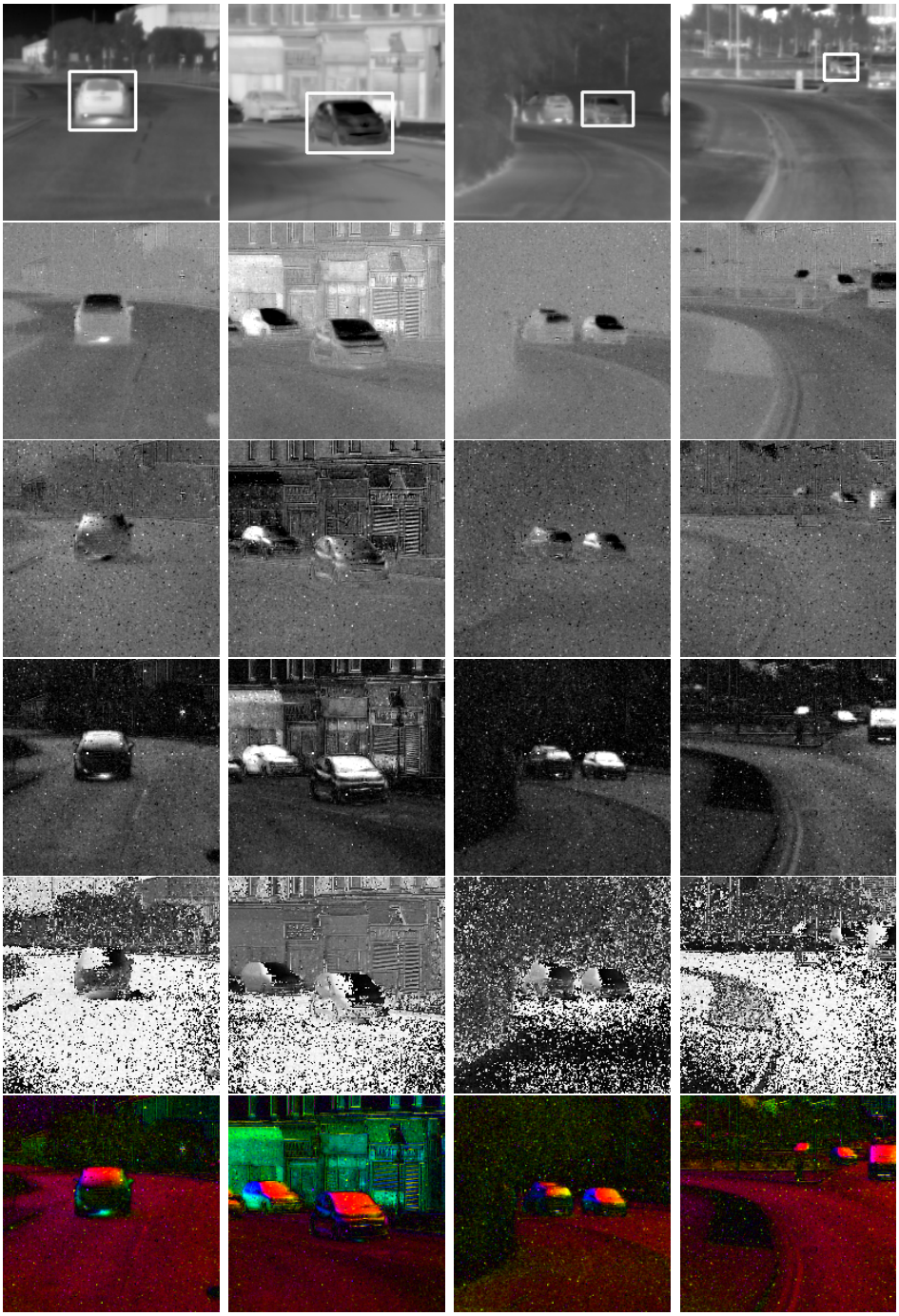
3.7.2 Data Set 2

Data set 2, DS2, was also collected with the Catherine MP LWIR polarimeter over a two day period. With the camera mounted in a forward-facing orientation on the trials van, static and moving vehicles were recorded while travelling through Glasgow on the 24th September 2013 (day one) and 1st October 2013 (day two). The weather on day one was overcast with light rain, and the weather on day two was sunny with a clear sky. The collected data contains examples of a number of different vehicle types, a variety of poses, some occluded vehicles, and vehicles at various distances from the camera against a complex and varied background providing a rich set of data for training and testing the classifier.

A number of continuous subsets from the 200,000 frames recorded on day two were annotated with bounding boxes around each vehicle present (positive examples). Different sections of the data were then partitioned into a training set of 29,000 positive examples, and a test set of 14,900 examples. Some examples frames from DS2, with ground truth annotations, are shown in Figure 3.18.

3.8 Summary

This chapter introduced the reader to the Catherine MP LWIR polarimeter used throughout this thesis. It explained why the QWIP detector elements are sensitive to polarisation, and how images of the first three Stokes components can be obtained from the output of the camera. Results were then presented from experiments performed to assess the performance of the camera in terms of its response to a range of irradiance values and its sensitivity to polarisation at different angles. The camera was then used to measure the Stokes parameters obtained when viewing different materials with different optical background temperatures (thus different ratios of emission to reflection) which was linked with the theory of reflection and emission of the previous chapter. Finally some example images obtained in an outdoor setting with the camera were presented along with the explanation of the phenomena encountered, and two datasets collected with the camera were introduced.



(a) Frame 1 (b) Frame 2 (c) Frame 3 (d) Frame 4

Figure 3.18: Example frames from DS2 with one ground truth bounding box marked on the I image, note that all vehicles in each frame are annotated in the dataset. From top to bottom, the I, Q, U, P, ϕ and coloured Polarimetric image are shown. The Polarimetric image contains (P, ϕ) information as described in Fig. 3.13.

Chapter 4
Object Detection: A Literature Review

This chapter presents an introduction to the field of object detection. Section 4.1 introduces the standard methodology for object detection, generally involving the extraction of some form of feature from the imagery, and the subsequent classification of this feature in order to determine object class. A selection of features which use two dimensional and three dimensional object representations are then discussed in Section 4.2 and three common classification algorithms are described in Section 4.3. In general, any feature can be combined with any classifier to form an object detection framework. Section 4.4 describes some standard methodologies for evaluating and comparing the accuracies of such detectors. Section 4.5 discusses some works where polarisation imagery has been used for object detection, with progression from segmentation, through anomaly detection to target detection. Finally, Section 4.6 presents a concluding discussion, highlighting areas for development.

4.1 Introduction

Object detection is the process of locating examples of objects such as pedestrians, vehicles, or faces within imagery and video. Object detection is a key component of many computer vision systems with applications including surveillance [45–47], driver assistance technologies [48, 49], and image retrieval systems [50, 51].

A typical object detection framework involves extracting some form of feature from the data, and using these features within a machine learning framework to classify objects as belonging to a particular class. As explained by Theodoridis and Koutroumbas [52] this process can be described as supervised, where an annotated training set exists with examples of both positive and negatives samples, or unsupervised, where the answer is not known *a priori*, and the goal is to cluster "similar" features together.

There are a number of inherent difficulties with object detection, including:

- different object-camera geometries (pose) can cause the scale and shape of an object to vary extensively;
- changes in lighting and weather conditions can have significant effects on the appearance of an object;
- occlusion, whether part or full, can disguise the object's appearance;
- distractors, or background clutter, can draw attention away from the object of interest, resulting in false detections.

Features are generally designed to overcome some of these difficulties. For example a feature can be designed to be scale or rotation invariant in order to reduce difficulties due to object pose; or use gradients rather than absolute intensity values to help overcome lighting effects.

Recent surveys on both vehicle detection [53] and pedestrian detection [54] highlight the significant progress made in detection algorithms over the last decade, with features evolving from the location of edges and symmetrical patterns such as the location of distinctive red brake lights to more complex features using frequency and gradient components.

4.2 Features

The typical input to an object detection algorithm is a grey-scale image consisting of a (potentially large) two dimensional array of intensity values. To reduce the complexity of the data, and to simplify the process of object detection, features (or descriptors) can be extracted from this imagery to overcome some of the inherent difficulties discussed above.

In this Section, two models of feature are described. In Section 4.2.1 a selection of three features, applicable to any class of object, are discussed. These include the Haar-like feature introduced by Viola and Jones [55]; the Histogram of Oriented Gradients algorithm introduced by Dalal and Triggs [56]; and finally the part-based detection method of Fergus *et al.* [57]. Following this, Section 4.2.2 describes two more complex algorithms which extend and tailor their features specifically for detecting and identifying vehicles. These more complex methods are potentially more expensive to compute, but offer a specific solution to a specific problem.

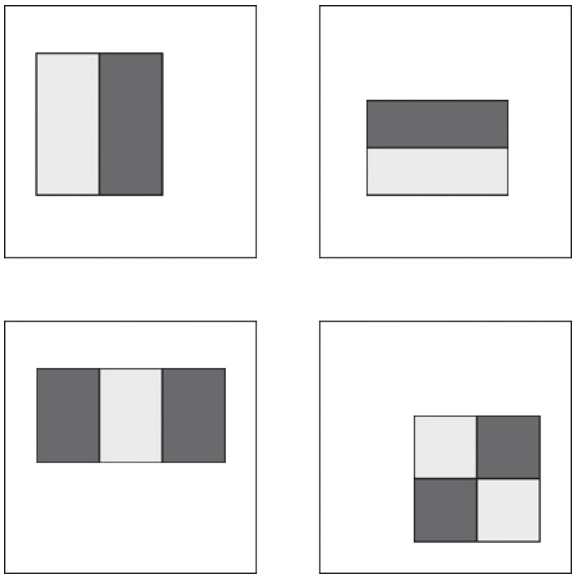


Figure 4.1: Examples of two-, three-, and four-rectangle Haar features. These are applied to the image and a value is calculated by subtracting the sum of pixels under light rectangles from the sum of pixels under dark rectangles.

4.2.1 Generic features

Haar-like Features

Viola and Jones [55] developed a popular algorithm originally designed for face detection using a boosted cascade of Haar-like features. These features consist of simple combinations of two-, three-, and four-rectangles of differing size and position, a selection of which are shown in Figure 4.1. Each feature is calculated by subtracting the sum of pixels within the light rectangular regions from the sum of those within dark rectangular regions.

The cascade is built iteratively by adding the most discriminative weak classifier (comprising a Haar-like feature and a threshold) to a stage in the cascade until it rejects a certain fraction of negative samples remaining after the previous stage. At the same time, the cascade is constrained to limit the number of positive samples rejected to a very small fraction. This cascade classifier is trained using a variation of AdaBoost [58] which is discussed in Section 4.3.1. The first two most discriminative features in the original face detection algorithm are shown in Figure 4.2. The most discriminative features are aligned with distinctive lines or edges within that class of object. In the case of face detection, a distinctive line is formed horizontally between the darker eyes and the lighter cheeks, while a distinctive edge is formed by the lighter nose between the darker eyes.

Although the original paper formulates the Viola and Jones algorithm (VJ) as a face detector, the procedure has been successfully applied by a number of authors

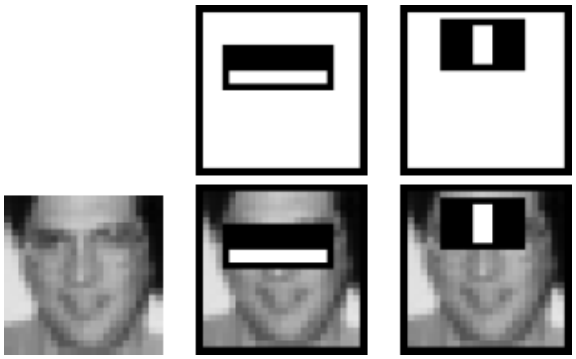


Figure 4.2: The first and second features selected for face detection in the original Viola and Jones paper [55]. The features, shown in the top row, are overlaid on a typical training face in the bottom row. Notice the feature identifying the eye region, where the eyes are typically darker than the cheeks (first feature) and nose (second feature). (Reproduced from [55, Figure 5] with kind permission from Springer Science and Business Media.)

for different detection tasks. Matzka [59] demonstrates three cascades of Haar-like features to detect pedestrians, cars, and lorries from grayscale images for an assisted driving application. Two limitations are mentioned, firstly the classifiers are not viewpoint independent. In this application the system was to be mounted forward facing on a vehicle, and as such vehicles were expected to be encountered from a very restricted range of viewpoints so this was deemed acceptable. A second limitation noted is a more general machine learning limitation in that because the training set used consisted only of daytime images. In hours of darkness the dominant set of features will change, and so the performance of the classifier will vary at different times of day.

Breckon *et al.* [60] have also used VJ to detect vehicles from an Unmanned Aerial Vehicle (UAV) for aerial reconnaissance and surveillance. Imagery of vehicles from the UAV is fairly uniform in elevation, but the azimuthal pose of the vehicles can vary which causes problems due to the classifiers being sensitive to orientation. To overcome this, four different classifier cascades were trained to detect vehicles at four different orientations, the training data for one of which is shown in Figure 4.3. In a later paper [61] additional cues, such as areas of strong thermal intensity compared to their surroundings, were used to verify the detections and reduce additional false positives. Within the current work, the combination of cues from different wavebands is discussed later in Chapter 7.

Integral Image In order to compute many thousand Haar-like features quickly, one of the novel contributions of the Viola and Jones paper was a new image representation known as the integral image. In this representation, each pixel in the integral image, $i_{\text{int}}(x, y)$, takes a value equal to the sum of pixels above and to the

Figure 4.3: Classifiers built from Haar-like cascades are not invariant to rotation. To overcome this Breckon *et al.* [60] split their vehicle training set into four based on the positional orientation to the horizontal. Shown is a subset of positive samples used to train the 135° classifier. (Reproduced from [60, Figure 2] © 2009 IEEE.)

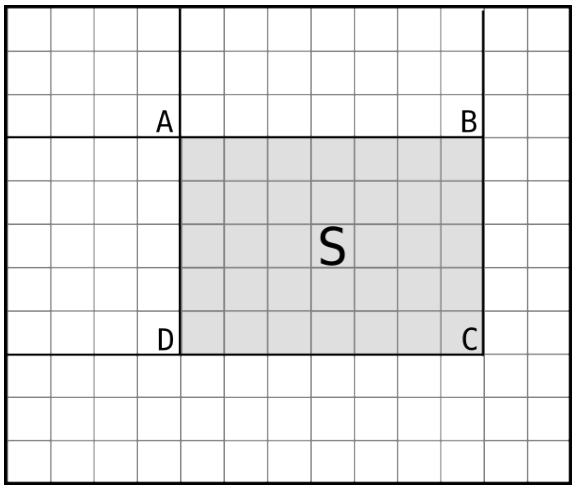


Figure 4.4: Using integral images the sum of the pixels within rectangle ABCD (S) can be computed using only three additions and four memory access operations: S = A + C − B − D.

left of it in the original image $i\,(x, y)$

$$i_{\text{int}}\,(x, y) = \sum_{x'=0}^{x} \sum_{y'=0}^{y} i\,(x', y')\,. \tag{4.1}$$

Once such an integral image has been constructed, the sum of any rectangular area can be computed from the values of the four corner pixels in the integral image (see Figure 4.4). Using this method the two-, three-, and four-rectangle features can be calculated efficiently with only six, eight, and nine references respectively.

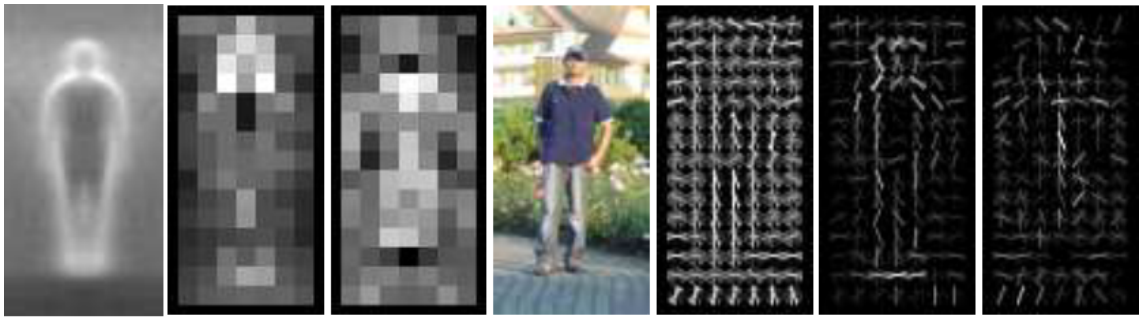


Figure 4.5: The HOG descriptor of Dalal and Triggs [56] mainly cues of silhouette contours. (a) The average gradient image over the training examples. (b) Each "pixel" shows the maximum positive SVM weight in the block centred on that pixel. (c) Likewise for the negative SVM weights. (d) A test image. (e) Its computed HOG descriptor. (f,g) The HOG descriptor weighted by the positive and negative SVM weights respectively. (Reproduced from [56, Figure 6] © 2005 IEEE.)

Histogram of Oriented Gradients

Dalal and Triggs [56] developed the Histogram of Oriented Gradients (HOG) algorithm as a discriminative method for pedestrian detection. The method is based around the assumption that an object can be robustly described by the distribution of intensity gradients across the object, with the main cues for pedestrian detection being found to come from silhouette contours.

To calculate the feature, gradients are computed across the greyscale input window. Following this, each pixel calculates a weighted vote for an edge orientation histogram based on the orientation of the gradient element centred on it. The votes (a function of gradient magnitude) are then accumulated into nine orientation bins in the range 0° to 180° over local spatial regions known as *cells*. Votes are interpolated bilinearly between neighbouring bin centres in both orientation and position to reduce aliasing. The gradient histograms in each cell are then normalised over larger overlapping spatial *blocks*. The final descriptor is then the vector of all components of the normalised cell responses from all of the blocks in the detection window. Figure 4.5 shows an example of the HOG descriptor. Notice that important positive cues come from gradients around the silhouette of the person, especially around the head and shoulders.

When published, the HOG algorithm outperformed other methods, giving near-perfect results on the MIT pedestrian database [62] leading the authors to collect the more challenging INRIA dataset. While more recent algorithms such as Integral Channel Features [63] have now outranked it in terms of increased detection accuracy, the use of some form of gradient histogram is still common. In a recent review on pedestrian detection [54] it is noted that the majority of pedestrian

detection algorithms incorporate gradient histograms into their features, with the state-of-the-art methods additionally combining them with other cues such as colour or texture. While the HOG algorithm was originally developed for pedestrian detection it has since been used for a number of object detection tasks such as vehicle detection [48, 64, 65], traffic sign detection [66], and detection in shadow [67].

Part-based Object Detection

As a third example, Fergus *et al.* [57] have developed a descriptor based on detecting and describing stable feature points within an image. In their approach objects are modelled as a flexible constellation of parts, where each part has a location, **X**, relative scale, **S**, appearance, **A**, and may or may not be occluded. Location, relative scale, and appearance are all modelled by Gaussian probability density functions characterised by a set of mean, covariance, and appearance parameters.

Stable features points are identified using the method of Kadir and Brady [68] which finds regions which are salient over location and scale. A histogram is made of the intensities within a circular region radius (scale) **S** at every point **X** in the image. The entropy of these histograms are computed, and local maxima within this scale-space are selected as features. An image patch is extracted from this location and scale and rescaled to a fixed (11×11 pixel) size to which principal component analysis (PCA [69], discussed later in Section 4.4.3) is applied to reduce the dimensionality.[1] The first 10-15 components are selected to define the appearance, **A**. This method provides 20-30 parameters per part, and it was found that classes of object can generally be represented with 6-7 parts.

After testing their model on a number of data sets (faces, motorbikes, planes, cars (side view), cars (rear view) and spotted cats) they show improvement over previous studies using these datasets by an average receiver-operating characteristic (ROC, see Section 4.4.1) equal error rate of 91% compared to 81% for the previous studies.

This feature is interesting due to the ability of the model to adapt to each dataset. The motorbike model, shown in Figure 4.6, features tight shape models, but some parts have highly variable appearance. For these parts, any feature in that location is accepted with little restriction on appearance. In contrast, Figure 4.7 shows the spotted cat model which consists of a loose shape model but a highly distinctive appearance for each patch. This model essentially looks for patches of spotted fur wherever they occur.

[1] This process allows the method to be scale invariant. In a subsequent publication [70], the use of an elliptical sampling window, and two scale factors additionally allows affine invariance.

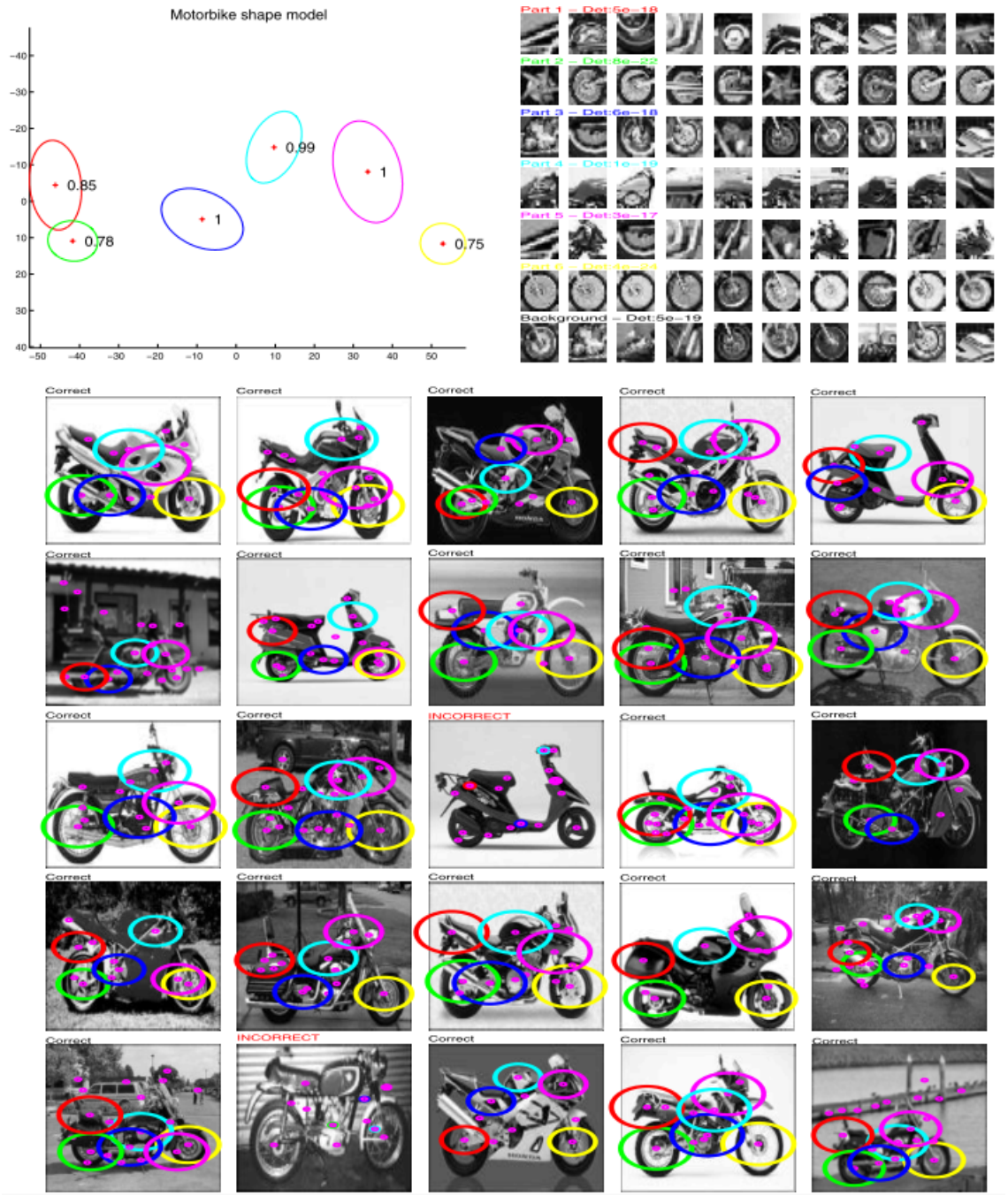


Figure 4.6: A 6 part motorbike model from the method of Fergus *et al.* [57]. The expected position of each part are indicated by the ellipses in the top left figure, while the appearance of image patches for each part are shown in the top right figure. The remainder of the figure indicates parts detected for a number of test images. Note the clear identification of the front wheel, and other parts of the motorbike. (Reproduced from [57, Figure 6] © 2003 IEEE.)

Figure 4.7: A 6 part spotted cat model from the method of Fergus *et al.* [57]. The expected position of each part are indicated by the ellipses in the top left figure, while the appearance of image patches for each part are shown in the top right figure. The remainder of the figure indicates parts detected for a number of test images. Note the loose shape distribution, but distinctive "spotted fur" appearance. (Reproduced from [57, Figure 7] © 2003 IEEE.)

4.2.2 Tailored features

The methods presented in the previous section were fairly generic in that they could be applied to detect any class of object, given suitable training data. In this section, features are presented which use the *a priori* knowledge of three dimensional object shape to offer more advanced detection. The three dimensional extent of objects can be considered a problem, in that viewing an object from different orientations can result in large differences in object appearance. Instead of considering this a problem, incorporating the three dimensional nature of objects into the model used can indeed help the detector. A number of new datasets have recently been published which include the annotations of 3-D pose to aid 3-D object detection algorithms [71, 72]. In this section two such methods of incorporating the three dimensional nature of vehicles are described, a three dimensional extension to HOG for vehicle detection, and an application of an adaptive three dimensional vehicle model.

3-D Extended Histogram of Oriented Gradients

Buch *et al.* [64, 73] have developed a method for detecting and classifying road-users with the generic categories of Bus/Lorry; Van; Car/Taxi; Motorbike/Bicycle and Pedestrian using three dimensional object models with two dimensional imagery. They found that while HOG had been demonstrated to be a reliable method, large pose variation significantly reduces the accuracy, and as such a method of normalising the three dimensional pose was required.

In their 3DHOG method, three dimensional models of each object class are created, and the object surfaces are divided into cells with known surface normal. The models are then aligned with moving blobs in the input imagery, and the pixels within each model cell are normalised based on the surface normal of that cell. Once normalised, a standard HOG feature is calculated within each cell. Different object classes require the use of different three dimensional models. An example of the different stages in this algorithm is presented in Figure 4.8.

The method was evaluated on the public i-LIDS dataset and was shown to outperform other reported results. An initial step in the algorithm is to identify foreground pixels based on a Gaussian Mixture Model of the scene. This step requires that background pixels remain approximately constant, and so would be difficult to apply to a moving camera. The method is also fairly complex, and so for it to perform robustly it would require an extensive set of training data.

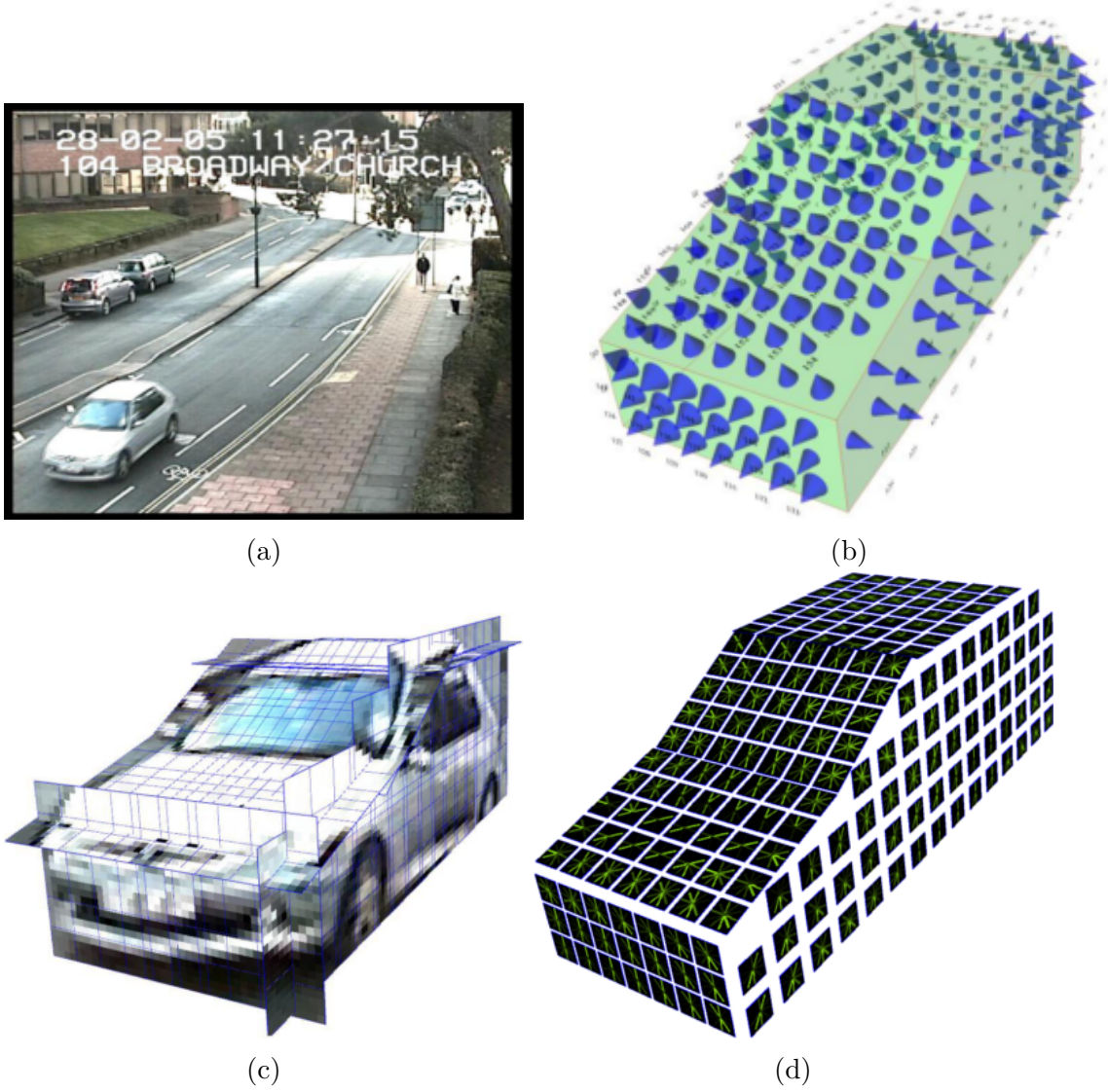


Figure 4.8: Stages in the 3DHOG method. A moving blob is detected in input image (a) with which a three dimensional vehicle model (b) is aligned (the cones indicate the surface normal orientation). (c) Patches are extracted and normalised based on the surface normal of the model patches. (d) Finally, HOG features are computed for each extracted patch. (Reproduced from [74, Figures 45 & 46].)

(a) (b)

Figure 4.9: An example of hypothesised model edges projected onto images (a) before, and (b) after the fitting process. Note that both the pose and shape of the model have been optimised to match the imagery. (Adapted from [47, Figure 9] © 2010 IEEE.)

Adaptive 3-D Vehicle Model

Many methods require a three dimensional vehicle model. However for objects such as vehicles, where the shapes of each make and model are different, the definition of a realistic model is hard to obtain. Leotta and Mundy [47] have developed an adaptive 3-D vehicle model which is able to mould itself to a particular vehicle instance in the image, in a similar manner to deformable models which are frequently used for real-time tracking of facial expressions [75].

The fitting process attempts to optimise the overlap between edges predicted using the model (such as the boundary between the windows and the body, or the outline of the vehicle), and edges detected in the imagery. The model consists of the three dimensional structure of a vehicle, along with part locations (such as windows) and is parametrised in terms of shape and pose. An example of hypothesised model edges projected onto images before and after fitting is shown in Figure 4.9.

The authors suggest such a deformable model could improve tracking and classification algorithms which utilise a three dimensional model of the vehicle. However, it is noted that the final result is sensitive to the initialisation of the model, with most examples in the paper using foreground detection and motion vectors to initialise the location and pose of the object or manual initialisation with specified location and pose. This means the method lends itself well to applications such as the identification of vehicle type, given a suitable initialisation, but is perhaps too computationally complex to detect vehicles in the first instance.

4.2.3 Discussion

In this section a number of different algorithms have been described which all seek to extract an invariant representation of a class object for use in a subsequent classification algorithm. Simple features such as Haar-like wavelets can be used in multiple locations and orientations to span the imagery and an intelligent training scheme can be used to select the best subset of these features, or more complex features which densely span an object such as HOG, or the three dimensional extension, 3DHOG can be used to extract a rich representation directly. The choice of which feature to use for a particular class of problem depends on a number of parameters including the volume of training data available, the similarity (in appearance, orientation, or external environment) of the class of objects, and the computational budget available for computing the feature (more complex features being more expensive to compute).

For tasks such as frontal face detection, where the pose is controlled to some extent and the lighting is fairly stable, the Haar-like feature approach of Viola and Jones (VJ) is shown to work well. However, this algorithm suffers if the object class contains significant variations in pose or appearance. The Histogram of Gradients (HOG) algorithm also suffers in this regard. The HOG algorithm has been shown to work well for pedestrian detection, however if a person is viewed from the side, or if they are sitting or in any other position, its performance quickly deteriorates. Methods such as HOG are robust and reliable for detecting rigid objects from a consistent viewpoint. However, to maintain performance for a generic viewpoint, either additional classifiers are required to cover different cases, or more expensive multi-view classifiers are required.

Parts-based methods, such as that of Fergus *et al.* discussed above, are certainly more complex than both VJ and HOG, but by being more flexible in their representation of an object are able to accommodate both a wide variation in object appearance, but also object pose to a certain extent. Full three dimensional approaches, involving detailed (or approximate but adaptable) three dimensional models of the object class are more complex still and, while this might result in improved accuracy for that particular class of objects, the difficulty in adapting such methods to different classes is seen as a disadvantage.

4.3 Classifiers

A classification algorithm can be trained to identify to which of a set of classes a new observation belongs. In general, any classification method can be combined with any set of features to build a complete detector. In the following section,

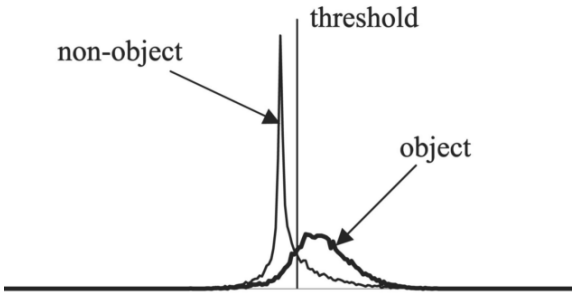


Figure 4.10: An example value distribution of a particular Haar-like feature. The Viola and Jones face detector [55] selects a series of single features and thresholds which best separate two classes. (Reproduced from [79, Figure 3] © 2008 IEEE.)

three such classification algorithms are introduced: AdaBoost [58], Support Vector Machines [76, 77], and Neural Networks [78].

4.3.1 AdaBoost

The Viola and Jones face detector [55], discussed above, made use of an adapted AdaBoost classifier [58] to identify which of the many thousand features were most discriminative, realising that among all the possible features (many times the number of pixels in the input window), only a few were expected to give consistent values for all faces.

Each Haar-like feature, when evaluating a particular window, x, returns a number $f(x)$ which can be used as an input to a weak classifier, $h(x, f, p, \theta)$. A weak classifier is formally defined as

$$h(x, f, p, \theta) = \begin{cases} 1 & \text{if } pf(x) < p\theta \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

where $f(x)$ is the current feature's response to pixel window x, p is the polarity (+1 or -1), and θ is the classification threshold. Figure 4.10 demonstrates a typical distribution for a feature, $f(x)$. The training process determines the best threshold to separate the two classes, and the appropriate polarity to select whether the object class is above or below the threshold.

While a single weak classifier is not very discriminative, a set of weak classifiers can be combined in a weighted sum to produce a stronger classifier. The AdaBoost algorithm, outlined in Algorithm 4.1, is designed to select the most discriminative of a large set of weak classifiers to achieve this. The Viola and Jones adaptation of AdaBoost used this mechanism to construct a cascade of classifiers (each stage consisting of a set of weak classifiers), where earlier stages in the cascade reject the majority of negative subwindows before subsequent more complex classifiers are

relied on to reduce the false positive rate, as represented in Figure 4.11. The Viola and Jones paper used a 38 layer cascade of classifiers consisting of two features in the first classifier (which rejected 50% of non-faces), ten features in the second classifier, two 25-feature classifiers and a total of 6,060 features in the whole cascade. However, due to the use of a boosted cascade, where windows were dismissed immediately when any stage of the classifier reported a negative, the original authors found that an average of only 10 feature evaluations per sub-window were required when tested on a difficult dataset containing 507 faces and 75 million sub-windows.

Algorithm 4.1 AdaBoost method for selecting key weak classifiers. (Reproduced from [55, Table 1] with kind permission from Springer Science and Business Media.)

INPUT: Example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples, x_i, respectively; T number of weak classifiers to choose.

OUTPUT: Strong classifier, $C(x)$, comprising a set of T key weak classifiers.

1: Initialise weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negative and positive samples respectively.

2: **for** $t = 1 \rightarrow T$ **do**

3: Normalise the weights,

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$

4: Select the best weak classifier with respect to the weighted error

$$\epsilon_t = \min_{f,p,\theta} \sum_i w_i \left| h(x_i, f, p, \theta) - y_i \right|.$$

5: Define $h_t(x) = h(x, f_t, p_t, \theta_t)$ where f_t, p_t, and θ_t are the minimisers of ϵ_t.

6: Update the weights

$$w_{t+1,i} = w_{t,i} \beta_i^{1-e_i}$$

where $e_i = 0$ if example x_i is classifier correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

7: **end for**

8: The final strong classifier is

$$C(x) = \begin{cases} 1 & \text{if } \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha_t = \log \frac{1}{\beta_t}$.

4.3.2 Support Vector Machine

The Support Vector Machine (SVM) [76, 77] is a commonly used supervised classification algorithm. Given a set of training samples $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ with $\mathbf{x}_i \in \mathbb{R}^p$ along with their labels $\{y_1, y_2, \ldots, y_N\}$ with $y_i \in \{-1, 1\}$ the method seeks to find a hyperplane,

$$\mathbf{x} : f(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta} + \beta_0 = 0, \tag{4.3}$$

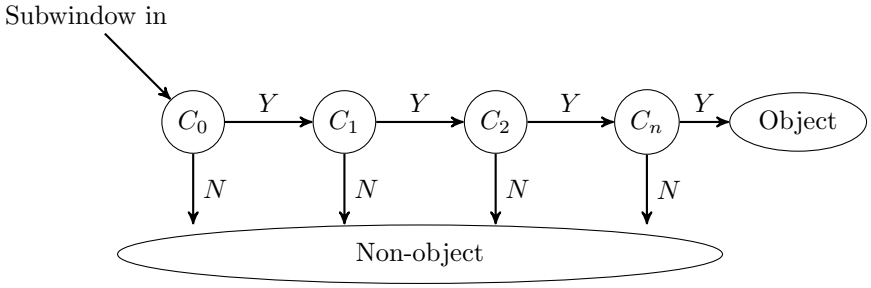


Figure 4.11: Schematic representation of the Viola and Jones classifier cascade. Each sub-window is passed through a series of classifiers. The initial classifier rejects a large number of negative examples with minimal processing, while subsequent more complex classifiers require additional processing but only evaluate a small subset of the assessed windows.

which creates the biggest margin between the training samples of the two classes (see Figure 4.12). This leads to the classification rule

$$G(\mathbf{x}) = \text{sign}\left[\mathbf{x}^T\boldsymbol{\beta} + \beta_0\right]. \tag{4.4}$$

When the classes are separable, a function

$$f(\mathbf{x}) \;=\; \mathbf{x}^T\boldsymbol{\beta} \;+\; \beta_0 \tag{4.5}$$

exists with

$$y_i f(\mathbf{x}_i) \;>\; 0 \qquad \forall\, i. \tag{4.6}$$

The task of the SVM is then to find the hyperplane which creates the biggest margin between training points for class 1 and -1. This is captured through the optimisation problem

$$\begin{aligned} &\min_{\boldsymbol{\beta},\beta_0} && ||\boldsymbol{\beta}|| \\ &\text{subject to} && y_i\left(\mathbf{x}_i^T\boldsymbol{\beta} + \beta_0\right) \geq M, i = 1, \ldots, N, \end{aligned} \tag{4.7}$$

where $M = \frac{1}{||\boldsymbol{\beta}||}$ is the size of the *margin* shown in Figure 4.12.

To accommodate cases where the two classes overlap in the feature space, slack variables $\xi = (\xi_1, \xi_2, \ldots, \xi_N)$ are defined to measure the amount by which the prediction is on the wrong side of its margin (see Figure 4.12). Bounding the sum $\sum \xi_i$ bounds the total number of training misclassifications, and alters the optimisation to

$$\begin{aligned} &\min_{\boldsymbol{\beta},\beta_0} && ||\boldsymbol{\beta}|| \\ &\text{subject to} && \begin{cases} y_i\left(\mathbf{x}_i^T\boldsymbol{\beta} + \beta_0\right) \geq 1 - \xi_i \quad \forall i, \\ \xi_i \geq 0, \quad \sum \xi_i \leq C. \end{cases} \end{aligned} \tag{4.8}$$

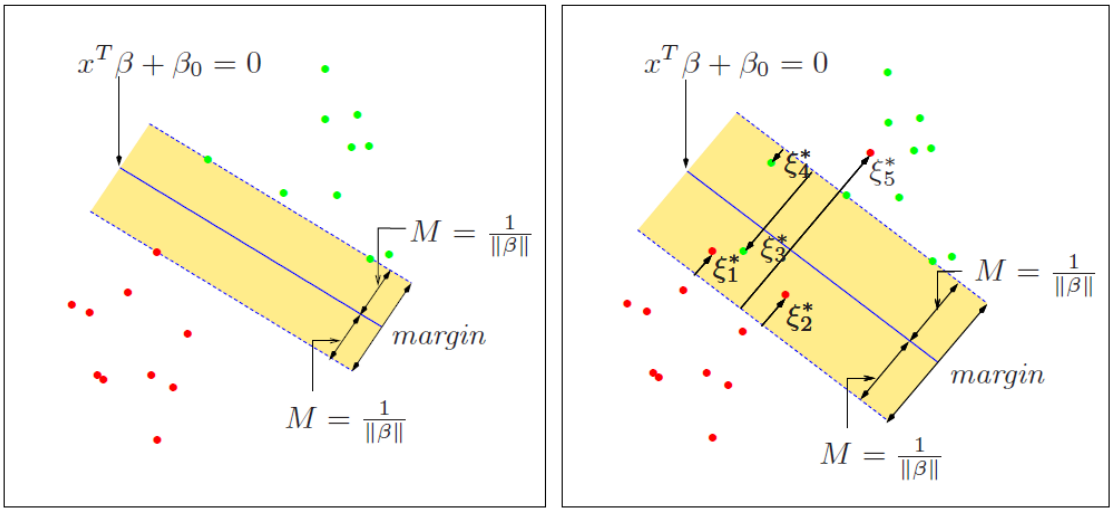


Figure 4.12: Support vector classifiers. The left panel shows the separable case. The decision boundary is the solid line, while broken lines bound the shaded maximal margin of width $2M = 2/\left|\left|\beta\right|\right|$. The right panel shows the non-separable (overlap) case. The points labelled $\xi_j^\star$ are on the wrong side of the their margin by an amount $\xi_j^\star = M\xi_j$; points on the correct side have $\xi_j^\star = 0$. The margin is maximised subject to a total budget $\sum \xi_i^\star \leq$ constant. Hence $\sum \xi_j^\star$ is the total distance of points on the wrong side of their margin. (Reproduced from [80, Figure 12.1], with kind permission from Springer Science+Business Media.)

Here $C > 0$ is the regularisation parameter, or penalty parameter of the error term. One common solution to this problem involves Lagrange multipliers and can be found in textbooks such as [80, Chapter 12].

To overcome cases where the data is not linearly separable, it is common to replace the training vectors $\mathbf{x}_i$ with $\phi\left(\mathbf{x}_i\right)$ where $\phi\left(\cdot\right)$ is a kernel function which maps $\mathbf{x}_i$ into a higher-dimensional space. In the standard implementation of an SVM the computation of $\phi\left(\mathbf{x}_i\right)$ is not directly required, rather the implementation only requires computation of the inner products in the transformed space,

$$K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left\langle \phi\left(\mathbf{x}_i\right), \phi\left(\mathbf{x}_j\right) \right\rangle. \tag{4.9}$$

Common choices for K include

- linear: $K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \mathbf{x}_i^T \mathbf{x}_j$,
- polynomial: $K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left(\gamma \mathbf{x}_i^T \mathbf{x}_j + r\right)^d$,
- radial basis function (RBF): $K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp\left(-\gamma \left|\left|\mathbf{x}_i - \mathbf{x}_j\right|\right|^2\right)$,
- sigmoid: $K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \tanh\left(\gamma \mathbf{x}_i^T \mathbf{x}_j + r\right)$.

where $\gamma > 0$, r, and d are kernel parameters.

The effectiveness of the SVM classifier is dependent on the choice of kernel, any kernel parameters, and the regularisation parameter C. The RBF kernel is generally considered a good choice but has a long run-time and so the linear SVM is the most common choice for object detection frameworks [54]. The choice of parameters is normally arrived at through a brute-force search over a grid of parameters (for example $C = 2^{-5}, 2^{-3}, \ldots, 2^{15}$ and $\gamma = 2^{-15}, 2^{-13}, \ldots, 2^{3}$), a process which is not too arduous and lends itself to parallel programming.

4.3.3 Neural Network

Neural networks [78], inspired by the workings of the brain, are another example of a nonlinear statistical model. They consist of a set of interconnected units, or neurons, which send messages to one another. Each neuron provides an output, y_j according to a function of their input, x_i, a set of weights, W_{ij}, and a choice of activation function, $f(x)$,

$$y_j = f\left(\sum W_{ij} x_i\right). \tag{4.10}$$

This activation function may take a number of forms, some of which include

- hard limit: $f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$,
- pure linear: $f(x) = x$,
- sigmoid: $f(x) = \frac{1}{1+e^{-x}}$.

The neural network consists of a set of neurons arranged into layers, with features flowing into an input layer, through any number of hidden layers, before a final output layer is reached. Each layer may consist of any number of neurons, with weights being applied at each transition. The output layer can consist of as many outputs as required, for example there could be a single output which is positive for positive samples and negative for negative samples or there could be one output per class, and the value at each output neuron would be the probability of the input vector belonging to that class. A graphical representation of a neural network is shown in Figure 4.13.

An important aspect of neural networks is the learning process, where the weights of each connection are determined. The back propagation (BP) neural network [81] is a relatively mature iterative training method which operates in two phases. Initially a training pattern is passed through the network and the corresponding output is generated. The output is then compared to the target output and the error is back propagated through the network, in order to update the weights for the next training iteration.

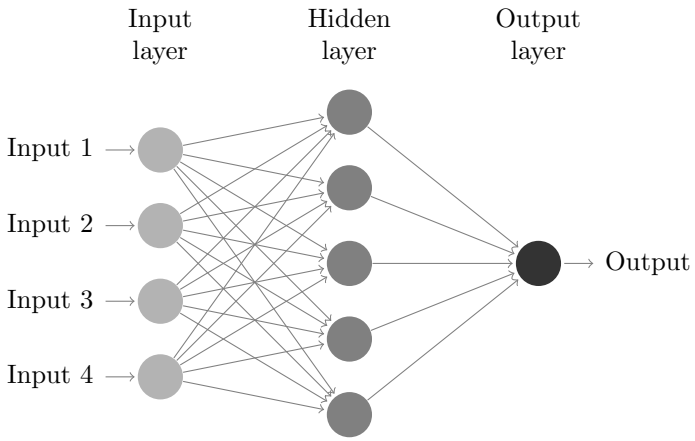


Figure 4.13: Example neural network structure. Each circle is a neuron which combines all inputs and provides an output according to an activation function. There can be any number of hidden layers consisting of any number of neurons per layer.

4.3.4 Discussion

From the literature it seems as though the choice of classifier is largely a matter of either personal preference or a trial-and-error process of finding the combination of feature and classifier which produces to best accuracy. The set of three classifiers presented in this section are certainly not an exhaustive list, with other classifiers such as Naive Bayes, k-nearest neighbours, Decision Trees, and Random Forests being used for a number of applications. Some considerations in choosing a classifier include how the classifier will be trained (will it be incrementally shown new examples, or is training a batch process where all examples are available at once), what is the required output from the classifier (categorical or numerical), speed (both training and at run-time), and complexity (are the classes linearly separable or not).

Support Vector Machines are designed to be batch trained, are suitable for both classification and regression (depending on the implementation), are comparatively slow to train but quick to run, and given different kernel functions can handle non-linear class separation. However, one of the main disadvantages of the SVM classifier is that due to the high-dimensional mapping of feature vectors by the kernel function, it can be very difficult to determine exactly why a certain feature was classified a certain way.

4.4 Performance Analysis

Given detection framework comprising a feature and classifier, it is important to be able to assess the performance in order to compare different choices of feature or

parameter. In this section the ROC curve for comparing classification performance is introduced as are two additional methods for examining how suitable different features are for distinguishing positive and negative classes.

4.4.1 Receiver Operating Characteristic

Receiver Operating Characteristic (ROC) curves are a common method of graphically comparing binary detectors based on the number of true positive (tp), false positive (fp), true negative (tn), and false negative (fn) responses at a series of sensitivity levels. They are constructed by plotting the true positive rate ($tpR = \frac{tp}{tp+fn}$) against the false positive rate ($fpR = \frac{fp}{fp+tn}$) at a number of sensitivity thresholds, t. A perfect detector is represented by a curve which passes through the point

$$(fpR, tpR) = (0.0, 1.0)\,. \tag{4.11}$$

For an object detection algorithm, producing the ROC curve requires the output set of bounding boxes detected from the test imagery (BB_{dt}) along with their associated confidences, and the set of ground truth bounding boxes (BB_{gt}). The detected bounding boxes are then partitioned based on a threshold, t, on the confidence returned. For detected bounding boxes with confidence greater than t a true positive (tp) is recorded if they overlap sufficiently with a ground truth bounding box, as defined by the PASCAL measure [82] which states that their overlap, a_o, must exceed 50% according to the equation:

$$a_o = \frac{\text{area}\,(BB_{dt} \cap BB_{gt})}{\text{area}\,(BB_{dt} \cup BB_{gt})} > 0.50 \tag{4.12}$$

otherwise a false positive (fp) is recorded. Similarly for detected bounding boxes with confidence less than t (no detection), a true negative (tn) is recorded if the overlap with any ground truth BB is less than 50% otherwise a false negative (fn) is recorded.

4.4.2 Descriptor Dimension Analysis

As an additional method of qualitatively analysing the ability for a particular descriptor to discriminate between positive and negative samples, different pairs of dimensions of the descriptor are plotted to examine the separation in different planes. In general this is only an indication of the multi-dimensional separation, but if the positive and negative samples appear separated in the projection, then they are likely to be separable with a multi-dimensional classifier such as the SVM. In these figures, descriptors of each positive and negative training sample are plotted in an

array of figures where each figure shows the projection onto each pair of dimensions within the descriptor.

4.4.3 Principal Component Analysis

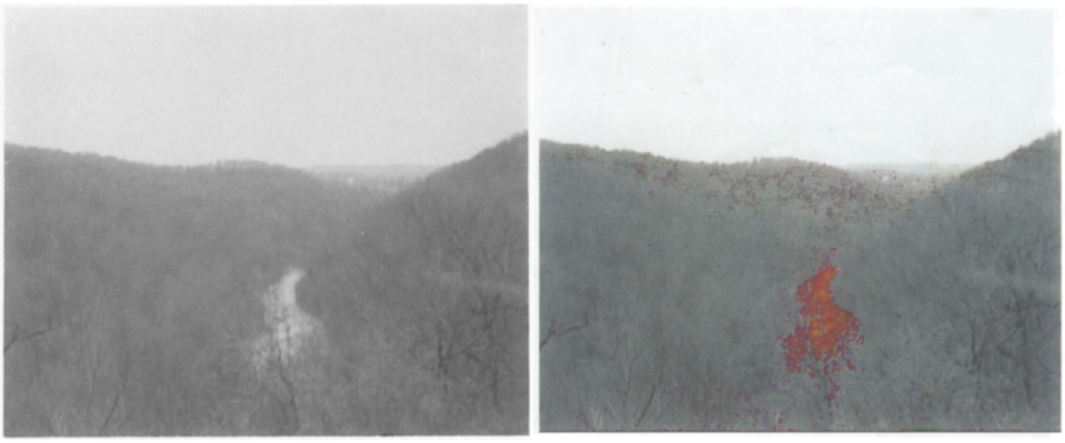
As an alternative method for visualising the separability of positive and negative descriptors, Principal Component Analysis (PCA) [69] is used to project the descriptors into a different space. PCA is a method which transforms multidimensional data into a set of linearly uncorrelated dimensions in such a way that the first component has the largest possible variance, and each subsequent component then has the highest possible variance under the constraint that they are orthogonal to each preceding component. By computing the principal components of the positive and negative descriptors, the largest variation (ideally the separation into positive and negative descriptors) is captured in the first few dimensions, and so plotting these can give a better qualitative interpretation of the overlap between these classes.

4.5 Use of Polarisation for Detection

4.5.1 Qualitative Assessment of Polarisation

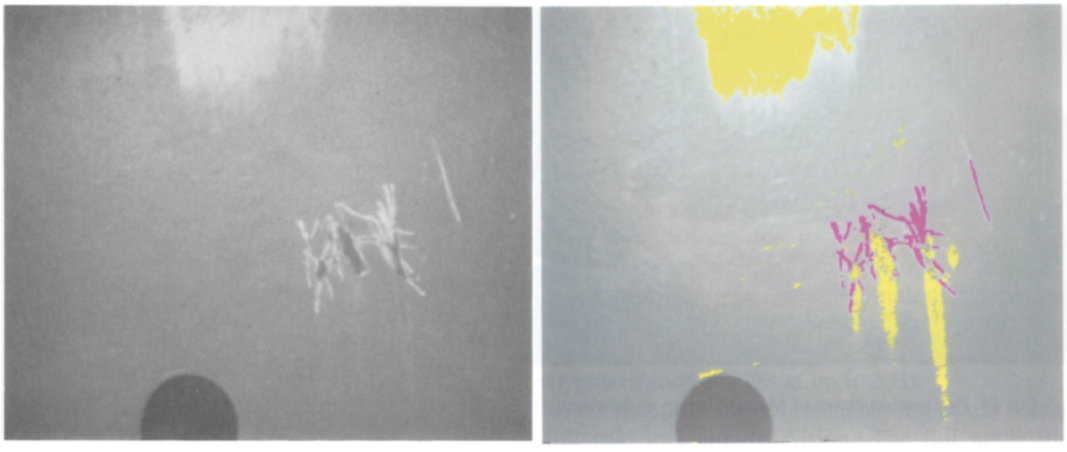
Polarimetric imaging, and its applicability to improving detection accuracy, has been of interest to researchers for a number of years. Many papers have been published which qualitatively assess the usefulness of this domain. Wolff [83] highlights a number of potential applications for polarisation including the segmentation of water based on the strong partial polarisation of the water compared to the scene background (Figure 4.14) and the segmentation of rusted scrapes from superficial scrapes on a ship hull based on the degree of depolarisation of reflected light (Figure 4.15). Similarly, Connor *et al.* [84] and Carrie and Connor [85] demonstrate the ability for polarimetric LWIR to discriminate man-made objects from cluttered backgrounds and to assist in detecting disturbed earth, important for land mine detection.

The detection and effective identification of land mines is a strong area of interest for the humanitarian community. A number of sensors are commonly used for this task including ground penetrating radar, metal detectors, and infrared sensors. A number of authors have additionally assessed the ability for polarisation measurements to improve detection performance. Barbour *et al.* [86] find a significant benefit in using IR polarisation above standard intensity images, with Degree of Linear Polarisation (P) images providing a higher signal-to-noise detection capability than the intensity images. Similarly Jong *et al.* [87] found benefit applying



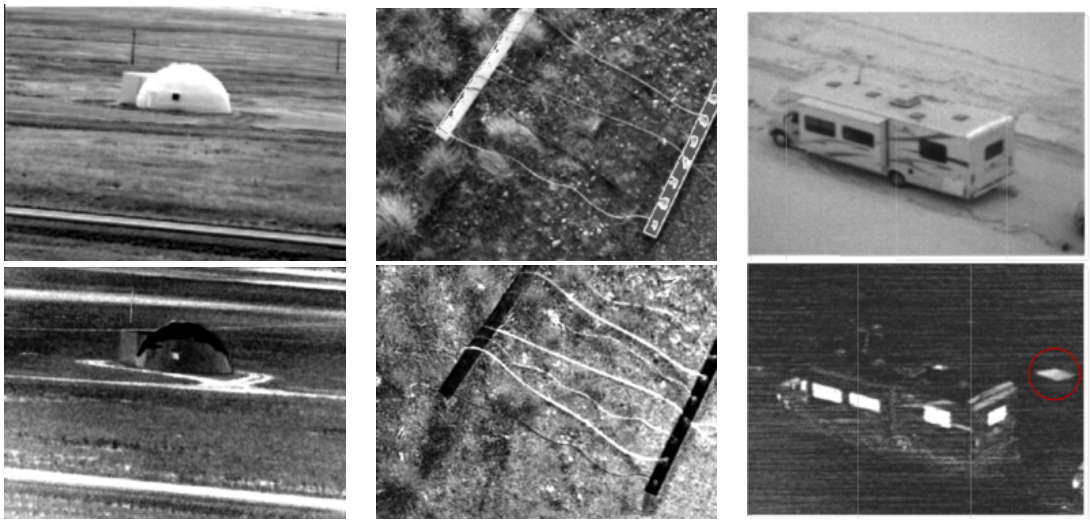
(a) (b)

Figure 4.14: (a) Intensity and (b) Polarisation image of a river running through wooded hills under cloudy sky. (Reprinted from [83, Figures 14, 15], with permission from Elsevier.)



(a) (b)

Figure 4.15: (a) Intensity image and (b) Polarisation-based segmentation of rust (yellow) and superficial scrapes (magenta). (Reprinted from [83, Figures 20, 21], with permission from Elsevier.)



(a) (b) (c)

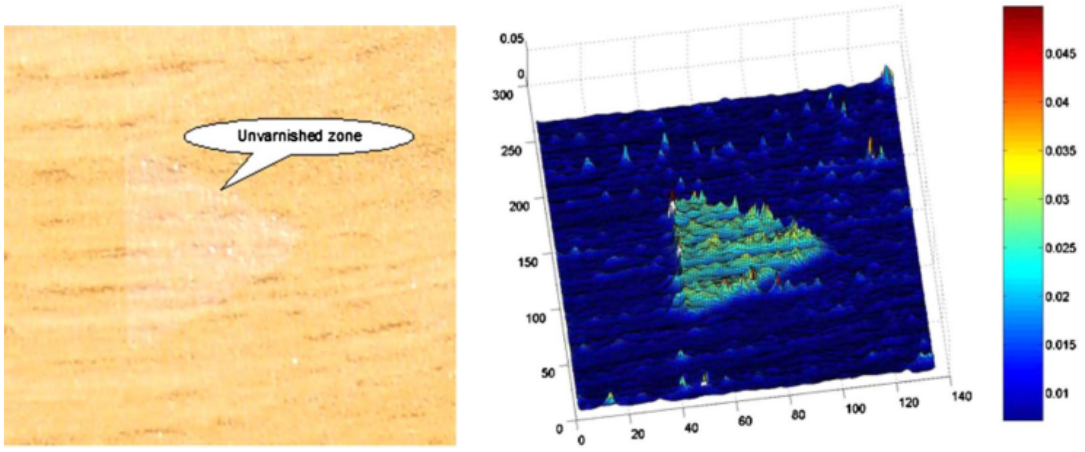
Figure 4.16: (a) Visible band intensity (top) and degree of linear polarisation (bottom) shows the ability of polarisation to aid track detection. (b) Visible band intensity (top) and degree of linear polarisation (bottom) shows the enhancement of wires against the natural background. (c) SWIR band intensity (top) and degree of linear polarisation (bottom) shows a Plexiglas plate being easily revealed in the DoLP image. (Reproduced from [89, Figures 6, 7, and 8], with kind permission from SPIE and D. Lavigne.)

a threshold to P imagery in visible and MWIR wavelengths to improve land mine detection.

Finally, Lavigne *et al.* [88,89] have performed a number of trials using both active and passive sensors in a range of wavelengths from visible to LWIR. They highlight a number of applications of polarisation imagery including the improved ability to detect roads (Figure 4.16a), wires (Figure 4.16b) and other hard-to-detect objects (Figure 4.16c). The improved ability to detect roads with polarisation is also used by Connor *et al.* [90] where a simple road region classifier based on the degree of polarisation is used as part of a scene-segmentation algorithm to allow subsequent context specific processing of different parts of the scene.

4.5.2 Polarisation-based segmentation

Polarisation has been found to be a strong cue to distinguish certain types of objects, an ability which has been exploited in segmentation algorithms to separate classes of objects. Chen and Wolff [91] use polarisation as an input to their metal/dielectric segmentation algorithm, improving on the earlier work of Wolff [92]. Using the premise that incident linear polarisation will be shifted to an elliptically polarised state by conducting materials while remaining in a linear polarisation state for non-conducting materials, they show a very clean segmentation of solder metal from



(a) (b)

Figure 4.17: Example of a polarisation-based method to detect surface roughness. (a) shows an intensity image of a wooden object with an unvarnished patch. (b) shows a map of s values, which relates to the surface roughness, estimated based on polarimetric imaging. (Reproduced from [5, Figure 7, 8], with kind permission from The Optical Society.)

plastic on a circuit board, and identification of metal in an outdoor scene.

The extent to which a surface polarises incident illumination is related to the degree of surface roughness of the reflecting surface. Terrier *et al.* [5] explore this, noting that if polarisation imagery could be used to measure surface roughness this would avoid the use of a stylus which may damage the surface, or laser methods which can only assess a single point at a time. Using their method, they successfully obtain an experimental measurement of the s parameter, a measure of surface roughness. Figure 4.17 shows an example of an observed piece of wood with a triangular section which is unvarnished. Using the method, they identify this region as having a higher s value than the surrounding varnished area.

4.5.3 Polarisation-based anomaly detection

It has been shown that polarisation can be a valuable cue to identify anomalous materials in cluttered scenes such as metallic objects in natural environments. Conveniently, it is often found that objects of interest are present as anomalous regions in polarimetric imagery, and so a number of polarisation-based anomaly detectors have been proposed.

The RX algorithm [93] and the Topological Anomaly Detector (TAD) [94] are two anomaly detection algorithms. The RX algorithm essentially computes the Mahalanobis distance from each pixel to the global distribution of the image,

$$R(\mathbf{x}) = (\mathbf{x} - \mathbf{m})^T \Sigma^{-1} (\mathbf{x} - \mathbf{m}), \tag{4.13}$$

where $\mathbf{x}$ is the pixel vector, $\mathbf{m}$ is the mean image vector, and Σ is the covariance of the distribution. The TAD algorithm defines background and anomalous points according to the number of nearest neighbours within a certain radius of each point. Gartley and Basener [95] and Bartlett *et al.* [96] have tested these algorithms, and a combination of them, on a number of spectral bands both with and without polarisation. In general they find that the use of polarisation improves results, but not all the time. They also find that the performance of each algorithm was highly dependant on the combination of spectral bands used, and the general scene conditions.

Using an approach designed to exploit the polarimetric variations within a scene throughout a diurnal cycle, Romano *et al.* [97] show the benefit of long wave infrared polarisation information for the detection of military tank surrogates and other man-made objects in a natural background. In their investigation, across a twenty four hour diurnal cycle, none of the Stokes parameters or the degree of linear polarisation was able to completely separate all objects from the background. However, considering the spatial covariance in the (I_{90}, I_0) space, they define the *Local-Global Covariance* metric which compares the covariance of points within a local sliding window to the global covariance of all points in the image. Using this metric, they are able to demonstrate class separation between object and background at all times of day (see Figure 4.18).

The detection of a foliage-obscured vehicle is accomplished by Tan *et al.* [98] using polarisation information from a LiDAR system.[1] They show that measuring the cross-polarisation ratio, defined as the ratio between the cross-polarised return and the co-polarised return, provides a strong distinguishing feature between vehicle and vegetation due to their different reflective properties. This trial from a ground-based lidar system successfully demonstrated the advantages of polarimetric lidar for vehicle detection, and the authors suggest further work on the detection of man-made objects under different types of camouflage, and airborne lidar applications. The use of the co- and cross-polarisation returns is also common in the radar domain for a variety of applications [99–101] where different targets alter the polarisation state to differing degrees.

4.5.4 Polarisation-based target detection

The majority of the discussed algorithms use polarisation imagery (such as degree of polarisation) as a distinguishing modality in which different object classes appear

[1] LiDAR is a remote sensing technique that can measure distance by illuminating a target with a laser and analysing the reflected light. This particular application illuminates a sensor with polarised light and measures the ratio of light returned with the same polarisation state, *co-polarised*, and the opposite polarisation state, *cross-polarised*.

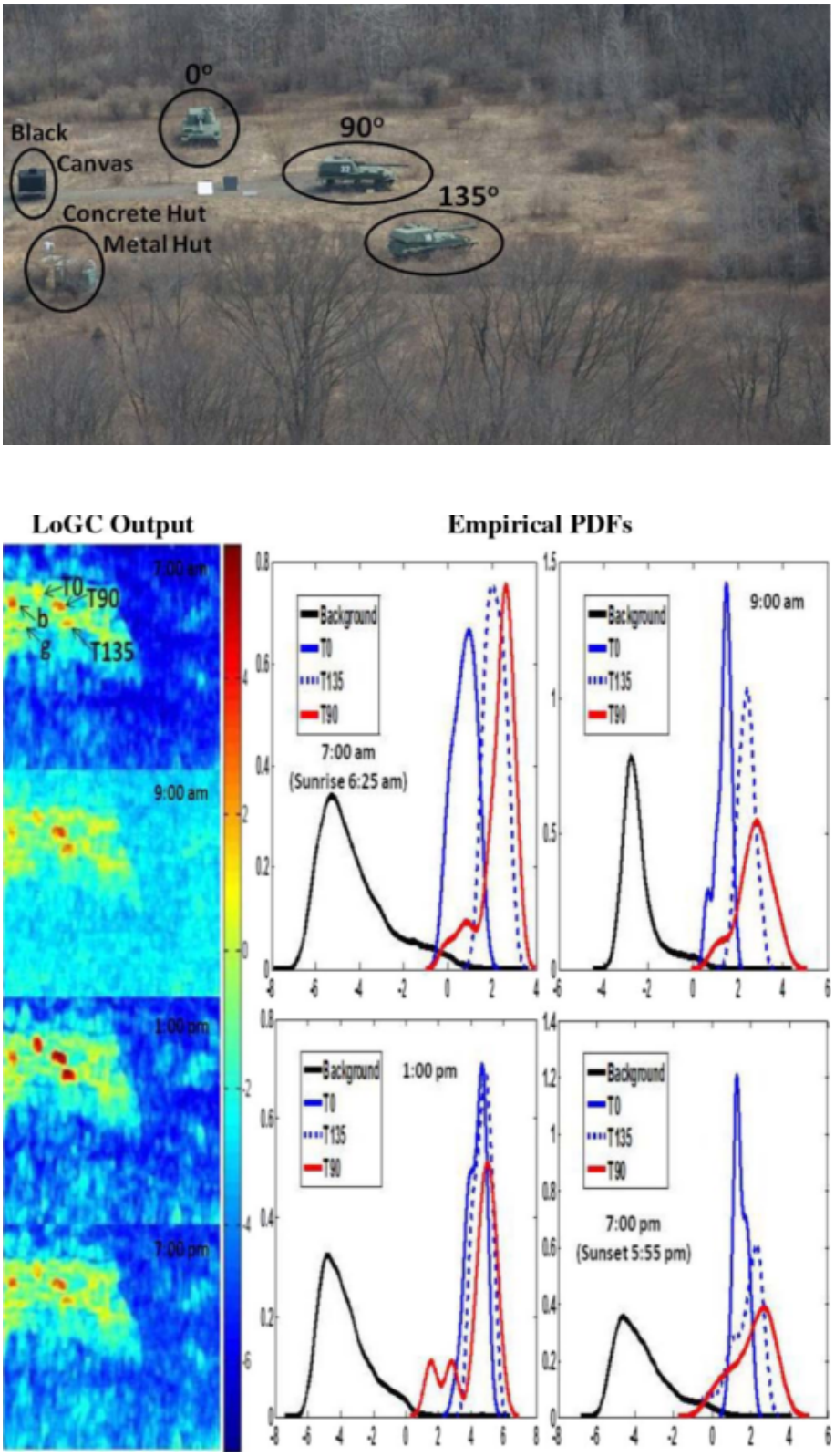


Figure 4.18: Top shows a visible image of the test site comprising military tank surrogates and other man-made objects within a natural cluttered background. Bottom left shows the Local-Global Covariance of the scene as computed from polarimetric LWIR images at four different times of day. Bottom right shows the probability density functions highlighting the separation of target from background throughout the day. (Reproduced from [97, Figures 1-6] © 2012 IEEE.)

at different intensity levels, or as anomalies in some manner. Some papers which begin to consider how polarisation can aid target classification include Chun and Sadjadi [102] and Sadjadi and Sadjadi [103] who present work on modelling the polarimetric appearance of targets and demonstrate a correlation technique which correlates targets with a model library, either matching extracted edges from intensity and polarimetric imagery, or a statistical pixel-based segmentation based on the similarity between the extracted statistics of the sensor image (I, P, and ϕ image) and the statistics of the true image of the target. They found that the use of the polarimetric imagery did improve performance compared with using only intensity imagery. However, beyond this there are no papers, to the author's knowledge, which use polarisation as an input modality to an object detection framework.

4.6 Discussion

The general pipeline for object detection involves extracting features and using a classification algorithm to determine whether those features belong to an object or not. In this chapter some examples of both features and classification algorithms were given. Polarisation imagery has also been shown to be more effective in distinguishing man-made objects and certain types of target from the background when compared to similar methods which rely on intensity imagery alone. The polarisation properties of different objects appear to naturally segment metal from dielectric, and vehicle from natural background.

Methods used to detect objects, such as vehicles, from the polarimetric imagery were described in Section 4.5. These methods generally involve the computation of some metric for each pixel in the image based on the properties of that pixel and its surrounding, such as the RX anomaly detection algorithm, or the Local-Global Covariance metric of Romano *et al.* However, these measures do not consider the spatial extent of objects directly, nor do they directly model the expected polarisation state of vehicles in polarimetric imagery.

Using knowledge about the polarimetric properties of vehicles, the following chapter describes the development of novel descriptors to capture this extended polarimetric signature in order to improve vehicle detection performance.

Chapter 5
Histogram-based Polarimetric Vehicle Descriptor

This chapter presents the progressive development of a series of novel histogram-based descriptors for vehicle detection within polarimetric infrared imagery. Section 5.1 motivates the use of polarimetric descriptors computed using histograms of different local quantities, and introduces the methodology by which the descriptors are assessed and compared. Section 5.2 then demonstrates the Histogram of Oriented Gradients (HOG) [56] algorithm, as introduced earlier in Section 4.2.1 as a benchmark intensity image descriptor. Following this Sections 5.3 to 5.6 present the Histogram of Polarisation Angle, the Polarimetric Histogram [104, 105], the Polarimetric Histogram with Centre of Mass, and the Histogram of Polarisation Angle Differences respectively. Each of these novel descriptors were motivated from the findings of the preceding algorithms, and represent a number of different methods of encapsulating the polarimetric content of an object. Finally, in Section 5.7 all the histogram-based descriptors are compared and thoughts for further improvement are discussed. It is found that the Polarimetric Histogram outperforms all other methods, including the intensity only HOG descriptor.

5.1 Introduction

The use of histograms in forming a descriptor for object detections is a common approach. The well known Histogram of Oriented Gradients (HOG) [56], as introduced earlier in Section 4.2.1, is a good example. With the histogram-based approach the input window is split into a number of spatial cells, and a histogram is computed within each of these cells. After some normalisation, the histograms are stacked together to form the descriptor vector for use in the classification algorithm. The spatial cells allow for some flexibility in the exact location of the objects, and the histograms provide for flexibility in the distribution of the quantity

being assessed (eg gradient orientation for HOG). In this chapter a progression of different histogram-based descriptors are presented, each of which have the same form, as described in Algorithm 5.1. Each descriptor is introduced in the context of building on the previous method, the parameters of each descriptor are assessed and the results and limitations of each descriptor are discussed. In each case, the results presented are obtained by using ground truth data from Data Set 2 (DS2, introduced in Section 3.7.2) to train and test a linear SVM classifier. Following this an overall discussion and summary of the histogram-based approach is presented.

Algorithm 5.1 Descriptor Template

INPUT: I 30×30 pixel input window, p any descriptor parameters
OUTPUT: $\mathbf{d}$ descriptor
1: Process I, using parameters specified, to produce descriptor $\mathbf{d}$

5.1.1 Methodology

The training stage, outlined in Algorithm 5.2, uses the set of training imagery from DS2, along with their known labels. Training samples are resized to 30×30 pixels and the descriptor algorithm is used to obtain a descriptor for each sample. These descriptor-label pairs are then used to train a linear SVM classifier. The SVM is trained using the `auto_train` method which employees 10-fold cross validation to find the optimal C parameter (an internal SVM parameter, as defined in Section 4.3.2).

Algorithm 5.2 Training Procedure

INPUT: $\{I_1, I_2, \ldots, I_n\}$ ground truth training image patches, $\{y_1, y_2, \ldots, y_n\}$ associated labels, $y_i \in \{-1, 1\}$.
OUTPUT: S trained SVM
1: **for** $i = 1 \rightarrow n$ **do** ▷ For each training sample
2: Resize sample to 30×30 pixels
3: $\mathbf{d}_i \leftarrow \text{compute}\,(I_i)$ ▷ Compute descriptor, Algorithm 5.1
4: **end for**
5: $S \leftarrow \text{auto\_train}\left[\{\mathbf{d}_1, \mathbf{d}_2, \ldots \mathbf{d}_n\}, \{y_1, y_2, \ldots, y_n\}\right]$

The descriptor is then assessed by using the trained classifier to detect objects from the set of test imagery. The test imagery consists of a number of images of size 240×240 pixels for which the ground truth location of vehicles are known. A 30×30 pixel sliding window is placed in one corner of the imagery and moved in 9 pixel steps across the full image. To detect vehicles at different scales, the test imagery is reduced in size and the 30×30 pixel sliding window is passed over the

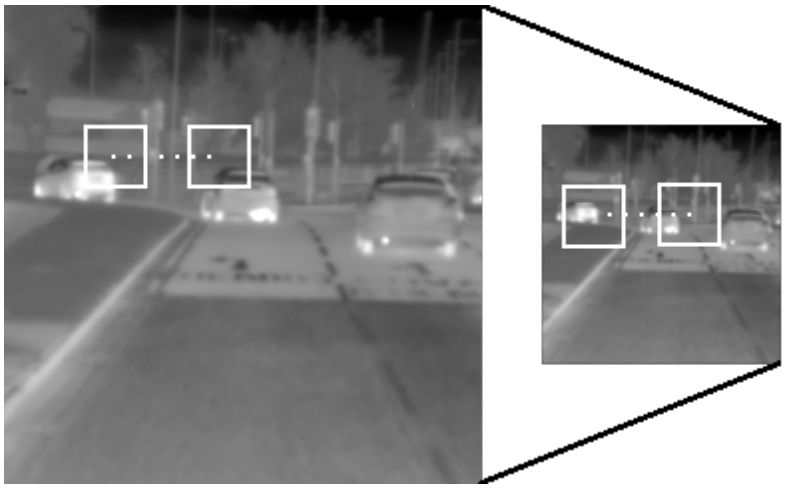


Figure 5.1: The input imagery is resized to a number of scales and for each scale a fixed size sliding window is passed spatially over the whole image. A descriptor is calculated and classified for each sliding window location.

image again (see Figure 5.1). This process is repeated over three octaves of scale[1] at four scales per octave to allow detections of objects between 30×30 pixels and 210×210 pixels. A descriptor is computed for each sliding window location, and this descriptor is classified with the trained SVM classifier.

The true identity of each sliding window location is confirmed by the PASCAL measure (4.12). The number of true positive, true negative, false positive, and false negative assessments are then recorded at a number of different distances to the SVM hyperplane. These statistics are used to calculate a ROC curve, as discussed in Section 4.4.1. The resulting ROC curves are then compared for different parameters and different descriptors to assess and compare detection performance.

5.2 Histogram of Oriented Gradients

The Histogram of Oriented Gradients (HOG) [56] descriptor was introduced earlier in Section 4.2.1. The algorithm, summarised in Algorithm 5.3, computes a descriptor from the spatial distribution of intensity gradients across the greyscale image. Gradient histograms are then accumulated into a number square cells of fixed size. The size of the cell determines the spatial granularity of the descriptor, with larger cells collecting gradient information from a larger spatial extent. For use with polarimetric imagery, classifiers trained with HOG features from each of the I, Q, U, P, and ϕ images were assessed.

[1] An octave is the scale factor between one image and another with half of its width and height.

Algorithm 5.3 Histogram of Oriented Gradients

INPUT: I input window, b number of histogram bins, c total number of cells
OUTPUT: h descriptor of length $4bc$

1: compute gradients for each pixel of image I
2: quantize gradient into one of b orientations in the range $[0, \pi)$
3: accumulate contribution to four neighbouring cell histograms based on gradient magnitude and distance from cell centre
4: define a block as a 2×2 sub-array of cells
5: calculate the norm of each block as the l^2 norm of the stacked histograms from the contained cells
6: calculate four normalisation factors for each cell as the inverse norm of the four blocks that contain the cell
7: the descriptor comprises four copies of each cell histogram, normalised using each of the four different normalisation factors.

5.2.1 Results

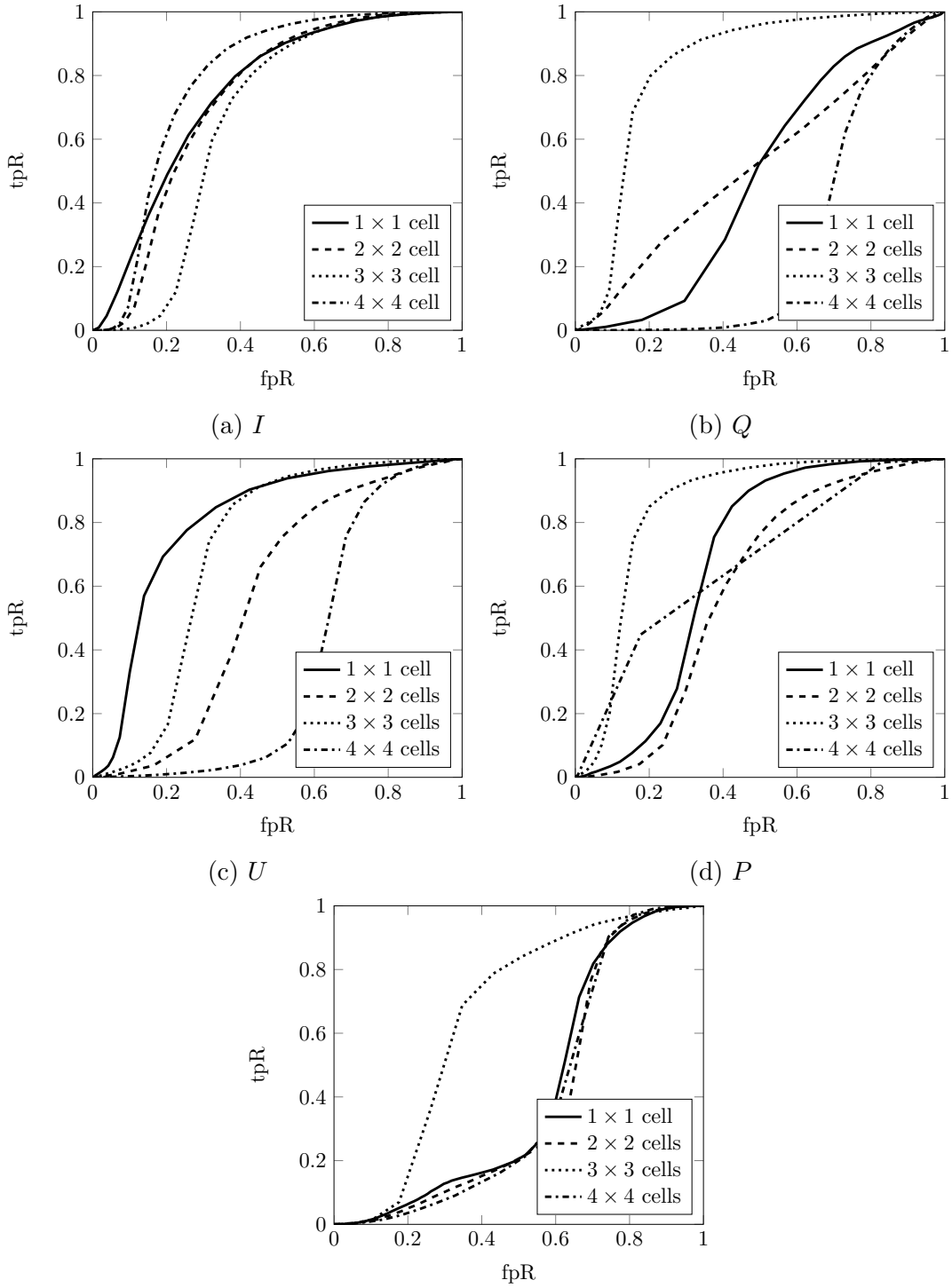
SVM classifiers were trained and tested using a standard HOG descriptor calculated with each of the I, Q, U, P, and ϕ images. For each image type (using an input window size of 30×30 pixels), classifiers were trained for windows spanned by 1×1, 2×2, 3×3 and 4×4 spatial cells. The results for each of these cases are shown in terms of ROC curves in Figure 5.2.

5.2.2 Discussion

The best performance is obtained for the P and Q imagery using 3×3 cells to span the window. The HOG algorithm identifies objects by learning the appearance of their gradient against the background. Referring back to some polarimetric imagery (for example Figure 3.18) it can be noticed that in both P and Q imagery, the vehicles do appear relatively distinct (and thus with a large gradient between the vehicle and the background).

From the P image a high degree of polarisation is observed from the vehicle surfaces, while the Q image highlights strong vertical polarisation. The U image is more noisy than the Q image, partly due to the sensor (in Section 3.4 it is mentioned that the 45° and 135° sensors have a lower polarisation contrast than the 0° and 90° sensors, which will also contribute to the noise). However, vehicles tend to have a dominant surface with a vertical surface normal which additionally lends itself to a stronger response from Q rather than U imagery. The ϕ image appears very noisy and so its poor performance is to be expected. The ϕ image is calculated as

$$\phi = \frac{1}{2} \tan^{-1} \left(\frac{U}{Q} \right) \tag{5.1}$$

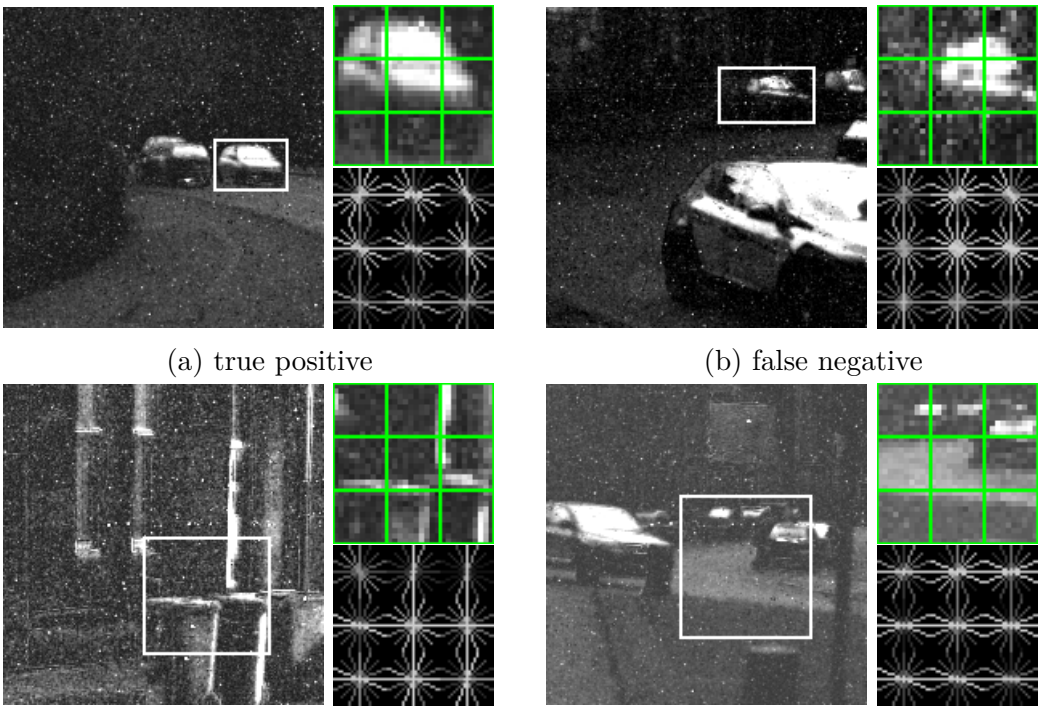


(a) I (b) Q

(c) U (d) P

(e) ϕ

Figure 5.2: ROC curve for HOG descriptor trained on different Stokes images using different numbers of cells.



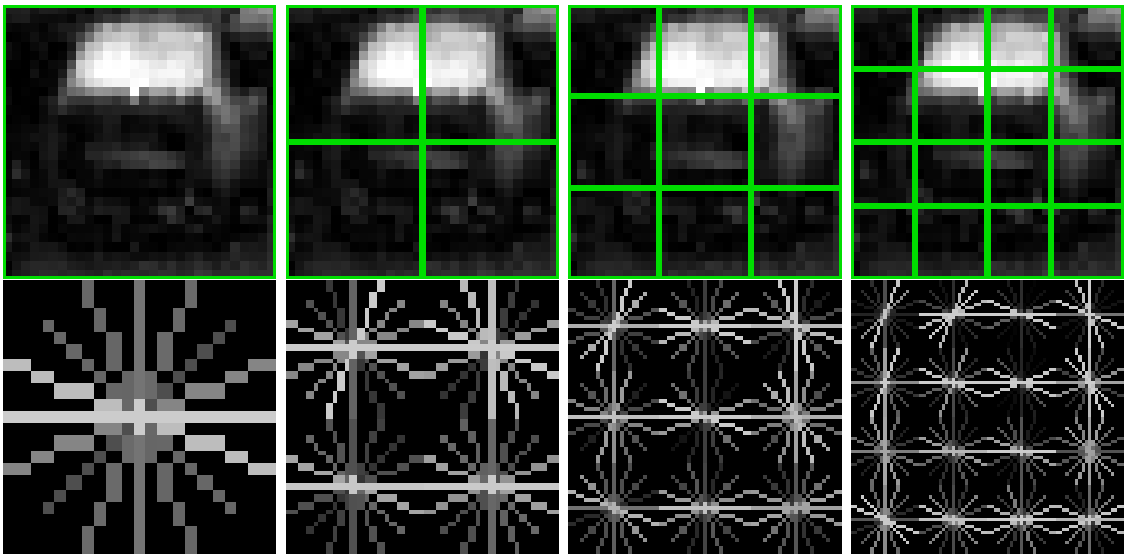
(a) true positive (b) false negative

(c) true negative (d) false positive

Figure 5.3: Output from the HOG classifier. In each figure, the image on the left is the full image with a marked bounding box. The upper right figure is the contents of the bounding box with the cells highlighted. The lower right figure is a graphical representation of the HOG descriptor for that image, with a histogram for each cell represented as a star. The brightness of each line in the star shape represents the frequency of gradients in that cell arising from edges with the same alignment as that line.

and so is very noisy when the degree of polarisation is low. Finally, from the I image it is observed that while the vehicles are fairly distinct there is a greater extent of background clutter than in the polarimetric images with every object in the scene having some form of thermal signature, thus making vehicles harder to distinguish.

In Figure 5.3 some sample output is shown from the 3×3 cell descriptor acting on P imagery. In this Figure, the trained detector is run on some test imagery, and the selected windows demonstrate examples of a true positive, false positive, true negative, and false negative detection. The full image is shown on the left with a bounding box surrounding the detection. On the right, the upper image shows an expanded view of the selected window, with the 3×3 arrangement of cells highlighted. The lower right image is a graphical representation of the HOG descriptor for this window. The gradients in each cell contribute to the gradient histogram, represented by the star shape in the cell location. If the cell contains strong gradients arising from a horizontal edge, the horizontal bar in the histogram will be more intense. From the true positive example, note that the car fills the entire window but only the top half is strongly polarised. The resulting descriptor



(a) 1×1 cell (b) 2×2 cells (c) 3×3 cells (d) 4×4 cells

Figure 5.4: Example of different HOG cell sizes. The top row indicates the spatial segmentation of a certain image into 1×1, 2×2, 3×3, and 4×4 cells, while the bottom row indicates the resulting gradient histogram for each cell.

thus shows a dominant orientation in each of the upper cells which follows the outline of the strongly polarised surfaces of the vehicle.

The number of cells has a large affect on the performance of the descriptor. Figure 5.4 shows the graphical representations of the P descriptor when using different numbers of cells for a particular vehicle sample. When the entire window is spanned with a single cell, the histogram representation cannot capture any spatial information from the samples, on which the HOG descriptor relies. Meanwhile, when the numbers of cells increases, the exact orientation of the vehicle within the window can cause the vehicle silhouette to move around and appear in different cells for different vehicle samples. These two factors suggest there is a trade-off where using fewer cells means losing spatial information while a greater number of cells cannot capture the variation in pose of different vehicle samples. When using P imagery, it would appear that the 3×3 arrangement of cells is most appropriate.

5.2.3 Summary

The results from using the Histogram of Oriented Gradients algorithm show that the polarimetric properties of vehicles do help distinguish them from the background, but the large variation in result with changes in cell size suggest that the HOG method does not make the best use of the information available.

5.3 Histogram of Polarisation Angle

The HOG descriptor builds histograms of gradient orientation, where each pixel location increases the appropriate histogram bin by a factor of its gradient magnitude. This has the result that strong gradients contribute significantly more to the resultant histogram than weaker gradients in that cell. The Histogram of Polarisation Angle (HPA) descriptor adapts the HOG approach to explicitly consider polarisation. The algorithm, described in Algorithm 5.4, assembles histograms based on quantised angles of polarisation, ϕ, and uses each pixel's degree of polarisation, P, as a weighting factor (in the original HOG, histograms are based on gradient angle, and use gradient magnitude as a weighting factor). This is intended to have the effect that locations which have a strong polarisation contribute to the descriptor with more weight than locations with a weaker degree of polarisation.

Algorithm 5.4 Histogram of Polarisation Angle

INPUT: P, ϕ input windows of degree and angle of polarisation, b number of histogram bins, c total number of cells

OUTPUT: h descriptor of length $4bc$

1: quantize angle of polarisation into one of b orientations in the range $[0, \pi)$
2: accumulate contribution to four neighbouring cell histograms based on degree of polarisation and distance from cell centre
3: define a block as a 2×2 sub-array of cells
4: calculate the norm of each block as the l^2 norm of the stacked histograms from the contained cells
5: calculate four normalisation factors for each cell as the inverse norm of the four blocks that contain the cell
6: the descriptor comprises four copies of each cell histogram, normalised using each of the four different normalisation factors.

5.3.1 Results

As with the HOG algorithm, one of the main parameters for this descriptor is the number of cells which span the image. With more cells, a finer representation of structure across the window can be captured, but with fewer cells a greater robustness to spatial arrangement can be tolerated. Figure 5.5 shows ROC curves for the Histogram of Polarisation Angle using 2×2, 3×3, 4×4, 5×5 and 6×6 cells spanning the window. The best result at a low false positive rate is obtained using 3×3 cells, while at higher false positive rates, the 5×5 and 6×6 arrangements show improved results.

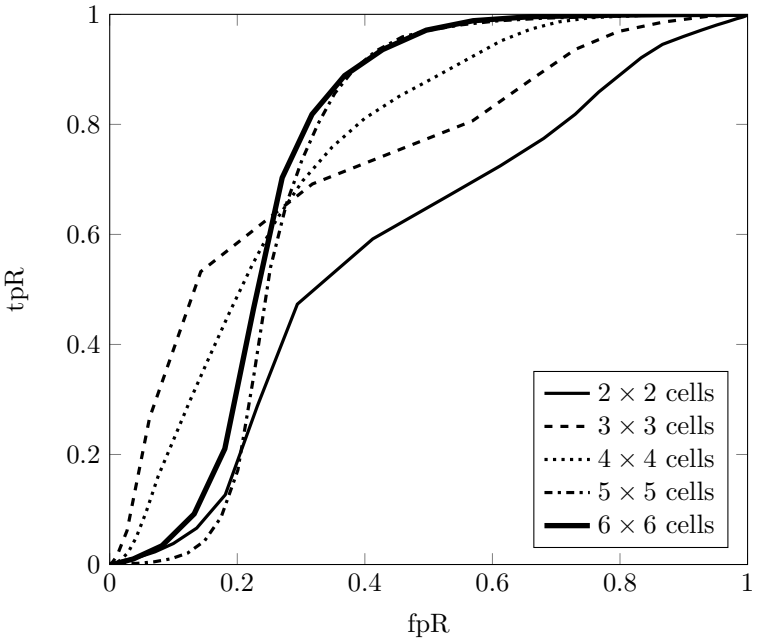


Figure 5.5: ROC for the Histogram of Polarisation Angle using different numbers of cells to span the window.

5.3.2 Discussion

In the current form, the HPA descriptor is not as robust as the HOG descriptor presented in the previous section. The HOG descriptor identifies objects by their silhouette against the background, while the HPA descriptor uses the polarimetric signature across the object. When vehicles are viewed from different orientations, the observed angle of polarisation can change significantly, leading to particular cells capturing a range of angles of polarisation depending on the pose of the vehicle.

Four positive and four negative windows are shown in Figure 5.6 along with a 3×3 cell glyph representing the 3×3 cell HPA descriptor. In this glyph, a histogram of angle of polarisation is collected for pixels within each cell, as marked in the images. The strong lines in the glyph correspond to the strong contributions from the angle of polarisation of the same orientation. Recall that each pixel contributes a magnitude based on the degree of polarisation to a particular angle of polarisation bin. Thus, for cells which predominately contain a single angle, particularly those with a strong degree of polarisation, a dominant histogram bin becomes apparent.

Pixels with a high degree of polarisation and a vertical angle of polarisation (bright red in the coloured image) appear as distinct vertical lines in the glyph, such as cells containing road surfaces, or the upper surfaces of vehicles. In contrast, regions with a low degree of polarisation (such as towards the bottom of Figure 5.6d) have little angular preference, and so all angles are equally weighted.

From the four positive samples shown, it is generally evident that the main

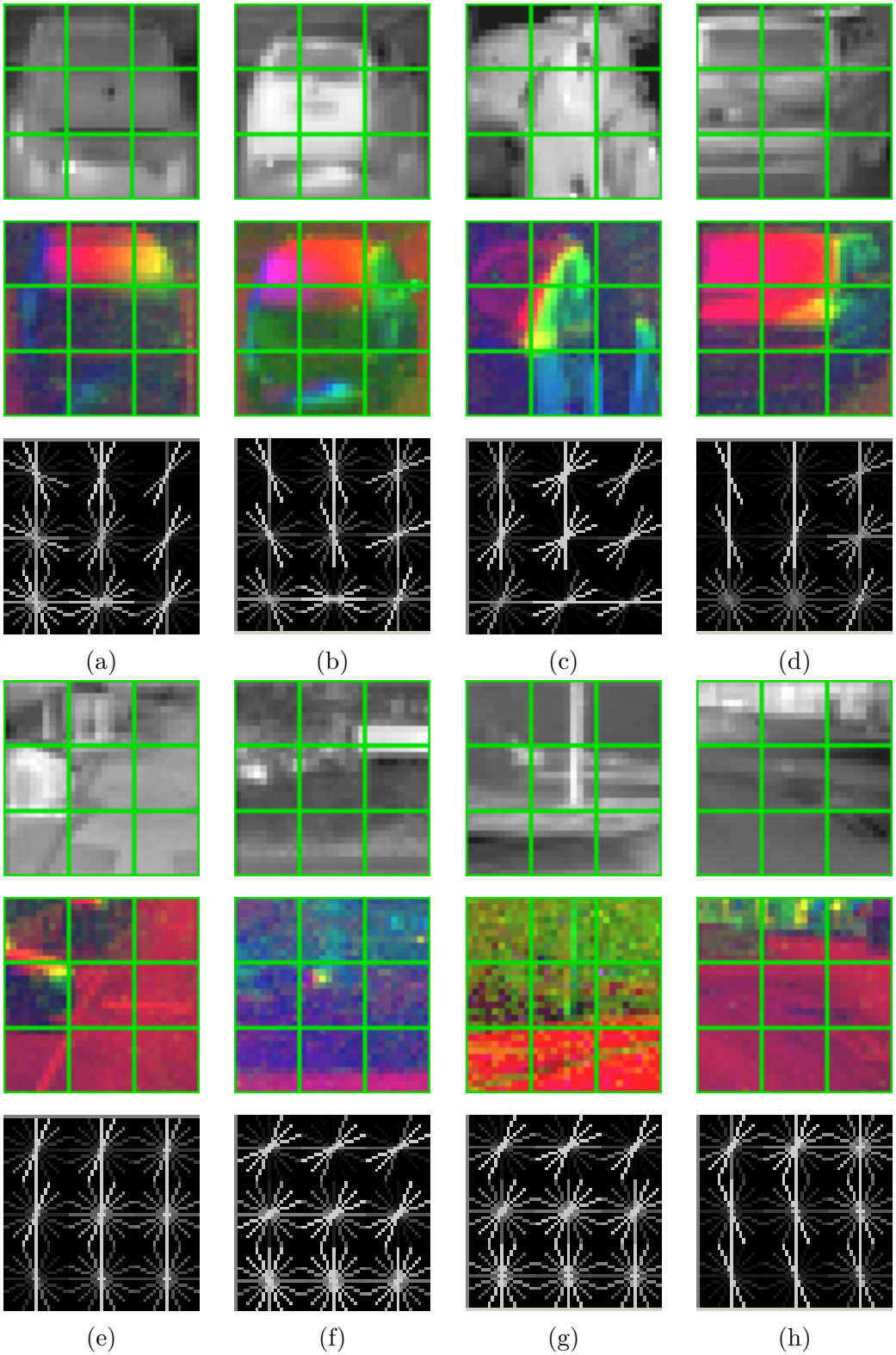


Figure 5.6: Examples of the Histogram of Polarisation Angle for (a)-(d) positive samples and (e)-(h) negative samples. In each case, the top image is the thermal intensity image, and the middle image is the false colour polarisation image, both of which have the 3×3 cell arrangement outlined. The bottom image shows a graphical representation of the feature, with each star representing an angle of polarisation histogram from the corresponding cell.

polarimetric content of descriptors calculated from the positive samples comes from the upper half of the window. The strong vehicle response arises from vertical angles of polarisation around the upper surface of the vehicle and towards the sides of the window the angle of polarisation rotates to reflect the surface normal of the vehicle sides. However, from the negative samples it is clear that there is a significant component of vertical polarisation arising from the road region (notice the strong vertical lines in Figures 5.6e and 5.6h and so the presence of predominantly vertical polarisation alone may not be enough to discriminate vehicles.

5.3.3 Summary

The Histogram of Polarisation Angle descriptor was developed to begin to capture the polarimetric signature of vehicles directly. The development led to the understanding that the pose of vehicles significantly affects the layout of polarisation angle observed.

5.4 Polarimetric Histogram

The Histogram of Polarisation Angles used the degree of polarisation to weight the contributions to the histogram while the histogram itself was based on angle alone. By retaining the degree of polarisation within the histogram representation, vehicles – which usually display high degree of polarisation – are hypothesised to be more separable. As such the Polarimetric Histogram (PH), summarised in Algorithm 5.5, instead forms a two dimensional histogram to record both P and ϕ, or alternatively the Stokes Q and U components.

Algorithm 5.5 Polarimetric Histogram

INPUT: (P, ϕ) or (Q, U) input windows, b_1, b_2 histogram bins in two dimensions, c total number of cells
OUTPUT: h descriptor of length $4b_1b_2c$

1: **for all** cells **do**
2: calculate two dimensional histogram from pixels within cell
3: **end for**
4: define a block as a 2×2 sub-array of cells
5: calculate the norm of each block as the l^2 norm of the stacked histograms from the contained cells
6: calculate four normalisation factors for each cell as the inverse norm of the four blocks that contain the cell
7: the descriptor comprises four copies of each cell histogram, normalised using each of the four different normalisation factors.

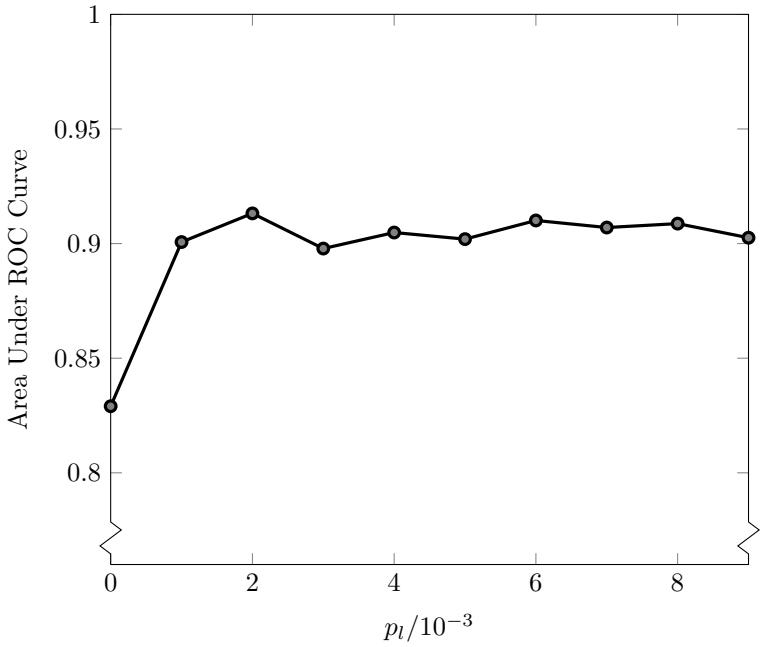
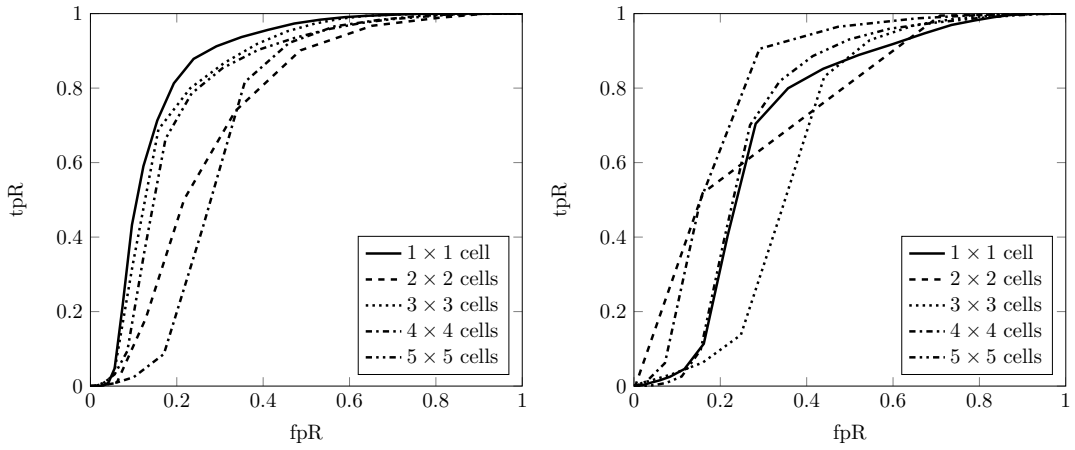


Figure 5.7: Area under the ROC curve with varying p_l, the lower limit on degree of polarisation considered.

5.4.1 Results

When computing a (P, ϕ) histogram, b_P bins were used to span P in the range $[p_l, p_u)$ and b_ϕ bins were used to span ϕ in the range $[-90°, 90°)$. The angle of polarisation is naturally bounded and so the range was easy to choose, but the absolute value of the degree of polarisation depends on a number of scene factors and so was more difficult to set. As vehicles were present at higher degree of polarisation it was advantageous to remove pixels with very low degrees of polarisation. Figure 5.7 shows the area under the ROC curve (a measure of performance) for the Polarimetric Histogram descriptor when using different values of p_l. It clearly shows the advantage of a non-zero p_l but shows limited performance gain for higher values of p_l and so a value of $p_l = 0.001$ was chosen. Based on the typical maximum degree of polarisation in the datasets, a value of $p_u = 0.03$ was chosen. If developed further, it may be advantageous to investigate how to algorithmically set these limits based on the current frame statistics.

The (Q, U) histogram was computed using b_Q and b_U bins for the Q and U parameters respectively. However, the histogram bounds were also difficult to choose for similar reason to the degree of polarisation, as the absolute values of these parameters depend on the scene. Based on typical ranges of these parameters within images in the dataset, the lower bound for both Q and U was set to -50 and the upper bound to 50.



(a) (Q, U) (b) (P, ϕ)

Figure 5.8: ROC curves for the Polarimetric Histogram with varying sizes of cell, and eight bins for each histogram dimension.

Number of cells Both (Q, U) and (P, ϕ) variants of the Polarimetric Histogram were trained and tested with eight bins for each dimension and different numbers of cells spanning the window, the ROC curves from which are shown in Figure 5.8. For the (Q, U) case improved results can be seen when using an odd number of cells, while for the (P, ϕ) case the results are less clear.

Number of bins Using a single cell to span the detection window, the Polarimetric Histogram was tested when using 6, 8, and 10 bins for each dimension for both (Q, U) and (P, ϕ) variants. The resulting set of ROC curves is shown in Figure 5.9. For both methods, the best performance is obtained when using ten bins, then six, and eight bins resulting in the poorest performance. This ordering is thought to be due to the angles at which the bin boundaries are defined, if a particular number of bins causes a surface orientation to overlap a bin boundary then this may affect the robustness of the descriptor. The (P, ϕ) histogram is shown to outperform the (Q, U) histogram for both six and ten bin cases.

5.4.2 Discussion

To fully evaluate the PH descriptor, a number of additional experiments would be required, to test every combination of parameter and discover any underlying trends. However, initially the best result were obtained for the (P, ϕ) histogram with ten histogram bins in each dimension and a single spatial cell. Figure 5.10 shows the detection results from an SVM classifier trained with these parameters on a number of unseen images. A sliding window approach is used where a 30×30 pixel window was moved in spatial steps of 5 pixels and scaled spatially over three octaves,

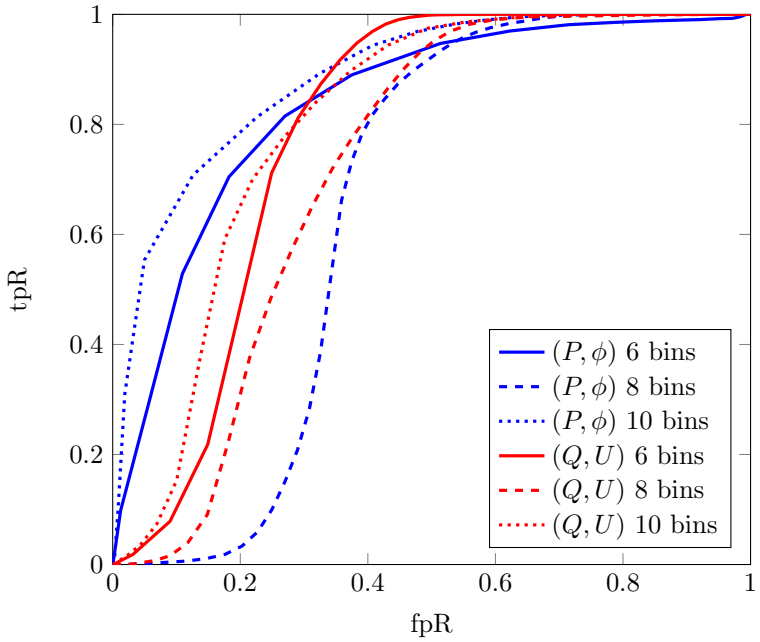
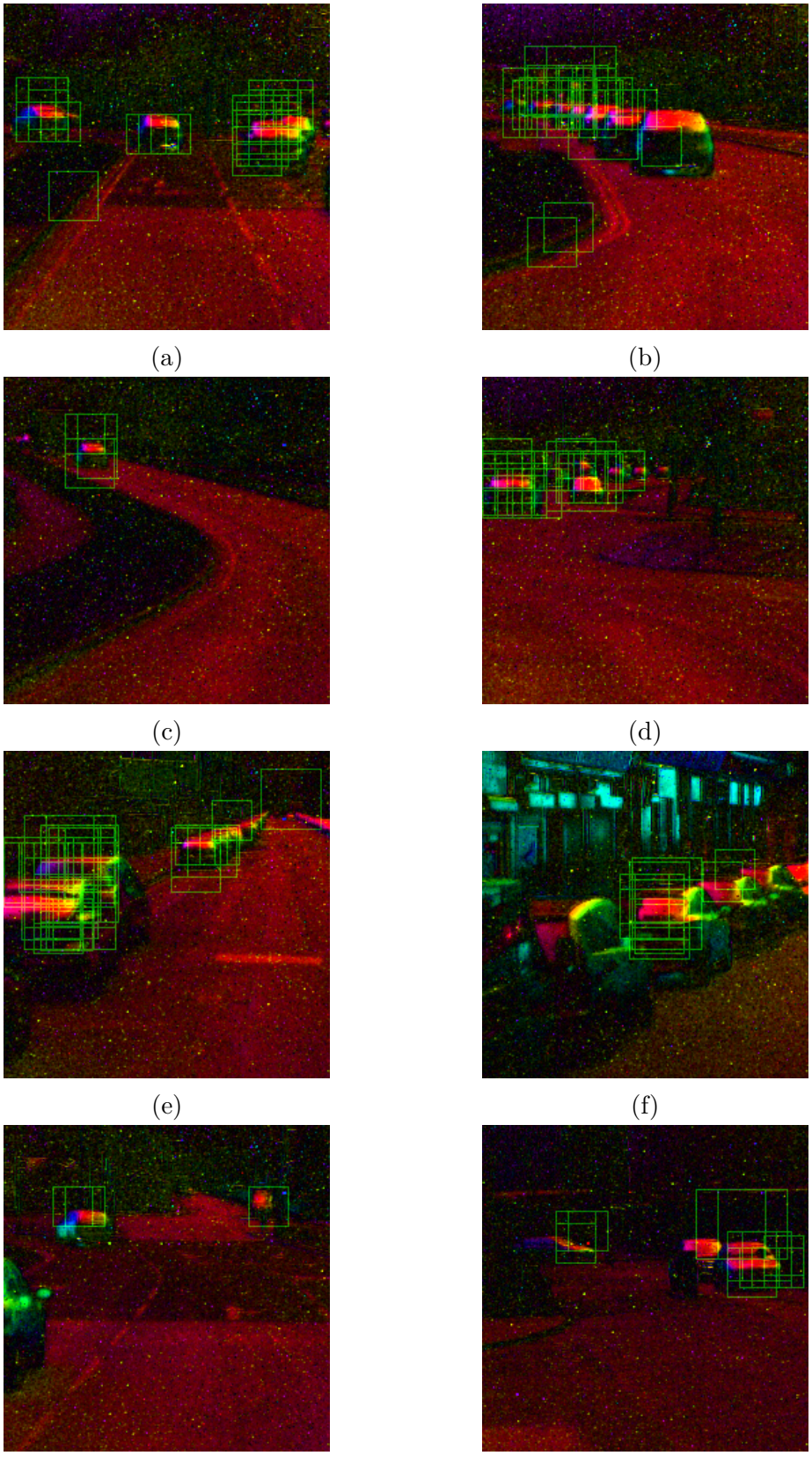


Figure 5.9: ROC curves for the Polarimetric Histogram descriptor with a single cell spanning the input window and different numbers of histogram bins (with equal numbers of bins in for each histogram dimension).

with four scales per octave, allowing detections of objects between 30×30 pixels and 210×210 pixels. Each window location was evaluated with the trained classifier and windows classified as vehicles are marked. No subsequent non-maximal suppression was applied to the detections, hence the large number of overlapping detections.

This output shows generally good performance, with bounding boxes around or near most vehicles in the scene. Most detections only span part of each vehicle (typically the higher degree of polarisation regions), which suggests that the detector mainly responds to high polarisation regions. However, it is encouraging to note that in Figure 5.10f no false positive detections are triggered on the shop fronts which display a strong horizontal polarisation characteristic, while some of the partially occluded vehicles in this image do trigger detections.

Individual polarimetric histograms The polarimetric histograms computed for a selection of windows are shown in Figure 5.11. The vehicle samples have a strong histogram response at a higher degree of polarisation compared to the negative samples. Also, the strong response within certain angular bins for each positive sample, due to each approximately flat vehicle surface polarising at an approximately constant angle is encouraging. However it is clear that the angular component of the polarimetric histogram is not always consistent. Indeed, as vehicles are observed from different orientations, different surface orientations mean that different angles of polarisation are observed, leading to significantly different descriptors.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5.10: Example output frames from the Polarimetric Histogram descriptor, parameters as described in text.

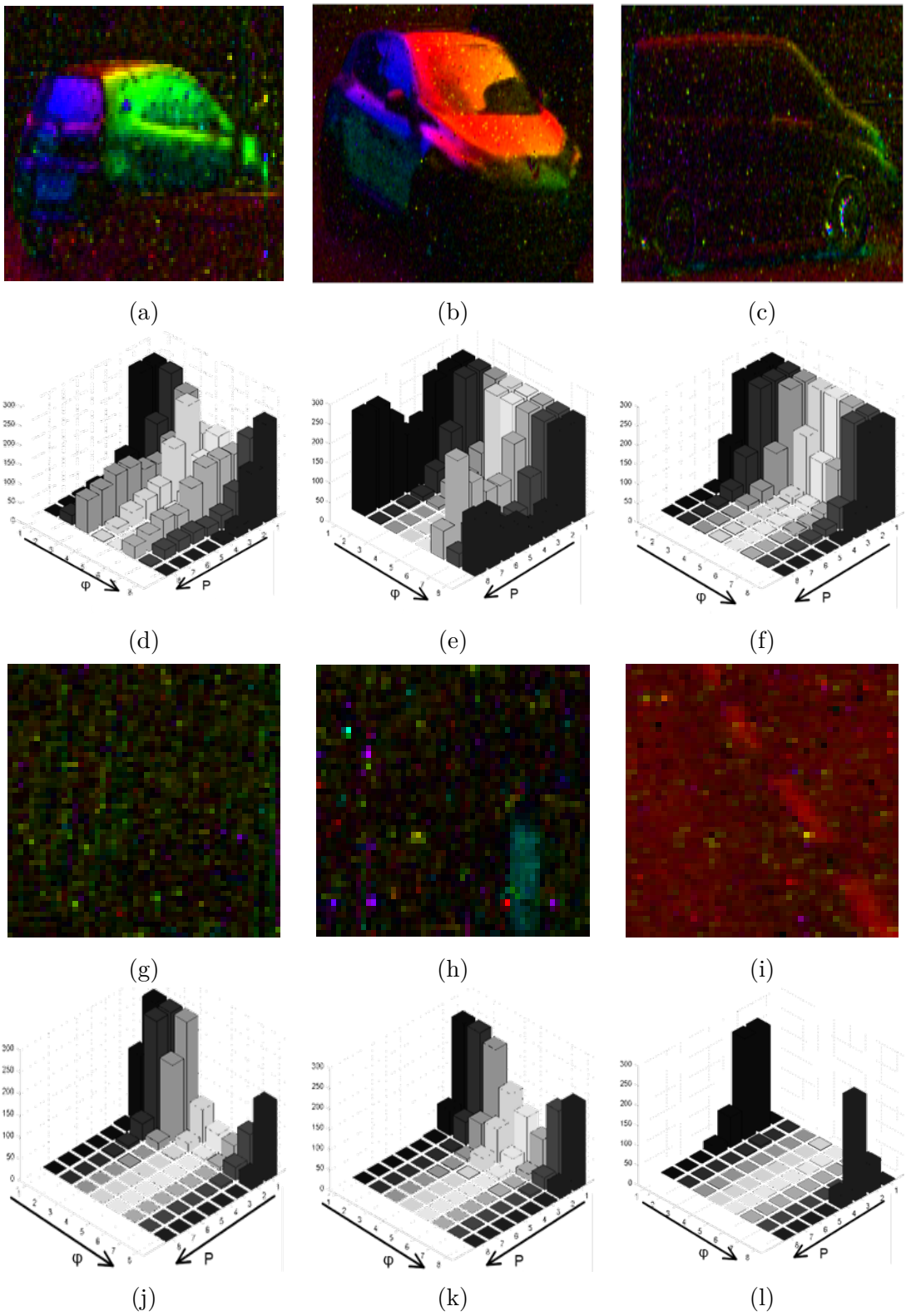
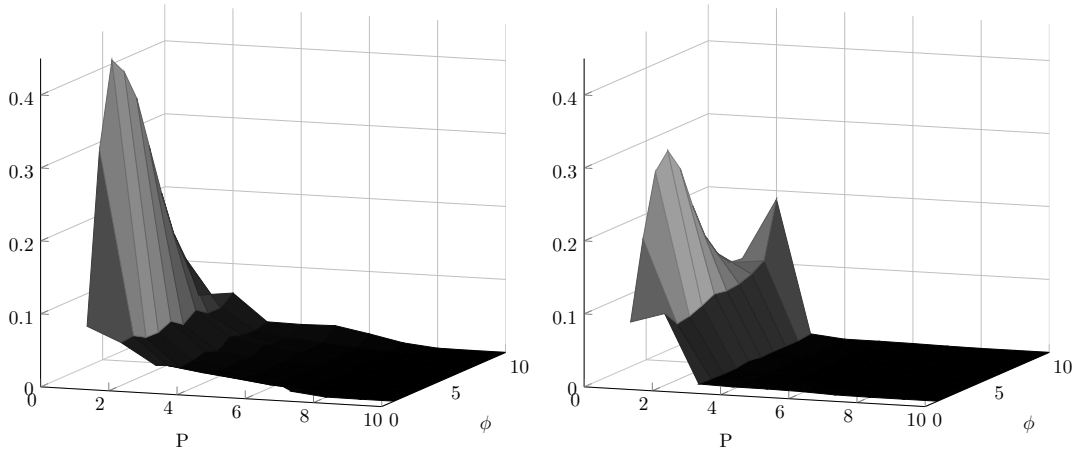


Figure 5.11: (a),(b),(c) images of positive samples; (d),(e),(f) corresponding positive polarimetric histograms; (g),(h),(i) images of negative examples; (j),(k),(l) corresponding negative polarimetric histograms.



(a) Positive (b) Negative

Figure 5.12: Mean (a) positive and (b) negative Polarimetric Histogram descriptors from training data in DS2.

Mean polarimetric histograms The mean positive and negative polarimetric histogram computed from the training data in DS2 are shown in Figure 5.12. The mean positive histogram shows a significant number of pixels within a few angular bins, while for the mean negative histogram a much flatter angular distribution is observed. Additionally, while the positive histogram shows a higher contribution from high degree of polarisation pixels, the specific angular dependences of the higher degree of polarisation pixels has been smoothed out.

5.4.3 Summary

In an initial evaluation of the PH descriptor over a selection of parameters, reasonable results were obtained for the (P, ϕ) histogram when using ten bins for each dimension and a single cell to span the window. By examining selected windows and their associated histogram representations, it was suggested that the degree of polarisation was a distinctive measure, allowing some separation of vehicle from non-vehicle. Additionally, pixels from each vehicle surface tend to be observed at a uniform angle of polarisation, equating to histograms peaked in certain angular bins for each example. However, as noted previously, the pose of the vehicle is strongly linked to the angles of polarisation observed. Using a single cell, no spatial information is retained in the descriptor, only the overall distribution of degree and angle of polarisation. This goes some way towards removing pose variation for vehicles with similar orientation, but also removes all information about the structural format of the polarimetric signature, including the spatial proximity of points with similar angles of polarisation.

5.5 Polarimetric Histogram with Centre of Mass

The best result for the Polarimetric Histogram method was found to use (P, ϕ) imagery in a single cell across the detection window. Without breaking the window into spatial cells, the spatial pattern of the polarimetric content of the window is lost. However, looking at examples of vehicles it would appear that the spatial distribution of polarisation is important, with high polarisation regions containing clusters of consistent angle of polarisation. Knowing that the angle of polarisation is related to the surface normal of the polarising surface, and considering vehicles as convex objects, it is reasonable to assume that the angle of polarisation visible at any particular location across an object will precess to some extent from $-90°$ towards the left of the vehicle to $90°$ towards the right of the vehicle. As such, in an attempt to compute the Polarimetric Histogram using a single cell, while retaining a representation of the spatial distribution of the angle of polarisation across the window, a centre of mass component was included in the computation of the descriptor, which is hypothesised to identify a left to right precession of polarisation angle across the image for positive samples.

The Polarimetric Histogram with Centre of Mass consists of a two dimensional (P, ϕ) histogram computed over a single spatial cell along with the spatial coordinate of the centre of mass, $\mathbf{R}$, of the degree of polarisation for pixels within each ϕ bin. This can be calculated as

$$\mathbf{R} = \frac{\sum_{i=1}^{n} P_i \mathbf{r}_i}{\sum_{i=1}^{n} P_i} \tag{5.2}$$

where the sums are over pixels with ϕ within the current angular bin, P_i is the degree of polarisation of pixel i, and $\mathbf{r}_i$ is the spatial position of pixel i within the window divided by the size of the window such that $(0, 0) \leq \mathbf{r}_i < (1, 1)$. The resulting descriptor can be visualised in Figure 5.13 the rows of which are then concatenated to form a $b_\phi \, (b_P + 2)$ dimension descriptor.

5.5.1 Results

A subset of DS2 was used to train and test the Polarimetric Histogram, using a single spatial cell, and ten bins for each of the P and ϕ dimensions for both the original case as discussed in Section 5.4 and with the additional centre of mass coordinate of each angular bin. The comparative ROC curve is shown in Figure 5.14, revealing that the addition of the centre of mass coordinate has very little effect on the accuracy of the Polarimetric Histogram method.

	$P(1)$	$P(2)$	$\ldots$	$P(b_P)$	R_x	R_y
$\phi(1)$						
$\phi(2)$						
$\vdots$						
$\phi(b_\phi)$						

Figure 5.13: Example of the Polarimetric Histogram with Centre of Mass. The main section is a regular two dimensional P, ϕ histogram, with the addition of the spatial location of the centre of mass of the degree of polarisation for pixels within that ϕ bin.

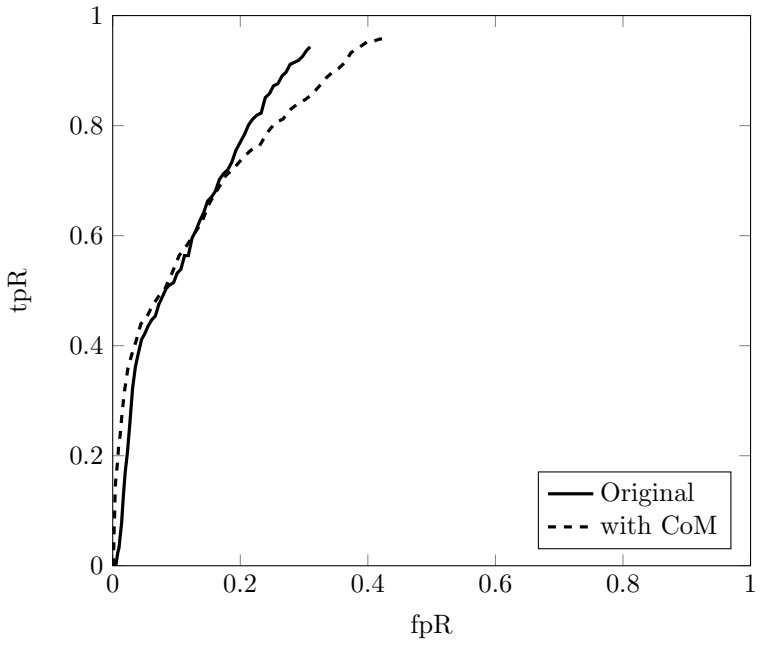


Figure 5.14: Comparison of the Polarimetric Histogram method using ten bins for each P and ϕ dimensions and a single cell with and without the center of mass component.

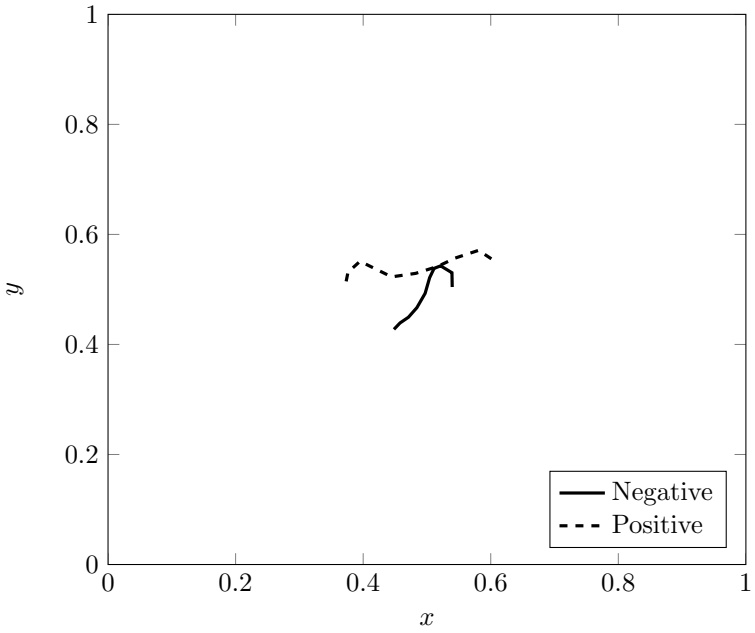


Figure 5.15: Graph of mean positions of pixels within each angular bin for positive and negative datasets. The positive samples do show a left to right pattern as the angle of polarisation increases, but this in not too dissimilar from the negative pattern.

5.5.2 Discussion

To analyse the centre of mass coordinate in more detail, the average positive and negative descriptors were computed over all training samples, and the mean x and y positions of each angular histogram bin extracted and plotted in Figure 5.15. The hypothesis was that due to the convex nature of vehicles, the mean position of each angular bin for positive samples will move left to right across the window, while for negative samples a more randomly position will be observed. Indeed, this is partially evident in Figure 5.15, however, the difference between positive and negative is smaller than anticipated. The positive training windows contain vehicles within background, and so each histogram bin is likely to contain contributions both from the vehicle and from some of the background. In addition, some positive samples do not follow this pattern (consider Figure 5.11c for example) which will also dilute the effect.

However, Figure 5.15 does reveal that the mean y component for all positive angular bins is further towards the top of the window than for the negative samples, which is consistent with the understanding that upper surfaces of vehicles are more strongly polarising than the lower surfaces.

5.5.3 Summary

The spatial position of the polarimetric signature is important, but it appears that there is too much variation throughout the vehicle class to summarise the appearance with a single histogram type representation. In attempting to represent vehicles in this way, much of the distinctive information is lost.

5.6 Histogram of Polarisation Angle Differences

The angle of polarisation observed for a particular surface of a vehicle is dependent on the direction from which it is viewed. As observed with the Polarimetric Histogram, this results in different angles of polarisation being observed for vehicles at different orientations. Invariant factors which could distinguish a vehicle from the background include

1. a high degree of polarisation compared with the background;
2. an approximately constant angle of polarisation for pixels on a particular vehicle surface;
3. a constant difference in angle of polarisation between pixels on two different surfaces of a vehicle.

The Polarimetric Histogram was able to exploit (1) by comparing samples along the P dimension alone, but building a histogram from the absolute angle of polarisation did not satisfactorily address (2) or (3) as different vehicles presented different angles resulting in significantly different descriptors.

By instead considering the difference in angle of polarisation, $\Delta\phi$, for pairs of pixels within the window, (2) would lead to positive samples showing a large component at $\Delta\phi \approx 0$ from pixel pairs from the same surface, and (3) would result in a significant contribution from $\Delta\phi \approx \phi'$ from pixel pairs from different vehicle surfaces. Thus building a histogram based on the difference in polarisation angle for pairs of pixels in the window was considered. To incorporate (1) in a similar manner to the Histogram of Polarisation Angle discussed previously, the contribution of each pixel pair was weighted by the combined degree of polarisation of that pair. These considerations led to the Histogram of Polarisation Angle Differences (HPAD), which is summarised in Algorithm 5.6.

5.6.1 Results

This descriptor was trained and tested on a reduced version of DS2 using four bins and sixteen bins, the resulting ROC curves are shown in Figure 5.16. Using four

Algorithm 5.6 Histogram of Polarisation Angle Differences

INPUT: b number of histogram bins, ϕ and P image windows
OUTPUT: h normalised histogram of polarisation angle differences

1: $h = (0, \ldots, 0)$ ▷ Zero initialised histogram length b
2: **for** $i = 1 \rightarrow$ number of pixels **do**
3: **for** $j = 1 \rightarrow$ number of pixels **do**
4: $\Delta\phi = \phi_i - \phi_j$ ▷ Calculate polarisation angle difference
5: $w = P_i P_j$ ▷ Calculate weighting factor
6: $h\left(\texttt{bin}\left(\Delta\phi\right)\right) \mathrel{+}= w$ ▷ Accumulate appropriate bin by w
7: **end for**
8: **end for**
9: $h \leftarrow \frac{h}{||h||}$ ▷ Normalise descriptor with l^2 norm

bins the performance is little better than chance, but improved performance is seen when using sixteen bins.

Rather than comparing ROC curves for additional numbers of histogram bins, a set of descriptors was computed using thirty bins to attempt to identify any signatures within this descriptor. This representation is an approximation of the continuous case, and was created such that any descriptor formed using fewer bins would consist of a subset of this result. The descriptor was computed over a subset of DS2 containing 1,000 positive and 1,000 negative training samples. Positive and negative descriptors are plotted in Figure 5.17 where the solid line indicates the median descriptor over all samples, and the dotted lines indicate the 0.25 and 0.75 percentiles which gives an indication of the overlap between positive and negative descriptors. Figure 5.17a shows the actual value of the descriptors, and Figure 5.17b shows the descriptors with each coordinate (histogram bin) normalised to have zero mean and a standard deviation of one (over all positive and negative samples).

5.6.2 Discussion

From Figure 5.17 it can be observed that the positive samples do indeed have a more consistent strong component at $\Delta\phi \approx 0^\circ$ than the negative samples. However, the second predicted feature arising from pixels on two different vehicle surfaces is not obvious. This suggests that the consistent angle of polarisation on each vehicle surface is a strong feature but the different in angle between two surfaces visible on a vehicle is not necessarily consistent for pairs of surfaces on different vehicles or from different points of view. To exploit this angle of polarisation difference between two surfaces, it may be more important to identify the existence of a pair of surfaces, each with a reasonably uniform angle of polarisation, rather than rely on the angle difference directly.

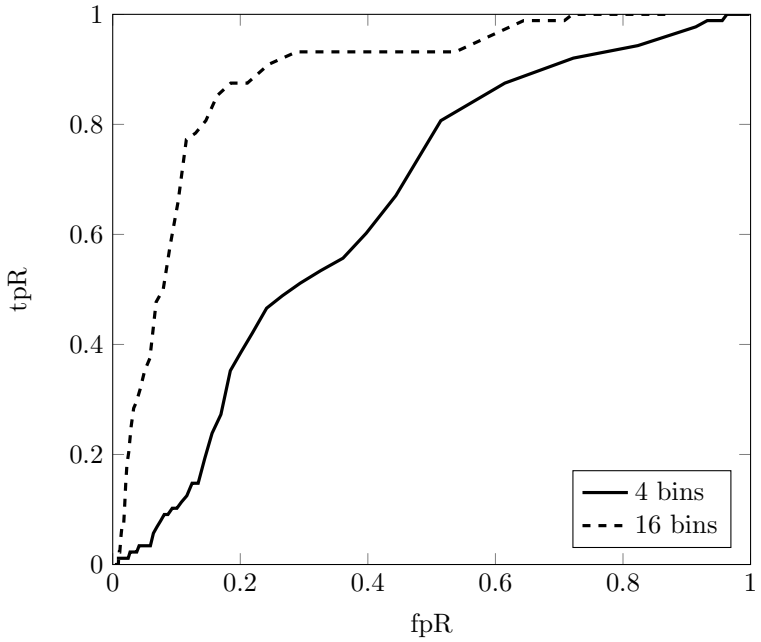


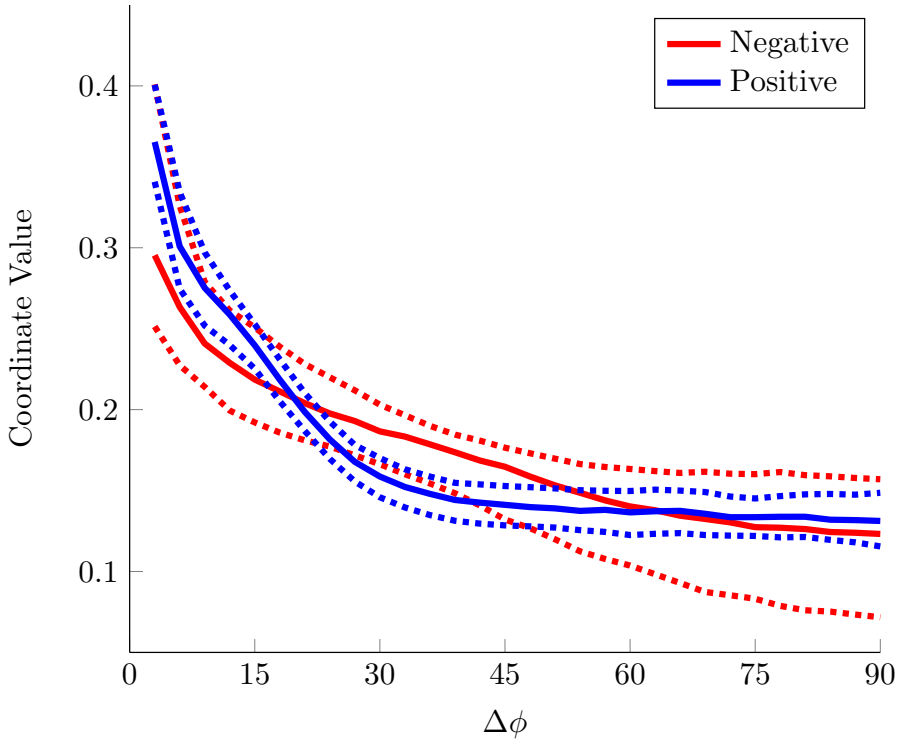
Figure 5.16: Comparison of choice of number of bins with the Histogram of Polarisation Angle Differences.

5.6.3 Summary

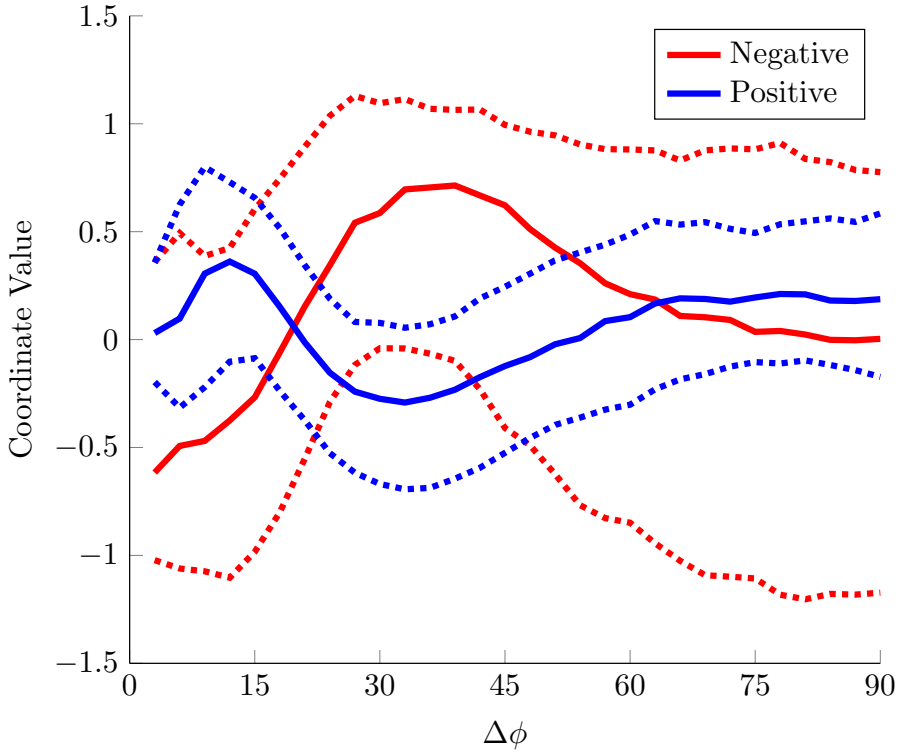
The HPAD descriptor helped develop the understanding that the polarimetric signature of vehicles can be summarised as consisting of a higher degree of polarisation than the background, approximately consistent angle of polarisation for pixels on each vehicle surface, and in some cases pairs of surfaces on the same vehicle with some offset in angle of polarisation depending on their relative orientation and the pose of the vehicle.

5.7 Discussion of the Histogram-based Approach

This chapter has explored the LWIR polarimetric signature of vehicles through the development of a series of histogram based descriptors. Vehicles are found to appear with a much higher degree of polarisation than the background, and a number of the descriptors have been proposed which aim to exploit this factor as the principal distinguishing feature. It was increasingly found that the pose of the observed vehicle with respect to the camera significantly altered the polarisation signature, with different orientations resulting in different angles of polarisation being observed. Towards the end of this chapter a descriptor based around the difference in angle of polarisation across the window (rather than the absolute angle of polarisation) was proposed. This method removed the reliance on the absolute angles of polarisation, which would change with the pose of the observed vehicle.



(a) Actual coordinate values.



(b) Coordinate values normalised to mean 0 and standard deviation 1.

Figure 5.17: Histogram of Polarisation Angle Differences descriptors over a set of 1,000 positive and 1,000 negative training samples from DS2. The solid line connects the median coordinate value for each bin, and the dashed lines connect the 0.25 and 0.75 percentile coordinate values. The x axis denotes the $\Delta\phi$ bin from 0° to 90°.

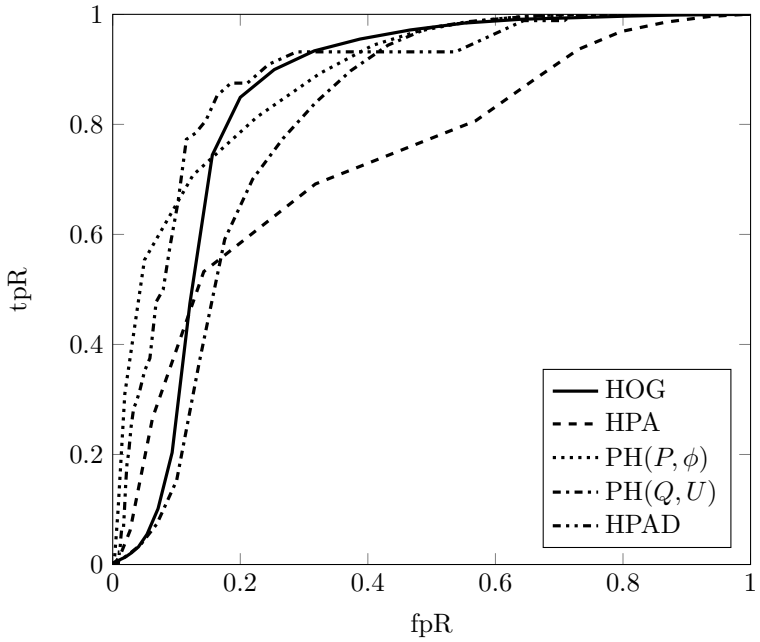


Figure 5.18: Comparison of all histogram based methods tested. HOG: Histogram of Oriented Gradients method on the P image with 3×3 cells; HPA: Histogram of Polarisation Angle using 3×3 cells; PH(P, ϕ): Polarimetric Histogram over (P, ϕ) with a single cell and 10 bins for each dimension; PH(Q, U): Polarimetric Histogram over Q, U with a single cell and 10 bins for each dimension; HPAD: Histogram of Polarisation Angle Differences using a 16 bin histogram.

A comparison of the best performing version of each histogram-based descriptor is shown in Figure 5.18. The best performance so far is achieved using either the Polarimetric Histogram method on (P, ϕ) (Section 5.4) or the Histogram of Polarisation Angle Differences method (Section 5.6). The main issue with a number of these methods, the Polarimetric Histogram and the Histogram of Polarisation Angle Differences included, is the lack of spatial information from the content of the window. Vehicles consist of flat surfaces at fixed orientation, and so pixels on a certain surface will not only have a high degree of polarisation and consistent angle of polarisation, but they will also be in close proximity to each other. In order to improve upon this, an alternative representation of the vehicle signature needs to be considered.

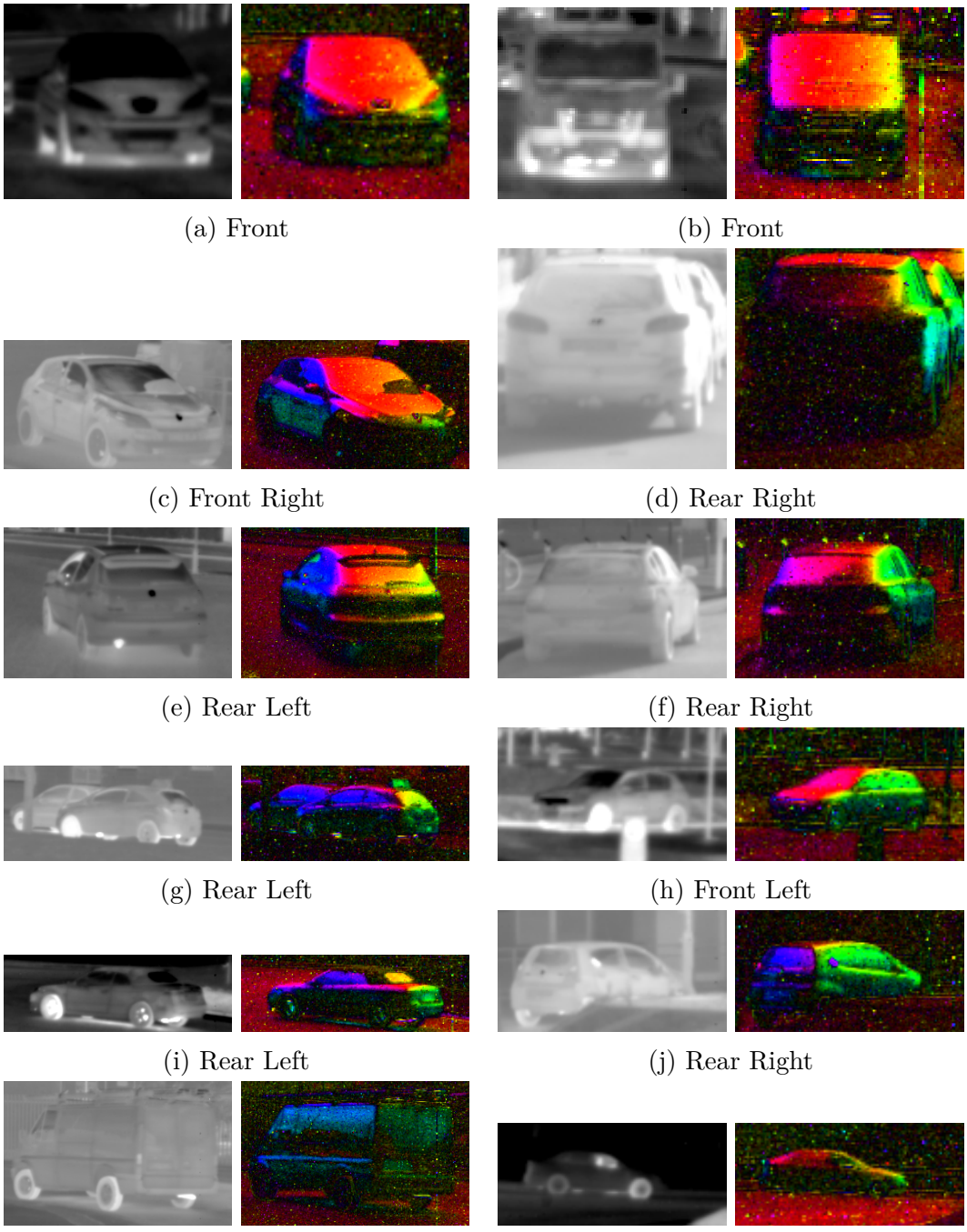
Chapter 6
Cluster-based Polarimetric Vehicle Descriptor

This chapter builds upon the findings of Chapter 5 and considers descriptors derived from clusters formed from the polarimetric data. These cluster-based descriptors are shown to further improve the performance of vehicle detection.

6.1 Introduction

From the material presented in Chapter 5 it can be said that vehicles are constructed from a set of strongly polarising surfaces, each fairly uniform in orientation and thus the observed angle of polarisation for a particular surface is largely uniform. Furthermore, while the pose of the vehicle with respect to the camera will alter the angle of polarisation observed for any particular surface, the relationship between surfaces depends on the 3-D vehicle shape, rather than on the pose. The following cluster-based approach aims to extract clusters of high degree of polarisation and consistent angle of polarisation, and build a descriptor based on the relationship between the extracted clusters. This differs from the histogram-based approach by allowing more flexibility in the spatial position of different surfaces through exploiting the high degree of polarisation of vehicles to extract dominant pixels from the vehicle surfaces.

Due to self-occlusion, only a subset of the vehicle surfaces will be visible at any particular vehicle-camera pose. With reference to Figure 6.1 notice that a vehicle observed from the front or rear appears to largely consist of a single vertically polarised region from the roof/bonnet; while when both the side and the front/rear of a vehicle can be observed, each surface appears to consist of a largely uniform angle of polarisation, with the absolute angle of polarisation depending on the geometry of the particular surface with respect to the camera. Specifically, compare Figure 6.1c with Figure 6.1h, referring to Figure 3.13 on page 34 to link colour with angle of



(a) Front (b) Front

(c) Front Right (d) Rear Right

(e) Rear Left (f) Rear Right

(g) Rear Left (h) Front Left

(i) Rear Left (j) Rear Right

(k) Van (l) Side

Figure 6.1: Example of vehicles when observed from different orientations. In each case, the image on the left is the LWIR intensity image, and the image on the right is the coloured representation of LWIR Polarisation, with colour relating to angle of polarisation as described in Section 3.6.

polarisation. The windscreen/bonnet region in both cases is approximately vertically polarised, while the vehicle sides appear polarised at angles of $\pm 45^\circ$ due to the different orientation.

In the following, Section 6.2 discusses each stage of the detector algorithm, Section 6.3 outlines the SVM training and testing process, some results are presented in Section 6.4 and Section 6.5 presents a discussion on the performance of this approach.

6.2 Cluster-based detection method

A detection window, I, contains n points, $I = \{p_1, p_2 \ldots, p_n\}$, where each point, p_i, is characterised by the location in the window, (x_i, y_i), as well as the angle and degree of polarisation, (ϕ_i, P_i), such that$p_i = (x_i, y_i, \phi_i, P_i)$. A threshold selects points with a high degree of polarisation. These are clustered according to their position and angle of polarisation, and the descriptor is constructed from the relation between these extracted clusters. In contrast to the Histogram-based approach, this cluster-based algorithm allows flexibility in the spatial positioning of the different surfaces, and removes the dependence on the exact angle of polarisation of each cluster.

6.2.1 Threshold

The objective of the cluster based approach is to detect clusters of consistent polarisation state, relating to approximately flat surfaces on vehicles. Making the assumption that pixels within vehicles tend to have a high degree of polarisation when compared to the background pixels, then when a window contains a vehicle, the degree of polarisation will separate it from the background while when a window contains only background a fairly uniform low degree of polarisation is observed. Therefore, as a first step in attempting to select vehicle pixels (if present) a threshold, P_t, is applied based on the degree of polarisation. Points with high polarisation are then used as input to the clustering stage, I_{in},

$$I_{\text{in}} = \{p_i \in I : P_i > P_t\}\,. \tag{6.1}$$

As discussed in Section 5.4 regarding selecting appropriate ranges for P in the Polarimetric Histogram, and from the theory of Section 2.4, it is known that the absolute degree of polarisation in a particular scene is hard to quantify and predict as it will depend on a number of conditions such as the temperature of the object and optical background, and the properties of the polarising surface. For these reasons, a fixed absolute threshold would not be appropriate for this stage in the

algorithm, and so instead it is proposed that P_t is chosen based on the relative degree of polarisation within a particular window in such a way that a fixed fraction of pixels from the window with the highest degree of polarisation are selected.

6.2.2 Clustering Method

The next stage in the algorithm is to cluster points in I_{in} based on their spatial proximity and angle of polarisation. Some of the points within I_{in} may not be part of a vehicle and will need to be ignored or removed, while some points covering the vehicle may not fit with this model and so will also need to be removed. As such, some form of additional outlier rejection will be needed to choose appropriate clusters.

Additionally, as the number of unoccluded clusters spanning a vehicle will depend on the pose of the vehicle, the clustering algorithm will either need to algorithmically choose or be guided to an appropriate number of clusters. Two different clustering methods are examined: a hierarchical method which partitions all input points into a hierarchy of different numbers of clusters, and a method based around the $k-$means algorithm, which requires the number of clusters, k, as an input.

For the majority of cases it is observed that vehicles may be characterised by one or two clusters, i.e. $k = 1$ or $k = 2$. Arguably there are cases where $k > 2$, as vehicles are generally composed of surfaces having normals in 3 principal directions. However, with data collected from a ground-level camera these cases are rare. This discussion is expanded upon when the results are discussed in Section 6.5.1.

Hierarchical clustering method

One approach tested for clustering points within a window involved initial outlier rejection followed by hierarchical clustering, and is discussed here. Hierarchical clustering [80, Section 14.3.12] is an agglomerative clustering method where initially all points are considered individual clusters. For each step in the algorithm, a distance measure is computed between all clusters and the two clusters with the smallest distance are combined, this process is then iterated until all clusters have been combined. The output from this method is a tree of cluster combinations which can be displayed in the form of a dendrogram (see Figure 6.2 for an example), where cluster index is depicted along the horizontal axis, and the vertical axis measures the total intra-cluster distance. This method requires the definition of a suitable distance function, and a method of comparing the distance between two clusters (for example the mean/median distance between points in each cluster or the longest/shortest distance between any points in two clusters).

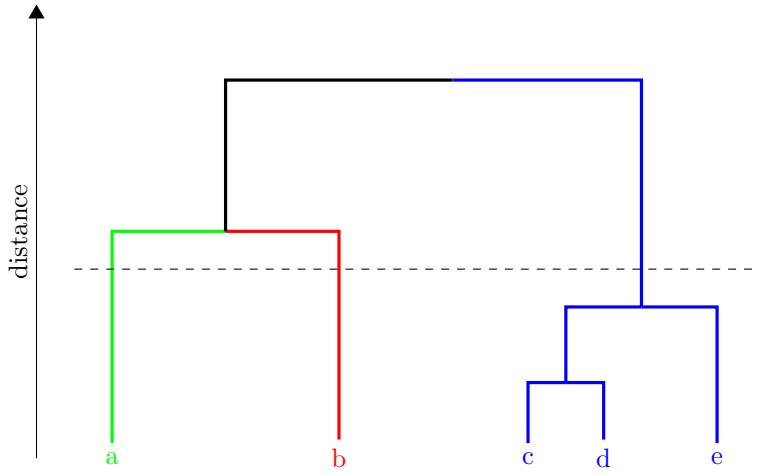


Figure 6.2: An example dendrogram. At the bottom of the dendrogram each point is in its own cluster. Clusters are joined in order of minimum distance until all points have been merged into a single cluster at the top of the dendrogram. The longer the vertical line joining two clusters, the more dissimilar they are. Notice how the user can choose to extract any number of clusters from the data by choosing a horizontal distance threshold.

When using the hierarchical clustering method, an initial outlier rejection method was implemented to remove points less likely to be of interest. The distance function used to compute the separation between points was also chosen to take into account the cyclical nature of the angle of polarisation, and weight the different components. Additionally, different linkage methods used in the cluster process to compute the distance between clusters were also explored. Each of these are discussed in the following paragraphs.

Outlier rejection method Hierarchical clustering uses all input data points to construct a hierarchical set of clusters. However, after a threshold is applied based on the degree of polarisation, some higher degree of polarisation points remain which are not of interest, either due to noise, or small sets of higher degree of polarisation points. To remove some of these additional points, the method of Almeida *et al.* [106] was followed, as summarised in Algorithm 6.1. This method considers outliers as points with low connectivity in opposition to higher connectivity points in the intra-cluster region. The connectivity, $c_i\,(R)$, of a point p_i, is a measure of the number of neighbouring points within a radius R to that point. Figure 6.3 compares the clustering results using hierarchical clustering on points before and after this outlier rejection method. From the lower images the effect of the outlier rejection process is obvious, with many of the background points remaining after the initial threshold being removed. This advantage is further confirmed by comparing the dendrograms, where the intra-cluster distances for the two-cluster segmentation are smaller after

the outlier rejection, suggesting that the clusters are tighter and more representative of the data.

Algorithm 6.1 Outlier Rejection for Hierarchical Clustering

INPUT: x input data
OUTPUT: x_{inliers}

for $j = 0 \rightarrow 2$ **do**
  $r \leftarrow 4 \times 2^{-j}$
  repeat
    Calculate $\bar{d}$ ▷ Average nearest-neighbour distance
    $R \leftarrow r\bar{d}$
    Calculate $\bar{c}(R)$ ▷ Average connectivity
    Discard object i if $c_i(R) < \frac{1}{3}\bar{c}(R)$
  until no objects discarded
end for

Distance function As points will be clustered in (x, y, ϕ), an extension to the standard Euclidean distance function must be considered in order to combine the linear x, y distances with the cyclic ϕ distances, and also to appropriately weight the three dimensions. The angle of polarisation, ϕ, spans $-90°$ to $90°$ in a cyclic manner, and so the absolute distance, $\Delta\phi$, between two angles ϕ_1 and ϕ_2 is calculated as

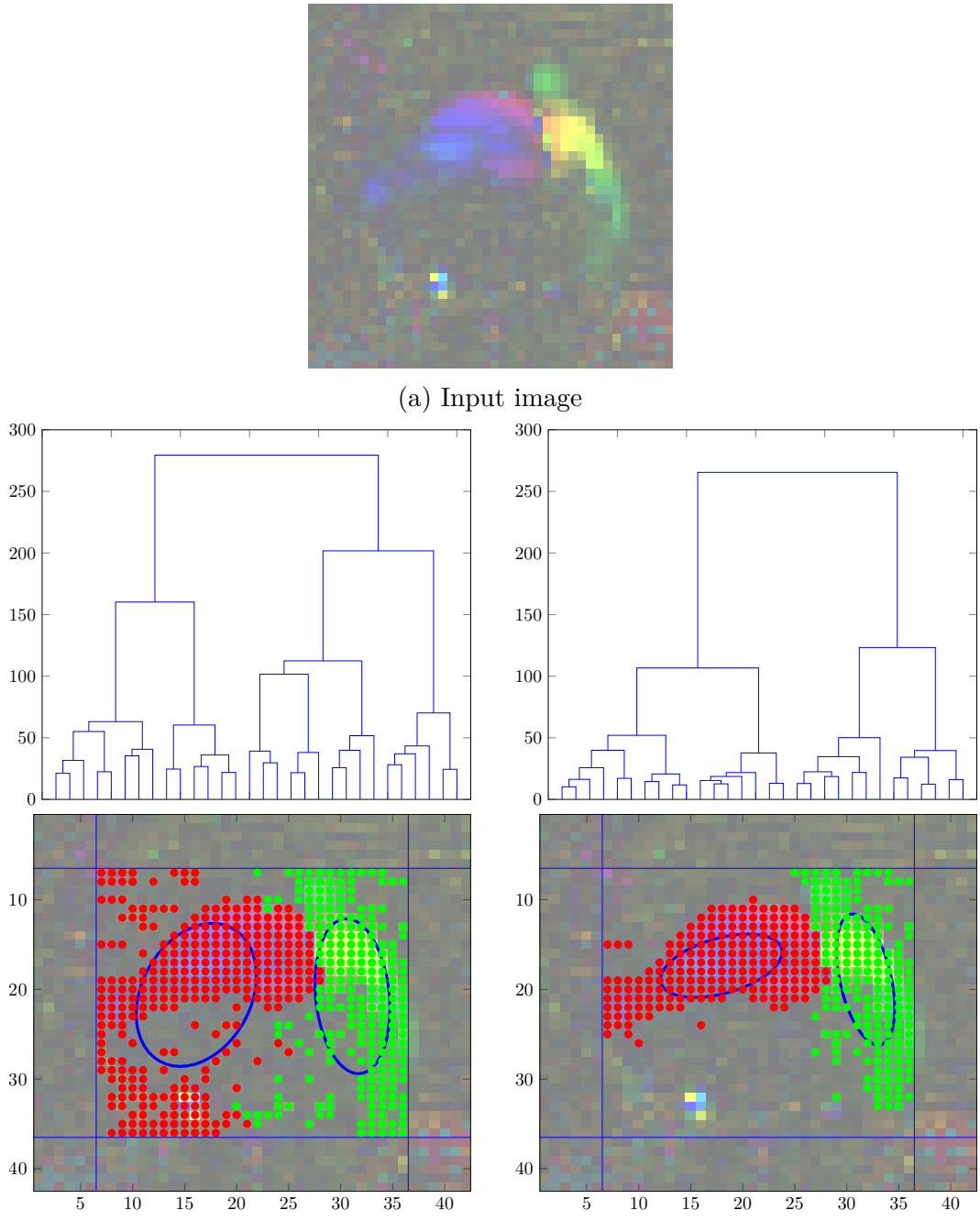
$$\Delta\phi = \min\left(\left|\phi_1 - \phi_2\right|, \left|\phi_1 - \phi_2 - 180°\right|, \left|\phi_1 - \phi_2 + 180°\right|\right) \tag{6.2}$$

where $|\cdot|$ indicates the absolute value. The distance between the x and y coordinates is simply their absolute difference ($\Delta x = |x_1 - x_2|$, $\Delta y = |y_1 - y_2|$). Combining these, the total distance between two points is calculated as

$$d = \sqrt{\left(w_x \Delta x\right)^2 + \left(w_y \Delta y\right)^2 + \left(w_\phi \Delta\phi\right)^2} \tag{6.3}$$

where (w_x, w_y, w_ϕ) is the set of normalisation/weighting factors. The input image is sized 30×30 pixels and so Δx and Δy can span 0 to 29; while $\Delta\phi$ can span $0°$ to $90°$. To normalise the coordinates to equal ranges, weighting factors of $w_x = \frac{1}{30}$, $w_y = \frac{1}{30}$, and $w_\phi = \frac{1}{90°}$ were chosen.

Linkage method In each stage of the Hierarchical clustering algorithm, the two clusters which are closest to each other are combined, as measured by the average linkage method. This method defines the distance between two clusters as the average distance between all pairs of points in the two clusters, with each distance calculated with (6.3). This is a convenient measure of distance which doesn't require any subsequent distance calculation after each pair of distances has been calculated,



(a) Input image

(b) After P threshold

(c) After outlier rejection

Figure 6.3: Example of outlier rejection stages. (a) Input window. (b) output from clustering based on points selected by initial P threshold; (c) output from clustering based on points remaining after outlier rejection. Note that the two clusters, indicated by red and green points are much tighter (the ellipses represent the 1σ standard deviation of each cluster), and the dendrograms above show more separation between the clusters after outlier rejection. Note that the images are of a vehicle in the $x - y$ plane, with the colour representing the ϕ coordinate.

while taking into account the location of each point within the cluster, and avoiding the calculation of the cluster centroid.

Robust $k-$means clustering method

An alternate clustering method, robust $k-$means was also considered. This method was based on the standard $k-$means algorithm, summarised in Algorithm 6.2, which partitions a set of points into k clusters so as to minimise the sum of within-cluster distances. As an extension, an additional outlier rejection method is used to iterate the algorithm after marking any points further than s standard deviations from their cluster centroid as outliers until no new outliers are defined. The robust $k-$means varient used was originally a subfunction from EEGLAB [107] and is summarised in Algorithm 6.3.

Algorithm 6.2 $k-$means

INPUT: x input data, k number of clusters, n number of iterations
OUTPUT: x partitioned into k clusters

1: **for** $i = 1 \rightarrow n$ **do**
2: Initialise k random cluster centres, $\mu_1, \ldots, \mu_k$
3: **repeat**
4: Assign x_i to nearest cluster centre μ_j
5: Recalculate μ_j as centroid of x in cluster j
6: **until** x assignments do not change
7: Record cluster results
8: **end for**
9: Return result with smallest within-cluster sums of point-to-centroid distances

Clustering using $k-$means requires a sensible definition of the mean of a set of points in order to work well. However, with the cyclic ϕ coordinate, where the angles $\phi = -90°$ and $\phi = 90°$ are identical, the mean of a set of points is ambiguous. For example the mean, μ_ϕ, of two points $\phi_1 = -70°$ and $\phi_2 = 80°$ would normally be calculated to be $\mu_\phi = 5°$ even though a value of $\mu_\phi = -85°$ would be more appropriate in this case. To highlight this, the front-view of a vehicle is shown in Figure 6.4a where points remaining after the initial outlier rejection method are marked. The $x-\phi$ projection of these points, as shown in Figure 6.4b, highlights the $90°$ to $-90°$ discontinuity. To overcome this, angles $\phi < 0°$ are referred to by their equivalent $\phi' = \phi + 180°$ coordinate, as shown in Figure 6.4c. While this method simply moves the location of the discontinuity to $0°$, this is considered acceptable due to the convex shape of vehicles. Surfaces on the left of a vehicle will tend to have a surface normal in the range $\phi' = 0°$ to $\phi' = 90°$, surfaces at the top of a vehicle will tend to have a surface normal approximately vertical ($\phi' = 45°$ to $\phi' = 135°$), and surfaces on the right of a vehicle will have a surface normal in the range $\phi' = 90°$ to

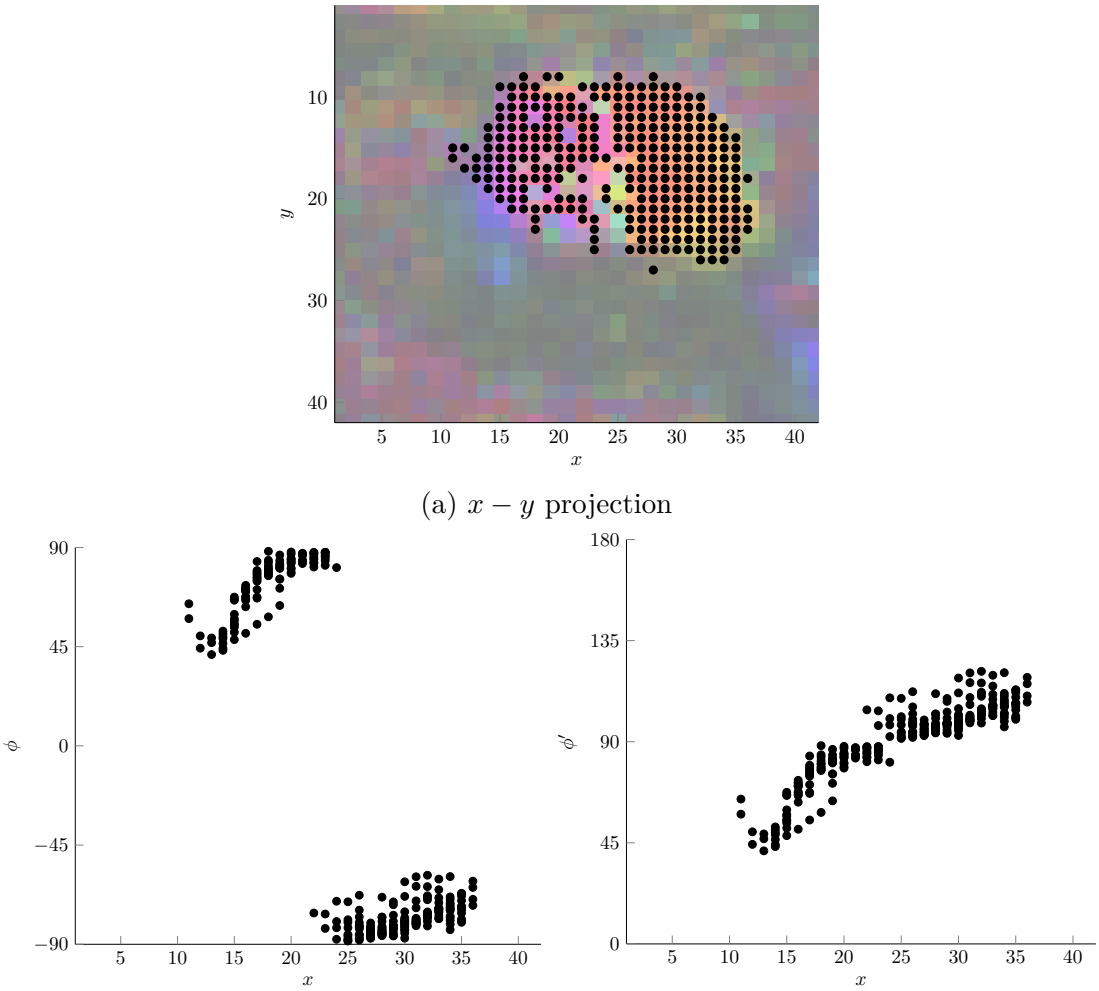
Algorithm 6.3 Robust $k-$means

INPUT: x input data, k number of clusters, s outlier distance
OUTPUT: x partitioned into k clusters

$x_{\text{inliers}} \leftarrow x$
repeat
  idx $\leftarrow$ `kmeans` (x_{inliers}, k) ▷ Run standard kmeans on inliers
  Compute standard deviation of each cluster, $\sigma_1, \ldots, \sigma_k$
  for all $x_i \in x_{\text{inliers}}$ **do**
    if $x_i - \mu_{\text{idx}(i)} > s\sigma_{\text{idx}(i)}$ **then** ▷ If x_i considered an outlier
      Move x_i from x_{inliers} to x_{outliers}
    end if
  end for
  for all $x_i \in x_{\text{outliers}}$ **do**
    for $j \leftarrow 1, k$ **do** ▷ For each cluster
      if $x_i - \mu_j < s\sigma_j$ **then** ▷ If x_i now considered an inlier
        Move x_i from x_{outliers} to x_{inliers}
      end if
    end for
  end for
until Assignments do not change

$\phi' = 180°$. Thus, having transformed the ϕ coordinate to ϕ', the assumption that points on a single vehicle surface will not span the discontinuity is made and as such the mean and variance remain good measures with which to define the cluster. Before the clustering process the ϕ coordinate is weighted by $\frac{1}{6°}$ to give it equal weighting with the spatial coordinates (both now spanning the range 0 to 30).

The only parameter required for robust $k-$means, in addition to the number of clusters, is s which controls the definition of an outlier in the algorithm. With reference to Algorithm 6.3 recall that a point is defined as an outlier if it is located further than s standard deviations from the cluster mean. If s is too small, the majority of points will be marked as outliers during the clustering process, making later description of the cluster less reliable as only based on a few points. Conversely, if s is too large, any outlier points belonging to the background region or noise may remain included in the cluster, also detracting from the cluster characteristics. This is highlighted in Figure 6.5 where a sample input image and the clustering results for a range of s values are shown. In these plots, all input (high polarisation) points are marked with a black circle if the robust $k-$means algorithm marks them as outliers, or with a coloured cross to show their inclusion in a cluster. At $s = 1.2$ very few points are classed as inliers at the end of the process, while for $s = 2.8$ the clusters are made up from almost every available point, including those some distance away. When comparing these plots, it is worth remembering that the third ϕ axis is also used in the clustering process, although is only represented through colour in these

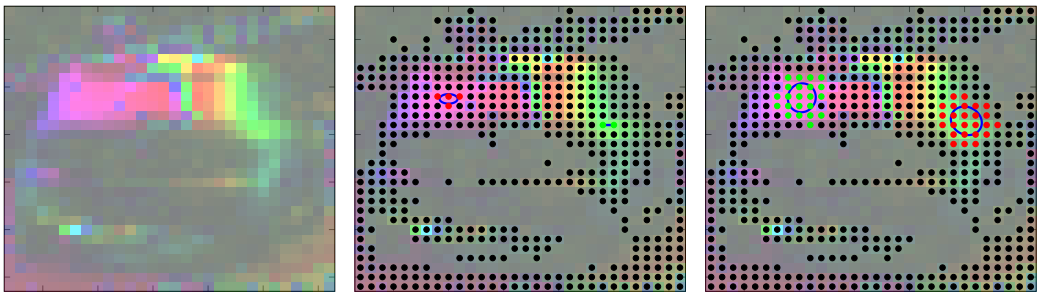


(a) $x - y$ projection

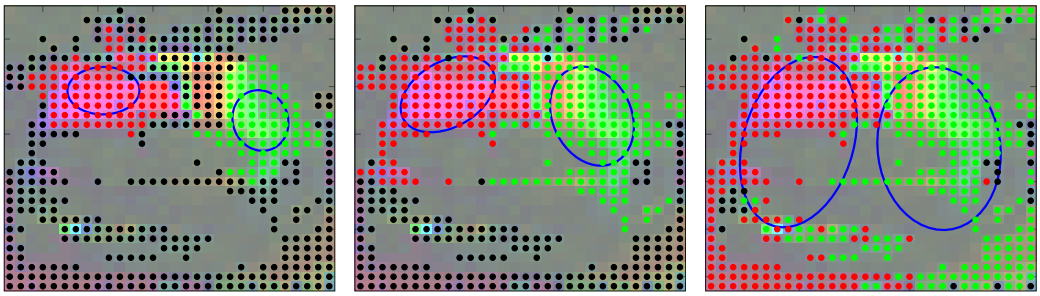
(b) $x - \phi$ projection

(c) $x - \phi'$ projection

Figure 6.4: (a) shows an image of a vehicle observed front-on. Each pixel is coloured according to the angle, and the points which have been marked as inliers are marked with black dots. (b), (c) show the $x-\phi$ projection and $x-\phi'$ projection respectively. Note that the degree of polarisation on the vehicle surface spans a discontinuity at $\phi = 90°$ which is overcome by defining ϕ over the range $0°$ to $180°$ rather than $-90°$ to $90°$.



(a) Input (b) $s = 1.2$ (c) $s = 1.6$



(d) $s = 2.0$ (e) $s = 2.4$ (f) $s = 2.8$

Figure 6.5: Example of robust $k-$means clustering with different values for the s parameter. Red and green marks indicate points included in each of the two clusters. Black marks indicate points marked as outliers. The 1σ standard deviation Gaussian distribution of each cluster is drawn.

plots.

By training and testing the full method using different s values on a larger set of samples, it was found that a value of $s = 1.5$ was preferable. This allowed most outlying points to be filtered out fairly effectively. While this means that fewer points are used to describe the final cluster, the tighter cluster was found to perform better for subsequent feature extraction.

6.2.3 Feature vector

The output from the clustering stage is a set of inlier points and their associated cluster index. From this, the feature vector was formed for use in the classification algorithm. For the majority of cases observed, vehicles could be characterised by either a single cluster, or two clusters. It is possible that cases exist for a greater number of clusters - as discussed later - but these were rare in the dataset available.

A particular cluster, $c^{(n)} = \left\{p_i^{(n)}\right\}$, can be characterised by its mean, $\mu^{(n)}$, and covariance, $\mathbf{\Sigma}^{(n)}$. The covariance matrix over the x, y, and ϕ parameters is defined

as

$$\begin{aligned}\mathbf{\Sigma}^{(n)} &= \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{x\phi} \\ \Sigma_{yx} & \Sigma_{yy} & \Sigma_{y\phi} \\ \Sigma_{\phi x} & \Sigma_{\phi y} & \Sigma_{\phi\phi} \end{bmatrix} \\ &= \begin{bmatrix} E\left[(x-\mu_x)(x-\mu_x)\right] & E\left[(x-\mu_x)(y-\mu_y)\right] & E\left[(x-\mu_x)(\phi-\mu_\phi)\right] \\ E\left[(y-\mu_y)(x-\mu_x)\right] & E\left[(y-\mu_y)(y-\mu_y)\right] & E\left[(y-\mu_y)(\phi-\mu_\phi)\right] \\ E\left[(\phi-\mu_\phi)(x-\mu_x)\right] & E\left[(\phi-\mu_\phi)(y-\mu_y)\right] & E\left[(\phi-\mu_\phi)(\phi-\mu_\phi)\right] \end{bmatrix}\end{aligned} \tag{6.4}$$

where $E(\cdot)$ indicates the expected value, and the cluster index has been dropped for convenience. Note that the covariance matrix is symmetric, such that $\Sigma_{xy} = \Sigma_{yx}$.

For convenience, the cluster mean is summarised in vector form as

$$\mu^{(n)} = \left(\mu_x^{(n)}, \mu_y^{(n)}, \mu_\phi^{(n)}\right) \tag{6.5}$$

and the cluster covariance as

$$\Sigma^{(n)} = \left(\Sigma_{xx}^{(n)}, \Sigma_{yy}^{(n)}, \Sigma_{\phi\phi}^{(n)}, \Sigma_{xy}^{(n)}, \Sigma_{x\phi}^{(n)}, \Sigma_{y\phi}^{(n)}\right). \tag{6.6}$$

One-cluster feature The 10-dimension single cluster feature vector, f_1, is defined to be

$$f_1 = \left(\mu^{(1)}, \Sigma^{(1)}, \frac{\mu_P^{(\mathrm{c})}}{\mu_P^{(\mathrm{s})}}\right) \tag{6.7}$$

where the term $\frac{\mu_P^{(\mathrm{c})}}{\mu_P^{(\mathrm{s})}}$ is a centre-surround metric which compares the mean degree of polarisation of all points within the window centre, $\mu_P^{(\mathrm{c})}$, to the mean degree of polarisation of points in a surround of equivalent area, $\mu_P^{(\mathrm{s})}$.

Two-cluster feature The 17-dimension two-cluster feature vector, f_2, is defined as

$$f_2 = \left(\frac{\mu^{(2)} - \mu^{(1)}}{\|\mu^{(2)} - \mu^{(1)}\|}, \left\|\mu^{(2)} - \mu^{(1)}\right\|, \Sigma^{(1)}, \Sigma^{(2)}, \frac{\mu_P^{(\mathrm{c})}}{\mu_P^{(\mathrm{s})}}\right), \tag{6.8}$$

with $\|\cdot\|$ denoting the l^2 norm and cluster one is defined to be that with the smallest μ_x. The first term in this feature vector is the normalised direction in (x, y, ϕ) between the two clusters, and the second term is a measure of the distance between the cluster centres. The choice to use the relative position between the clusters rather than their absolute position arose from the previous discussion about how vehicles observed from different orientation may present different absolute angles of polarisation, but the relative offset in angles between surfaces should be approxi-

mately constant.

6.2.4 Summary

In this section, a cluster-based descriptor has been introduced, with a number of different parameters, each of which are examined in more detail in Section 6.4. The formation of the one- and two-cluster descriptors, f_1 and f_2, were arrived at through a combination of hypotheses about which parameters would be discriminative, guided by the intuition gained through the previous histogram development and understanding of the physics of polarisation as presented in previous chapters. The discriminative ability of these features is assessed in the following pages.

6.3 Training

While developing the cluster based methods, it was advantageous to choose a subset of the training data with which to guide development. This subset, taken from DS2 and termed DS2-subset, consists of 417 positive samples, and 500 negative samples. The positive samples were further manually categorised into 206 one-cluster positive samples, and 211 two-cluster positives samples. A 5-folds training method was used, such that the dataset was randomly partitioned into five sets and, selecting each of these sets in turn as a test set, an RBF SVM[1] was trained on the remaining four and then tested on the fifth. This process was repeated for all five permutations, and an average of all results was taken.

6.4 Results

The details of the implementation described above are now explored in more depth. The choice of different components are compared, with reference to the default descriptor which consists of the following. The initial threshold on degree of polarisation, P_t, was calculated to retain points within the highest 60th percentile of P in each window; robust $k-$means clustering was used with $s = 1.5$.

[1]The histogram-based methods from Chapter 5 used a linear SVM by default. However, the SVM implementation used for cluster development (`libsvm` within MATLAB) is not particularly efficient for a linear kernel so here the RBF kernel is preferred. Where the histogram and cluster methods are compared in Section 6.5.2 this method is altered to ensure the training and testing procedures are consistent.

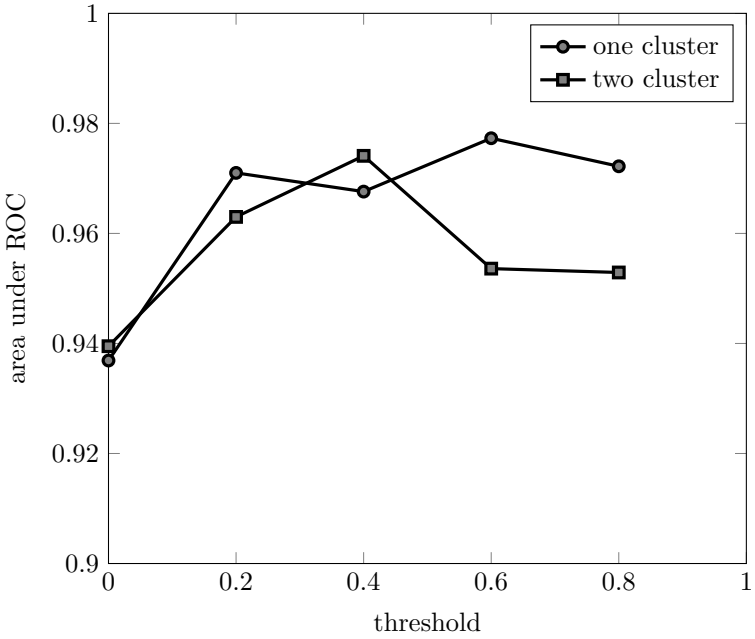


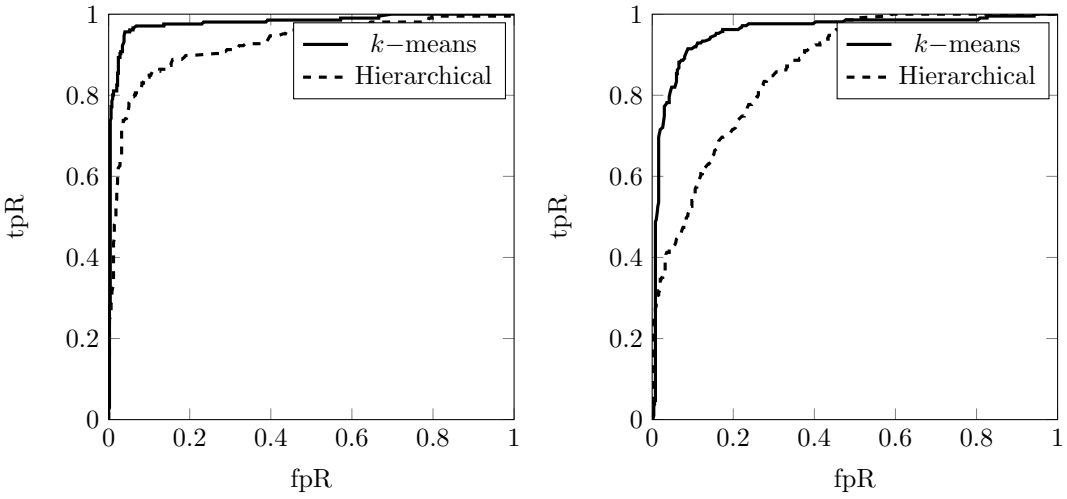
Figure 6.6: Threshold parameter

6.4.1 Threshold

A threshold was applied to the input window to select points with a higher degree of polarisation and disregard background points with a lower degree of polarisation. The desired positive windows contain vehicles which are centred and are at a consistent size within the window. As the degree of polarisation of a vehicle may not be constant across different images, the threshold was chosen such that a fixed number of points were selected, based on a percentile of the degree of polarisation within the window. The percentile above which points were selected was varied between $t = 0\%$ (select all points in the window), and $t = 80\%$ (select points with degree of polarisation within the top 20% of that window). Using the default parameters, an SVM was trained for each one- and two-cluster method and the area under the ROC curve is plotted for different thresholds in Figure 6.6. A significant improvement is seen when the pixels with the lowest degree of polarisation are removed, but there is little gain in removing further pixels. As a result, a threshold of $t = 40\%$ was chosen in order to eliminate many of the low degree of polarisation background pixels, while retaining a large enough number of pixels to ensure significant coverage of any high degree of polarisation objects in the window.

6.4.2 Choice of clustering method

Two different approaches for extracting clusters from the input window have been described: an outlier rejection method followed by hierarchical clustering; and a



(a) One-cluster feature

(b) Two-cluster feature

Figure 6.7: Comparison of clustering method

robust $k-$means method which includes outlier rejection as part of the clustering process. Using these methods, with the parameters as described above, classifiers were trained and tested for the one- and two-cluster features, the ROCs of which are shown in Figure 6.7. From these curves, the robust $k-$means method appears to have a better performance for both one- and two-cluster features.

To further explore the ability of each of these methods to extract suitable clusters on positive samples, Figure 6.8 shows clusters obtained from one-cluster samples for each method, and Figure 6.9 shows clusters obtained from two-cluster samples for each method. In both the one- and two-cluster cases, it is clear that the Hierarchical method retains a much larger proportion of the input samples points, while the robust $k-$means algorithm rejects a large proportion of input points as outliers. This leads to much smaller, tighter clusters with $k-$means, with each cluster largely centred on each vehicle surface (with the exception of the third $k-$means sample in Figure 6.9 which has detected some road as the second cluster), while the hierarchical method returns large clusters of points with a much greater spatial polarimetric spread of values.

Improvements to the hierarchical method could include improving the outlier rejection module to encourage more compact clusters. Improvements to the robust $k-$means algorithm might include adapting the outlier rejection method to self-tune the s parameter which defines how far points can deviate from the cluster mean before being defined as an outlier. This may allow the clusters formed on vehicle surfaces to grow to cover a larger portion of the surface, making the later description of the cluster more robust to any noisy pixels within the cluster data. For the rest of this section clustering will be performed using the robust $k-$means

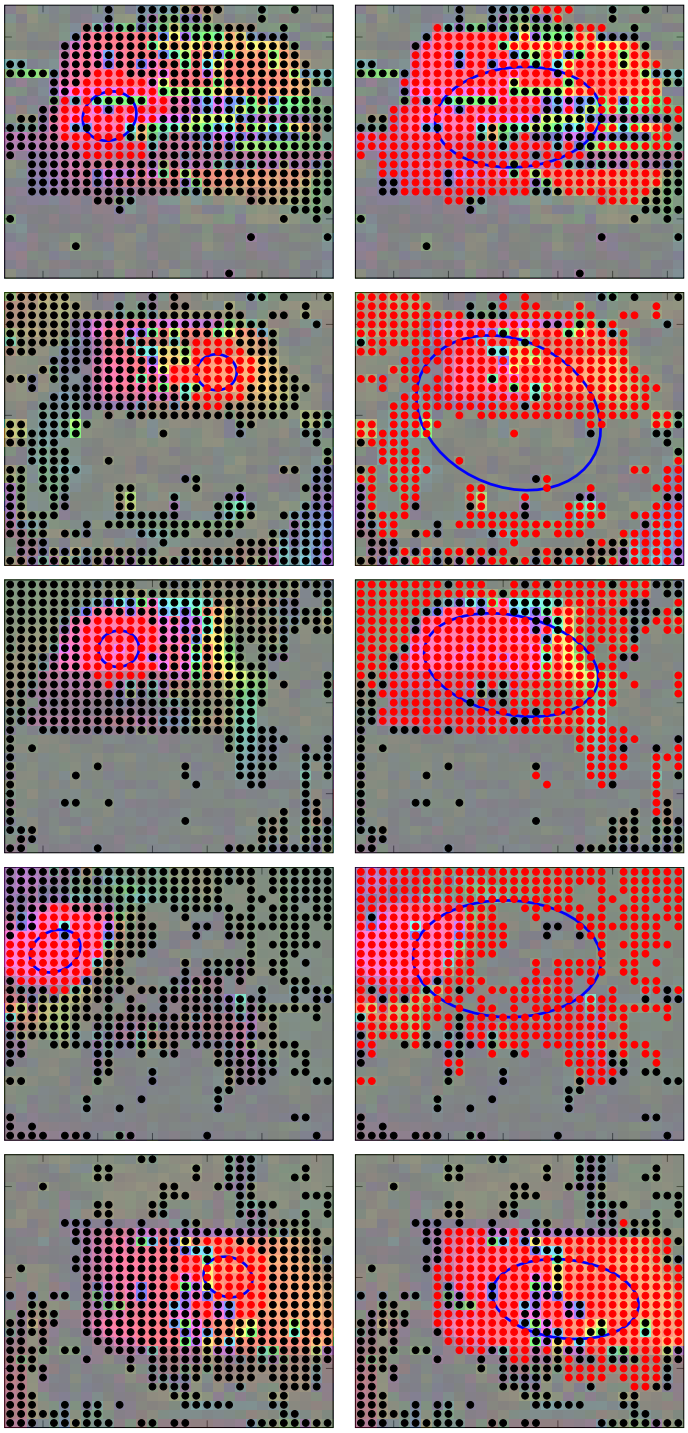


Figure 6.8: One-cluster clustering results on a set of five vehicle images using (left) robust $k-$means and (right) hierarchical clustering methods. Outlier points are marked in black, inlier points are marked in red, the blue oval identifies the 1σ standard deviation of the cluster.

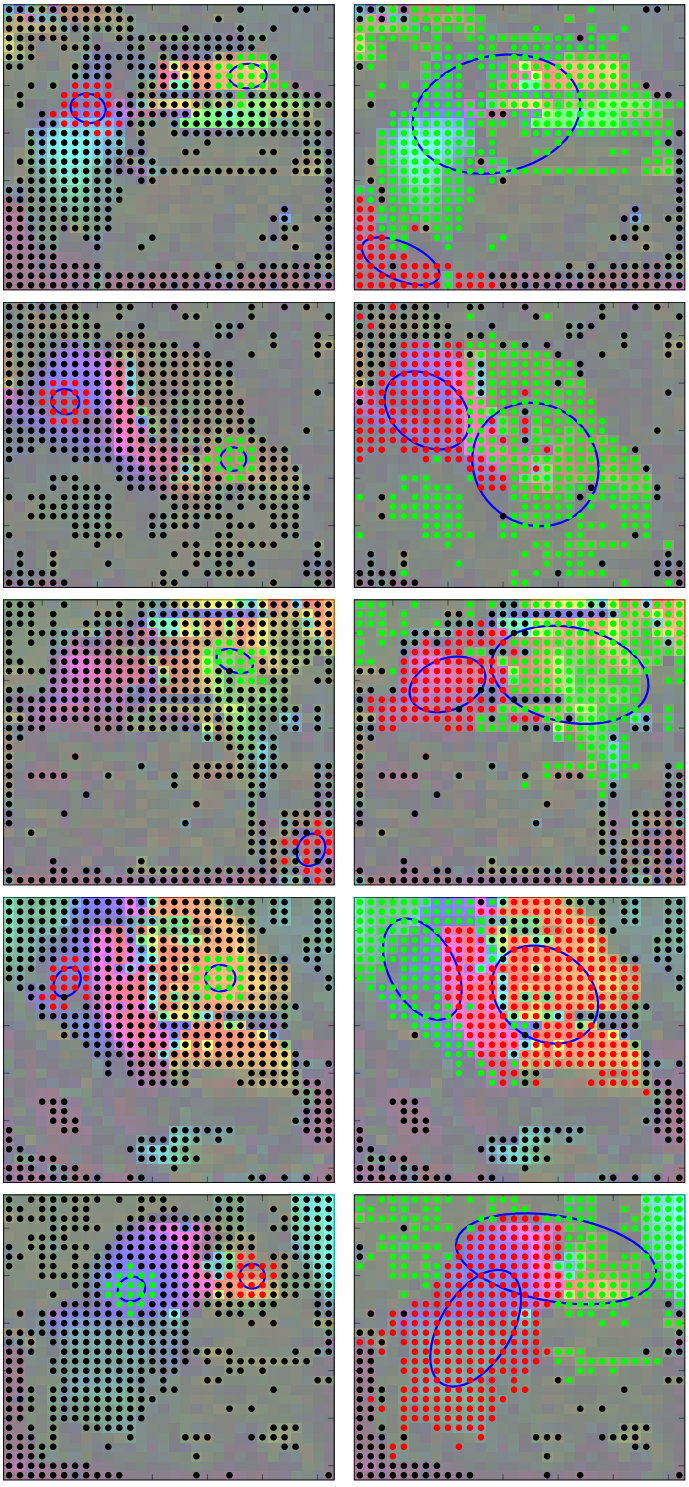


Figure 6.9: Two-cluster clustering results on a set of five vehicle images using (left) robust $k-$means and (right) hierarchical clustering methods. Outlier points are marked in black circles, inlier points for each cluster are marked in red and green, the blue ovals identifies the 1σ standard deviation of each cluster.

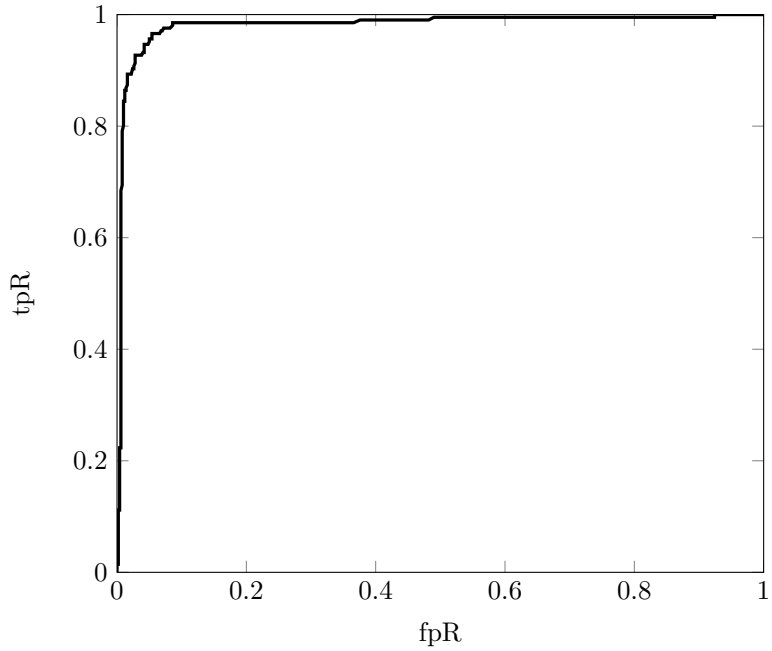


Figure 6.10: ROC curve of SVM trained on one-cluster feature.

method.

6.4.3 One-cluster feature evaluation

The 10-dimension one-cluster feature defined in equation (6.7) consists of the mean position of the cluster in (x, y, ϕ), the covariance of the cluster, and the centre-surround ratio of the degree of polarisation. A dataset containing 206 positive one-cluster vehicle samples and 500 negative samples was used to train an SVM classifier using 5-fold cross validation. The resulting ROC curve of this classifier is shown in Figure 6.10.

To further investigate the ability of this representation to partition positive and negative samples, Figure 6.11 shows a two-dimensional projection of positive and negative sample points for each pair of components. Each plot in this figure shows the projection of the indicated pair of components and allows some analysis on how well each component is able to separate positive and negative samples.

Mean cluster position Looking at the first three components, the mean (x, y, ϕ) position of the cluster, observe that there is little difference in the mean x position of positive and negative clusters; the mean y component of positive clusters is usually at low pixel indices (towards the top of the image) while there is no trend for the negative clusters; finally the mean ϕ component is centred around 80° to 100° for the positive samples while the negative samples have a greater variance. These values largely agree with the single cluster hypothesis case which consists of a single visible

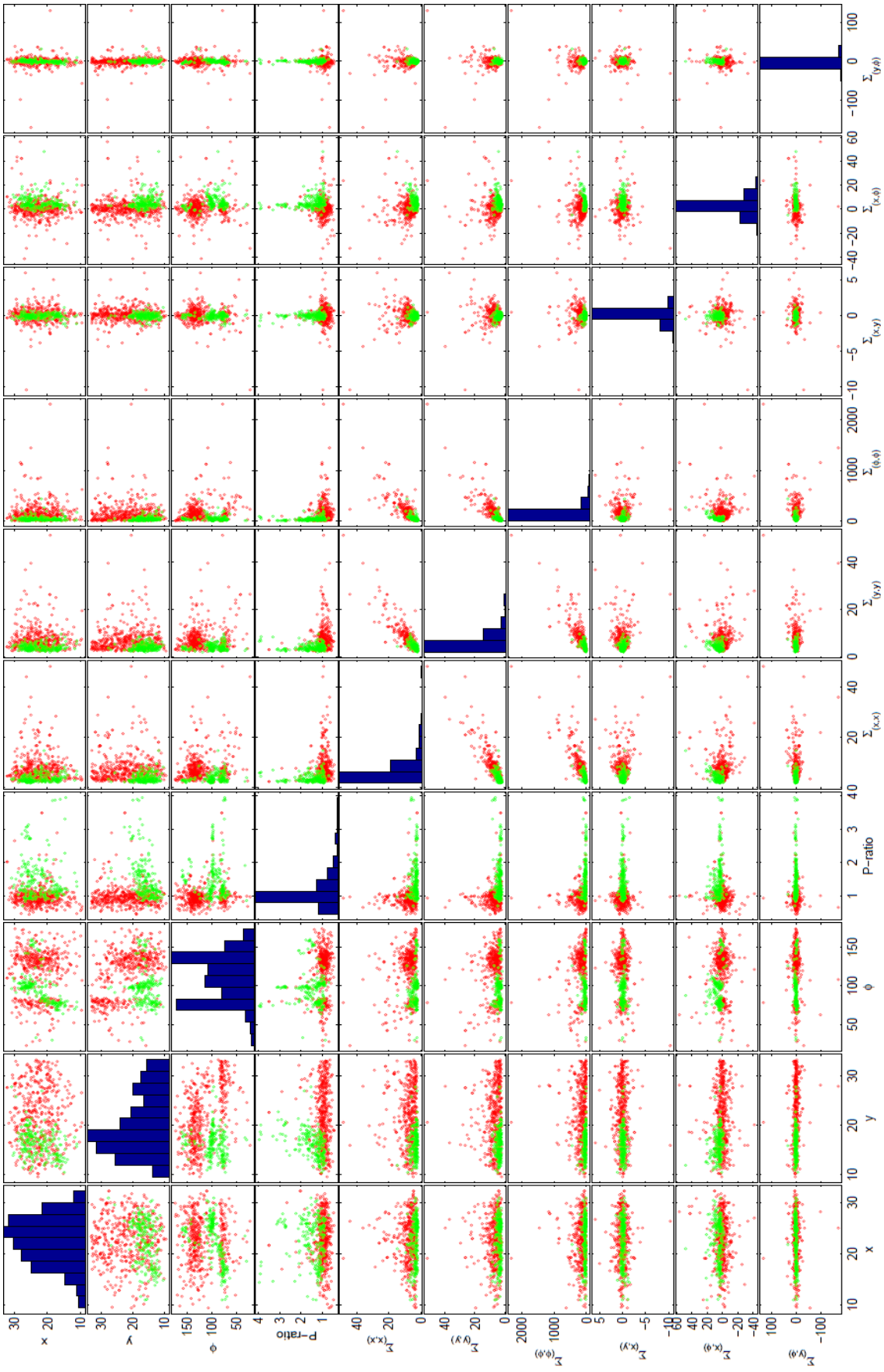


Figure 6.11: Projection of positive (green) and negative (red) sample points for each pair of components in the one-cluster feature vector.

vertically polarised cluster from the vehicle roof (such as Figures 6.1a and 6.1b), while the negative mean values have no strong feature and so have much greater variation.

Centre-surround degree of polarisation The coordinate termed $\mathrm{P}-\mathrm{ratio}$ is the ratio of the average degree of polarisation within the central window, to the average degree of polarisation within a surrounding window of equal area. It is expected that high degree of polarisation vehicles within a low degree of polarisation background would be described by a $\mathrm{P}-\mathrm{ratio} > 1$ while background regions could generally be described by $\mathrm{P}-\mathrm{ratio} \approx 1$. From the $\mathrm{P}-\mathrm{ratio}$ plots of Figure 6.11 this is largely what is observed.

Cluster covariance The rationale behind using the cluster covariance in the feature vector was that positive samples would result in tightly centred clusters on the vehicle surface which they were describing, while background samples would result in much looser clusters. From the first three diagonal terms in Figure 6.11 (comparing $\Sigma_{(x,x)}$, $\Sigma_{(y,y)}$, and $\Sigma_{(\phi,\phi)}$) it is seen that the positive samples are generally centred around the points $(0, 0)$, indicating tight clusters, while the negative samples have much greater spread.

Principal Component Analysis Additionally, Principal Component Analysis was performed on all features (positive and negative), and the second and fourth principal component representation of each sample feature is plotted in Figure 6.12. These components are shown as they appear to have the most significant separation of the first few components. However, with additional dimensions the features should be easier to separate.

6.4.4 Two-cluster feature evaluation

The 17-dimension two-cluster feature defined in (6.8) consists of the vector between the mean (x, y, ϕ) positions of each cluster (such that $\vec{x}$ is positive), the distance between the clusters, the covariance of each cluster, and the centre-surround ratio of the degree of polarisation. A dataset containing 211 positive two-cluster vehicle samples and 500 negative samples was used to train an SVM classifier using 5-fold cross validation. The resulting ROC curve of this classifier is shown in Figure 6.13.

As above, Figure 6.14 shows a two-dimensional projection of positive and negative sample points for each pair of components. Each plot in this figure shows the projection of the indicated pair of coordinates.

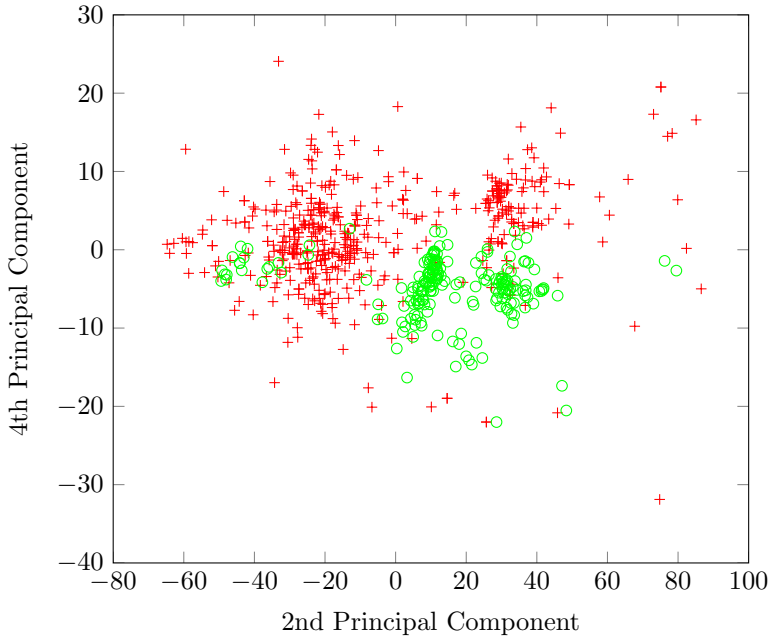


Figure 6.12: Principal component projection of positive (green) and negative (red) sample points for the one-cluster feature.

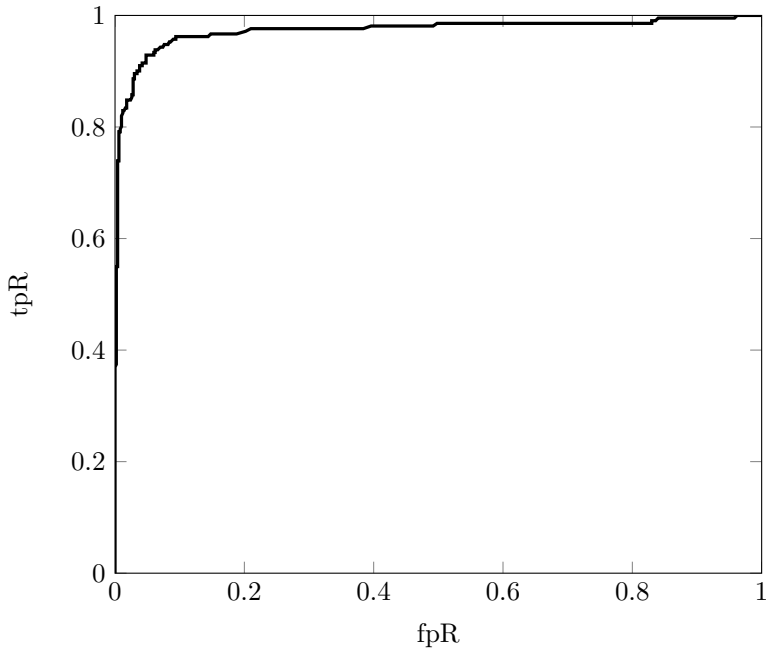


Figure 6.13: ROC curve of SVM trained on two-cluster feature.

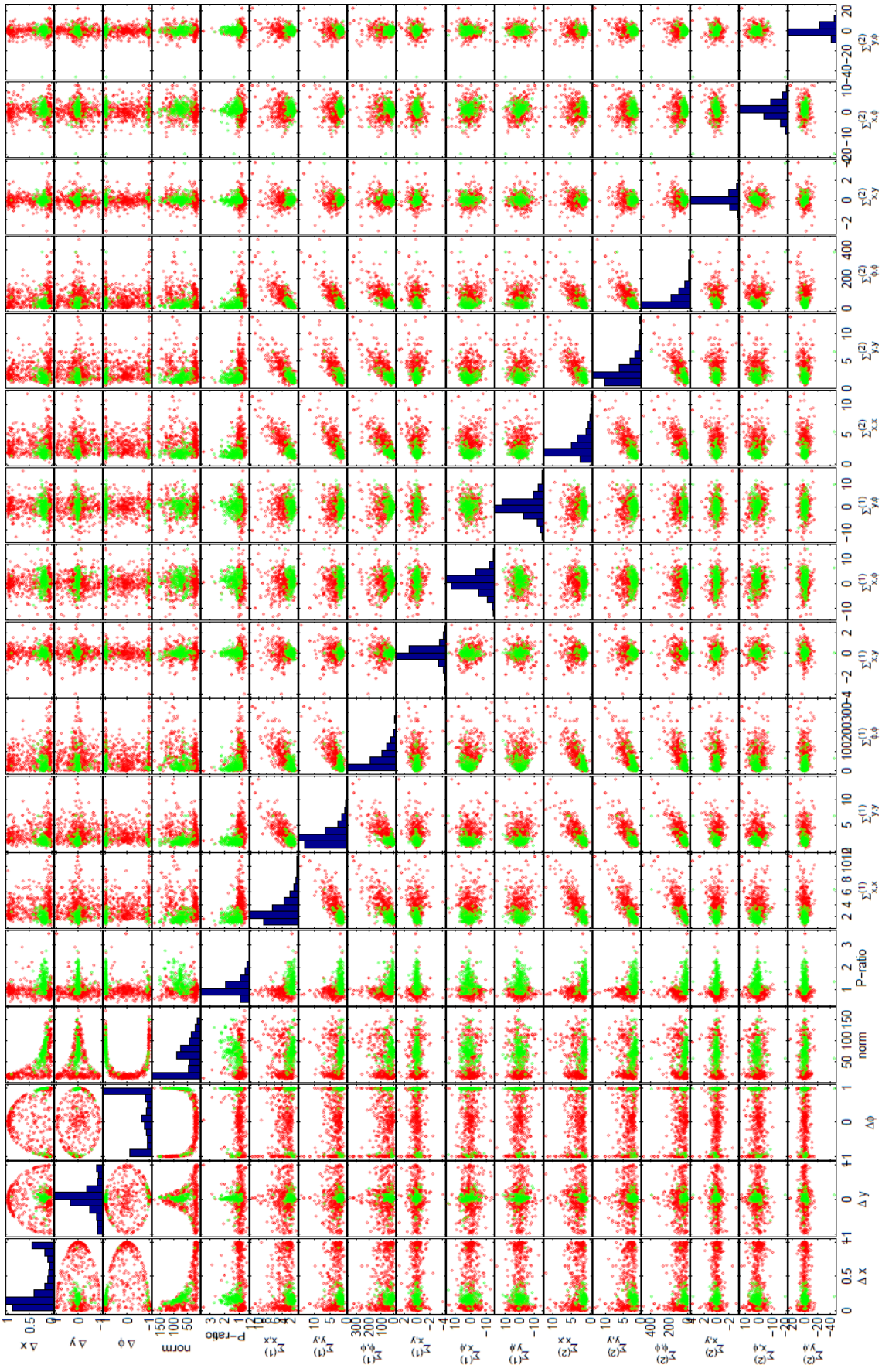


Figure 6.14: Projection of positive (green) and negative (red) sample points for each pair of components in the two-cluster feature vector.

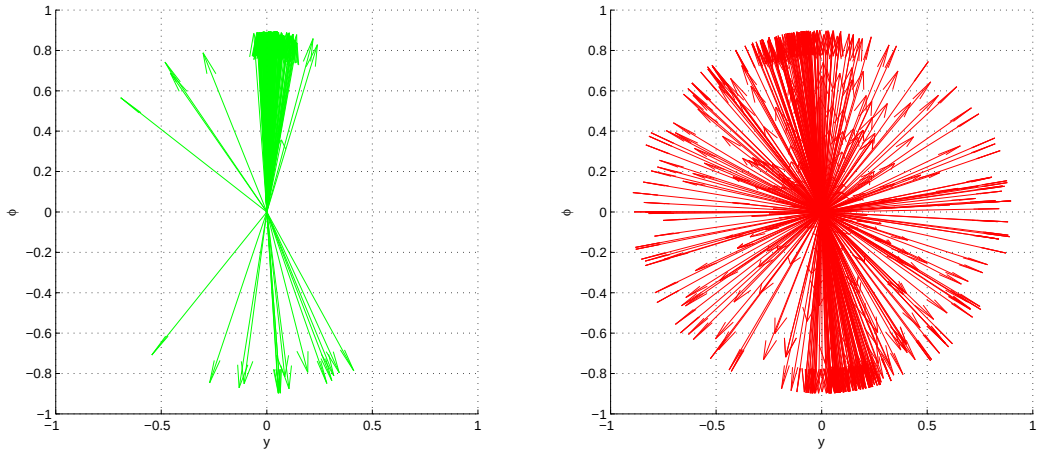


Figure 6.15: Projection of direction between clusters for each positive (left) and negative (right) sample.

Vector direction The first three plots in Figure 6.14 show the direction in normalised (x, y, ϕ) coordinates. The positive samples show that the two clusters have a small separation in x, and y, while a significant separation in ϕ. This is in contrast to the negative samples, which show a much broader distribution in all three coordinates. Figure 6.15 shows the $y - \phi$ projection of the inter-cluster direction for positive and negative samples. This further highlights the small variation of directions between two clusters detected from the positive samples, while the direction between two clusters detected from the negative samples is much more variable.

Vector length While the direction between cluster centres appears to be a strong distinguishing feature, the distance between cluster centres is less strong. The fourth coordinate plotted in Figure 6.14 indicates that a large number of negative samples have a very small distance between cluster centres, although there are also a large number of positive and negative samples with a larger separation and so this coordinate is not considered very strong at distinguishing between positive and negative samples and should most likely be removed for future work.

Centre-surround degree of polarisation As for the one-cluster case it is observed that the majority of negative samples have $\mathrm{P-ratio} \approx 1$, while the positive samples have $\mathrm{P-ratio} > 1$. The rationale is no different in this two-cluster case, and so the similarity is to be expected.

Cluster covariance The two-cluster feature vector also contains the covariance of both clusters, with the hypotheses that clusters on vehicle surfaces will be tighter than those on background regions. Compared to the one-cluster feature vector, this

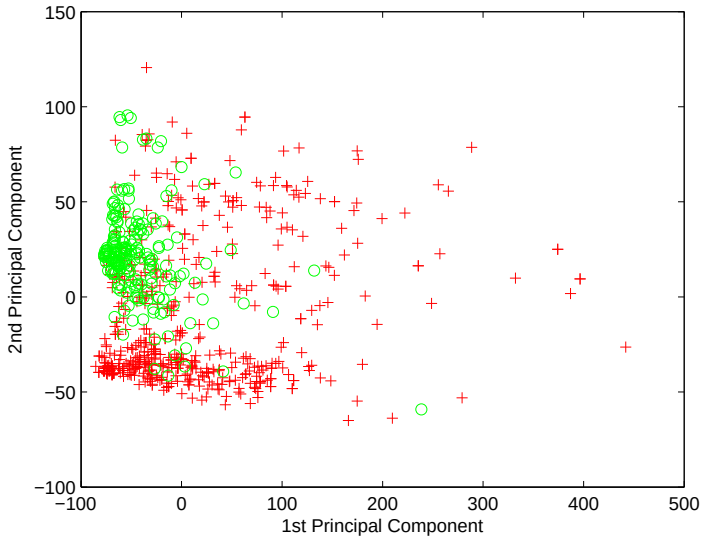


Figure 6.16: Principal component projection of positive (green) and negative (red) sample points for the two-cluster feature.

is less obvious, but is still the case as can be observed particularly in the three diagonal terms of each cluster.

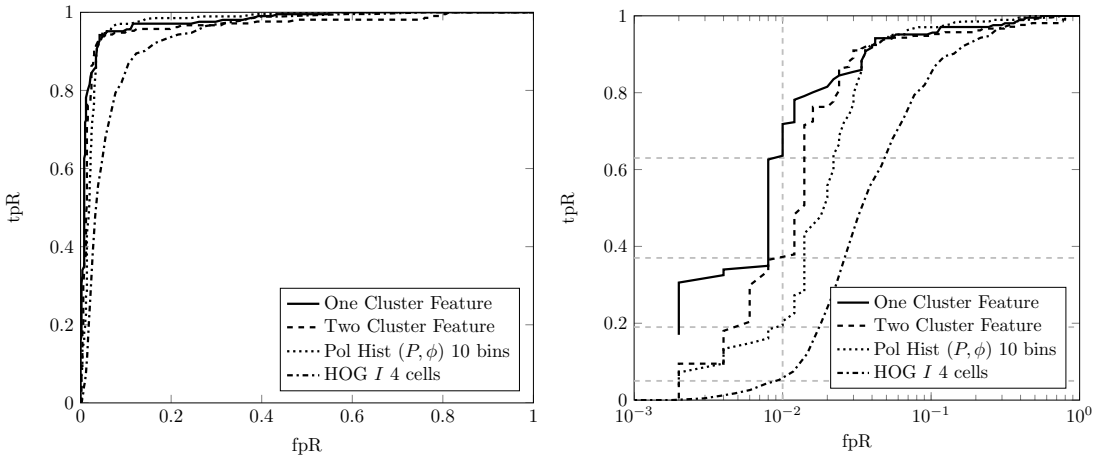
Principal Component Analysis Additionally, the first and second principal component representation of each sample feature are plotted in Figure 6.16. The PCA projection does show reasonable separation between the positive and negative samples indicating that the majority of sample points are well separated in this representation, but there are some cases where positive and negative feature vectors are similar.

6.5 Discussion

6.5.1 Three cluster feature

As mentioned earlier, cases involving three or more clusters have not been considered in the current work. The cluster based detector relies on being able to identify vehicle surfaces for which clusters of pixels at high degree of polarisation and similar spatial location and angle of polarisation exist.

The data collected from which to train and test the developed polarimetric vehicle detector (examples of vehicles therein are shown in Figure 6.1) was collected from a fixed height (a camera mounted on the roof of the trials van), and as such the variation in vehicle pose collected is mainly azimuthal while there is little variation in elevation. A three cluster case may exist when a vehicle is observed from



(a) Linear scale

(b) Logarithmic scale

Figure 6.17: Comparison of the best performing histogram-based descriptor and the cluster-based descriptors shown with the false positive rate on (a) linear and (b) log scales. Note that bot the Polarimetric Histogram and HOG descriptors have been trained and tested using the reduced dataset used with the cluster method.

an angle from which an equal number of pixels would span three vehicle surfaces, such as observing Figure 6.1g, 6.1j, or 6.1k from a higher elevation. However, as a limitation of the dataset, these cases are not observed and so development of a three cluster case is left as future work.

One logical extension to this work could consider vehicles (or any strongly polarising object class) as a three dimensional structure, and aim to match the trained three dimensional object class to the polarimetric imagery in a similar manner to Leotta and Mundy [47], whose three dimensional vehicle model (discussed in Section 4.2.2) was adapted to fit edges detected in the imagery. The cluster based approach presented in this section is a simpler model and so less knowledge about the intricate three dimensional nature of vehicles is required.

6.5.2 Comparison with histogram approach

As mentioned previously, the histogram-based descriptors were trained and tested on all the ground truth imagery available within DS2, while the cluster-based descriptors only used a subset (473 positive and 500 negative samples). In order to obtain a fair comparison, the best performing descriptor which uses only intensity imagery and the best performing histogram-based descriptor – the Histogram of Gradients with 4×4 cells and the Polarimetric Histogram with a single spatial cell, and ten histogram bins in each of the P and ϕ dimensions respectively – were trained and tested using the same dataset and classifier as used for the cluster approach[1]. The

[1]While the cluster based descriptor breaks the positive training data down into one- and two-cluster subsets, the Polarimetric Histogram and HOG methods use the entire positive dataset. If

Table 6.1: True positive rate of cluster and histogram-based descriptors at a false positive rate of 10^{-2}.

Descriptor	tpR
HOG on I imagery	0.05
Polarimetric Histogram	0.19
One-Cluster Feature	0.37
Two-Cluster Feature	0.63

resulting ROC curves are shown in Figure 6.17 alongside results from the one- and two-cluster ROC curves, repeated from Figures 6.10 and 6.13. The false positive rate is shown on both linear (6.17a) and logarithmic (6.17b) scales to highlight the differences between the plots. At a representative false positive rate of 10^{-2}, the true positive rates for each descriptor are shown in Table 6.1. The advantage of the polarimetric approach is immediately apparent, at a false positive rate of 10^{-2} the Polarimetric Histogram outperforms the intensity-only HOG descriptor by a factor of four, and the cluster-based approach shows an improvement of approximately an order of magnitude over the intensity-only descriptor.

6.5.3 Video Sequences

Finally, some examples of the application of the cluster-based approach to a a number of video sequences are presented, highlighting some of the failure modes which occur. In the first set of examples, all positive detections are shown, in the latter examples non-maximum suppression is applied to merge multiple detections covering a single object.

Figure 6.18 shows some output from the one-cluster descriptor, the two-cluster descriptor and the (P, ϕ) Polarimetric Histogram (PH, Section 5.4) with a single spatial cell and ten bins in each dimension. In these images, a sliding window is scanned across the whole image, with a 5 pixel step size, and 10 scales between 42×42 pixels (a central 30×30 pixel square with an equal area surround) and 210×210 pixels (the images are 240×240), leading to 4,000 window locations per image. A feature is calculated for each window location, and the label of that feature is predicted using the associated SVM classifier. All positive detections are shown in these images, although for use as a detector additional processing including non-maximal suppression would normally be applied to remove any isolated detections, and merge multiple detections covering a single object.

In general, the cluster based descriptors tend to be more aligned with the vehicles, while the PH descriptor does mark some windows as detections if they only partially

the comparison is extended to view-dependent HOG / Polarimetric Histogram descriptors this may require more detailed analysis of precisely which training sets should be used in which scenarios.

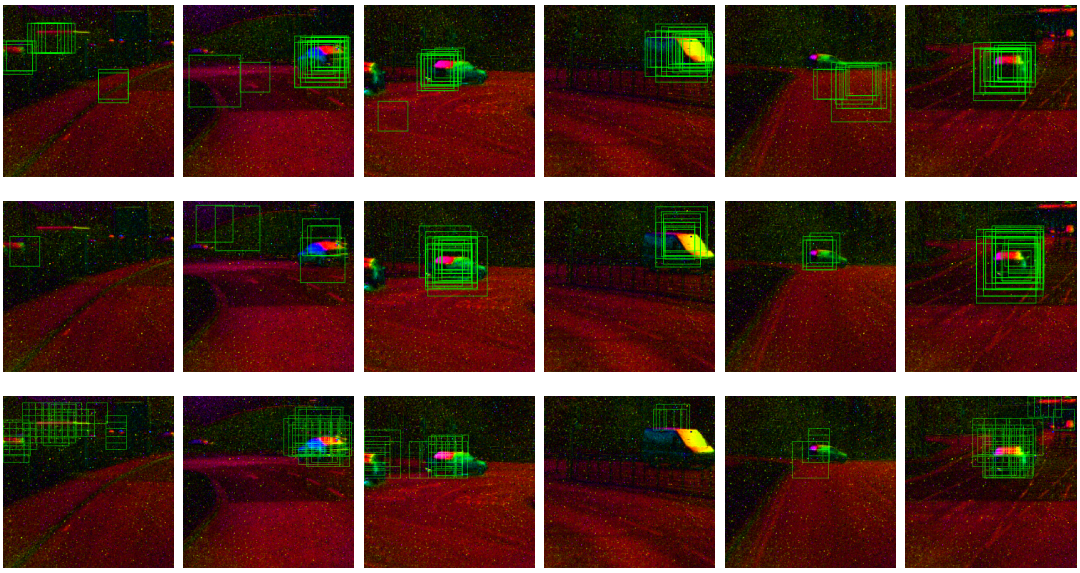


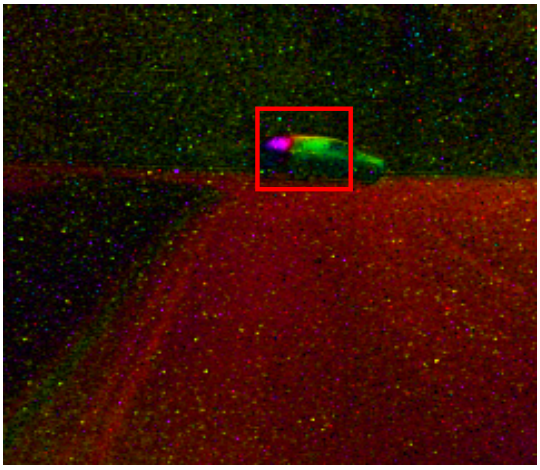
Figure 6.18: Example output frames from the One-Cluster (top), Two-Cluster (middle), and Polarimetric Histogram (bottom) descriptors.

overlap the vehicle. This is particularly evident in the second and sixth images.

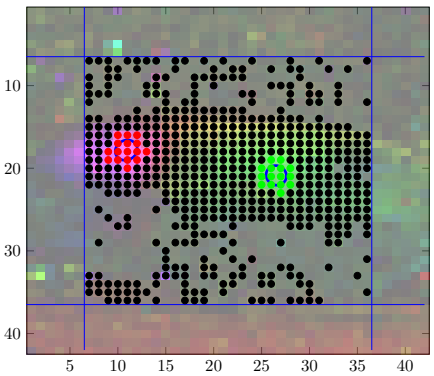
The van image (the fourth image) proves hard to detect with the PH descriptor, while the cluster approaches (both one- and two-cluster) mark some detections here. The particular surface orientations visible for the van are different to those normally observed for cars. The PH descriptor mainly looks for a strong 90° component (red in these images), which is only present to a very small amount on the roof of the van. However, the cluster approaches combine their models of the expected angle of polarisation (or relation between the average angle of two clusters) with parameters describing the degree of polarisation compared with the background, and the compactness of the clusters in terms of the covariance. This extra information seems to help the cluster-based approaches detect this type of vehicle.

While a number of the example images show the same vehicle being detected by both the one- and two-cluster detector (due to strong one- and two-cluster features), an exception is the fifth example in Figure 6.18 which contains a vehicle which is reasonably distant. As this vehicle has a small scale the dominant cluster detected by the single cluster detector is the green cluster on the vehicle side. As this feature is not consistent with the one-cluster case (where typically a red cluster with $\phi \approx 90°$ will be detected) it is rejected as a negative sample. However, the two-cluster feature is able to detect both the green and red regions (shown in Figure 6.19) and as this fits the two-cluster model well, it is marked as a detection.

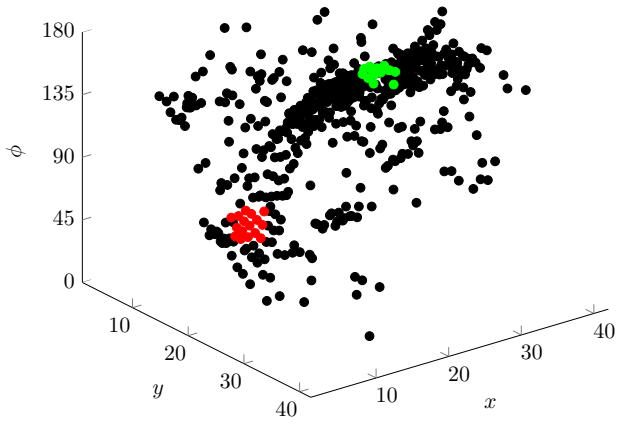
Some additional more cluttered examples that include non-maximum suppression are shown in Figure 6.20. In these cases, the red/green boxes denote a positive one-/two-cluster detection respectively. Figure 6.20a shows that the approach is effective in detecting a single car, which has a pronounced two-cluster signature,



(a) Detected region

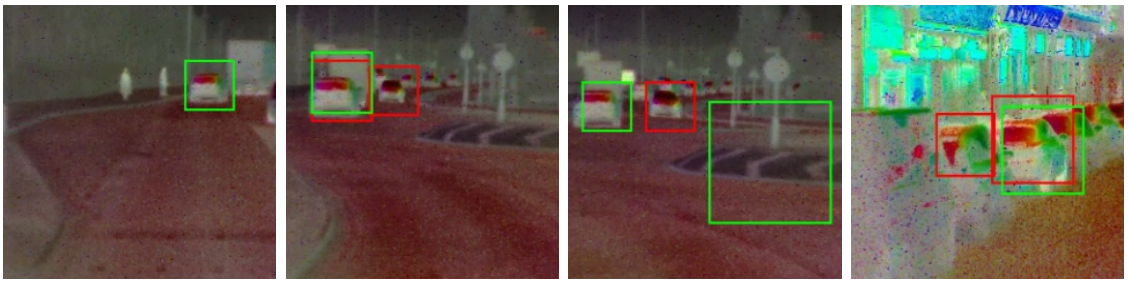


(b) Clusters detected within central region



(c) (x, y, ϕ) projection of clustered points.

Figure 6.19: Example of the two-clusters extracted from the fifth example image in Figure 6.18.



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

Figure 6.20: Examples of video frames with non maximum suppression, red: one-cluster, green: two-cluster detections.

but advancing a few frames to Figure 6.20b, it is seen that both one- and two-cluster detections are triggered on the same car, whereas the next car in front is only triggered by a single cluster. As yet no consideration has been given on how to combine results from the one- and two-cluster detectors. The large false positive in Figure 6.20c arises because a two-cluster feature is detected from the road and traffic island. However, what is immediately apparent is the scale of detection - currently, camera and road context are not used to eliminate this kind of false positive. The other two vehicles in this frame are well detected. Finally, Figure 6.20d shows a very cluttered and noisy scene, with several parked cars by the side of the road, and shop fronts which also give strong polarising returns. The algorithm does succeed in identifying the presence of cars, and it does not detect the shop fronts as false positives. However, it cannot determine the number of vehicles, nor separate them satisfactorily. To separate closely coupled targets is a common problem in all forms of video object detection, and is not yet satisfactorily modelled by the spatial arrangement of clusters defined here.

6.6 Conclusion

In this chapter further developments of the histogram-based approach presented previously re-examined the polarimetric structure of vehicles, and led to the development of an alternative cluster-based approach to vehicle detection. This method aimed to detect clusters of spatially close pixels with similar angle of polarisation and high degree of polarisation. Two clustering methods were explored, however both required the user to select a specific number of clusters to extract which led to two different descriptors - the one- and two-cluster descriptors, which were shown to be particularly successful in detecting vehicles based on the polarimetric content of the scene.

A comparison between the HOG I descriptor, the Polarimetric Histogram, and

the cluster based approach showed a significant increase in true positive rate for the descriptors which identified vehicles using their polarimetric signature. At a false positive rate of 10^{-2}, the two-cluster approach exhibited a true positive rate of 0.63, more than one order of magnitude above the HOG I descriptor which had a true positive rate of 0.05. This highlights the strength of the polarimetric signature of objects such as vehicles, and confirms that an understanding of the polarimetric properties of such vehicles can aid the development of effective detectors for these objects.

Two extensions to the current work could include analysis on whether additional clusters (for example a three cluster case) further improve results, and whether a method can be developed which detects an appropriate number of clusters and matches them to an overall vehicle model. It is possible that an alternative detection framework, other than an SVM classifier, may better suit this model based approach.

Chapter 7
Multi-modal Vehicle Detection: A Case Study

This chapter presents a case study into a vehicle detection framework which combines input from different sensors with processing using different hypothesis generation algorithms (HGs). As a case study, simple detection algorithms are used, but the work provides the scope for expansion to include the detectors discussed in the preceding chapters.

7.1 Introduction

Multi-sensor fusion can be used to integrate complementary information from a variety of sensors to enhance the performance of tasks beyond the capabilities of a single sensor. Letham *et al.* [2] demonstrate a system which makes use of a number of sensor modalities to improve a region segmentation task. The system combines a sky region detector which uses RGB imagery, a road region detector which uses polarimetric thermal imagery, and a vegetation region detector which uses a combination of NIR and colour imagery. The output from each region detector is combined in a Bayesian formulation using a prior based on the known performance of each detector, and the historic segmentation of the scene. Using a set of detectors in combination is shown to provide an improved result when compared to each detector working independently.

When data is collected from multiple sensors the stage at which it is combined must be carefully considered as there are implications on the volume of data requiring processing, and the sensitivity of the output to errors in misalignment. The stage at which fusion occurs may be approximately categorised into three levels of increasing abstraction [108].

Pixel-level fusion combines the raw (or pre-processed) imagery to form a multi-dimensional datacube.

Feature-level fusion combines features which are extracted from the inputs before being subsequently classified.

Decision-level fusion combines decisions (binary or otherwise) which are derived from single input streams.

Fusion at lower levels of abstraction requires more accurate registration methods to align the data and is more computationally intensive as a greater volume of data is carried through the system. Examples of pixel level fusion include the use of NIR sensors in combination with RGB sensors for object segmentation [109], image categorisation [110], photographic skin enhancement [111], and de-hazing colour images [112].

Fusion at higher levels of abstraction can be more readily adaptable to different sensor types for which accurate sub-pixel registration is not possible. D'Arca [113] demonstrates improved person tracking via the fusion of audio and video sensors. In this work, tracks are hypothesised using both a mean-shift tracker processing the visible sensor, and triangulation of the time differences of arrival across pairs of microphones. These different track hypotheses are then combined in the fusion stage.

Finally, it is also worth noting that when multiple sensors or algorithms are used in conjunction, each source of information should not necessarily be trusted equally. For example, visible imagery may provide the best performance during daylight, while when used during the hours of darkness, an infrared sensor may be more preferable. As such, the notion of trust in a source of information is an important consideration, as is the ability to assess the trustworthiness of a particular information source while a system is in operation.

In this work, the input from two sensors are considered.

Jai AD-080 GE 2-CCD multi-spectral camera provides access to both the visible and near infrared (NIR) bands (1024×768 pixels, NIR at 750 nm to 1000 nm, 30 frames/s) through a single optical system.

Thales Catherine MP LWIR Polarimeter as introduced previously in Chapter 3 measures intensity and polarimetric imagery in the longwave infrared.

The following sections explore a vehicle detection system using different combinations of input imagery, and different types of detector. The architecture of the vehicle detection system is described in more detail in Section 7.2. The experimental methodology is outlined in Section 7.3 and the results are discussed in Section 7.4. Finally, Section 7.5 summarises with some closing remarks and paths for further investigation.

7.2 Vehicle Detection System

The proposed vehicle detection system consists of a set of hypothesis generators (HGs) which process the input imagery to form a set of output detection maps. A detection map is a binary image produced by each HG to highlight pixels which are determined to be vehicle pixels. These detection maps are subsequently registered and combined with a hypothesis combination method to provide a final set of vehicle pixel detections for that frame. This process is summarised in Figure 7.1.

In the hypothesis generation stage, each of n HGs operate on a single image I_j from the set of m available images, $\mathbf{I} = [I_0, I_1, \ldots, I_m]$. In the current implementation the set of input imagery include visible, near infrared (NIR), LWIR Intensity (TI), and LWIR Q (TQ, the Stokes Q image), but this could be extended with other polarimetric imagery (such as P and ϕ), or imagery from other wavebands and even non-imaging sensors such as radar or lidar. Each input image j is processed by hypothesis generator i, $f_i(.)$, to produce an output detection map, $y_{i,j}$,

$$y_{i,j} = f_i\left(I_j\right), \quad i \in \{0 \ldots n\}. \tag{7.1}$$

The HG algorithms used to process the available imagery are described in Section 7.2.1.

The output detection maps are binary images in the local co-ordinate system of the sensor from which they were obtained. Before combining them, a registration process is applied to align these detections into a common coordinate system

$$y'_{i,j}\left(\mathbf{x}'\right) = y_{i,j}\left(t^j_{x_0}\left(\mathbf{x}\right), t^j_{x_1}\left(\mathbf{x}\right)\right). \tag{7.2}$$

In this equation, the transformation $t^j_{x_0,x_1}$ transforms images from sensor j into a common coordinate system by mapping pixels $\mathbf{x}' \leftarrow \mathbf{x}$, $\mathbf{x} = (x_0, x_1)^{\mathrm{T}}$. The registration process is described in Section 7.2.2.

Finally, the set of registered detection maps, $\mathbf{y}' = \left[y'_{0,0}, \ldots, y'_{n,m}\right]^{\mathrm{T}}$, are combined using a hypothesis combination (HC) method.

$$z = g\left(\mathbf{y}'\right) \tag{7.3}$$

to produce the final output detection image, z. The hypothesis combination algorithm is discussed in Section 7.2.3.

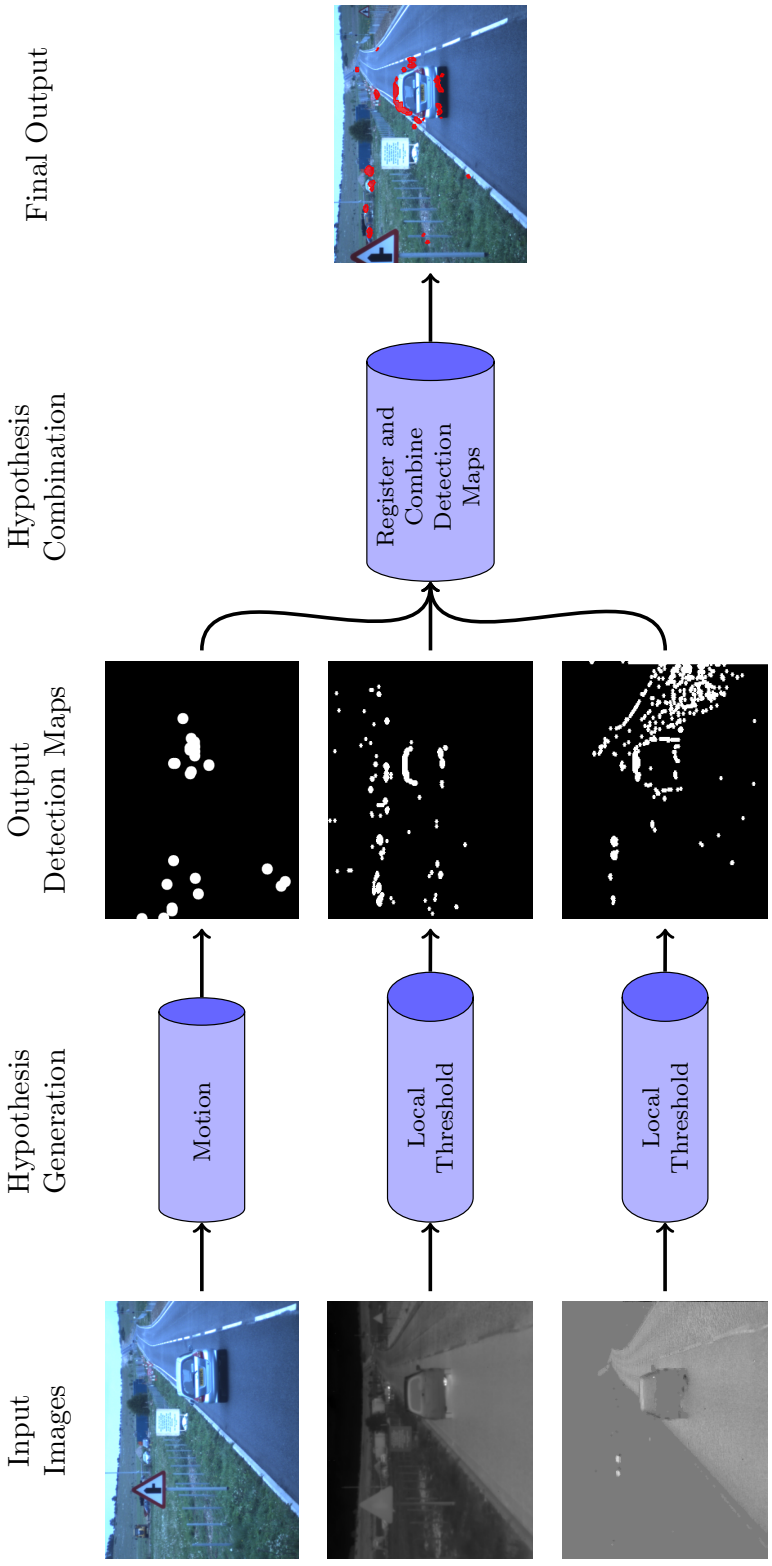


Figure 7.1: Flowchart demonstrating operation of algorithm. The input images are (top to bottom) Visible, LWIR Intensity, and LWIR Stokes Q. In this example, at least two detection maps must agree for a final detection to be marked.

7.2.1 Hypothesis Generation

The first stage in this detection system involves processing the input imagery with a number of different hypothesis generators (HGs) which each transform a particular input image into a binary detection map. Each detection map represents a hypothesis about the location of vehicles within the imagery, and by reaching these hypotheses with different algorithms processing different input data, the hypothesis is that HGs will agree on the position of vehicles, but will additionally output noise unique to each HG.

This section details three different HG methods, two consisting of simple thresholds which are used to process the thermal intensity and the thermal Q inputs, and one more complex HG which aims to detect moving objects from the visible imagery.

Global threshold

Due to their nature, vehicles can often be identified as local heat sources, or as regions of high linear polarisation. To exploit this a simple HG is proposed which uses a global threshold to form a detection map from the input imagery and is designed to be used with the thermal I, P or Q imagery.

The global threshold method produces a binary detection map controlled by a single parameter, p_{Global}. It is defined as

$$f_{\text{TGlobal}}(\mathbf{x}) = \begin{cases} 1 & I(\mathbf{x}) \geq p_{\text{Global}} \\ 0 & \text{otherwise} \end{cases} \tag{7.4}$$

where $I(\mathbf{x})$ is the input image.

Local threshold

Vehicles are distinguished from the background by being brighter than their local neighbourhood, so it may be more appropriate to consider a local threshold than a global threshold. The local threshold method extends the global threshold HG by refining the definition of the threshold level such that it is based on the local content of the image. The local threshold method also produces a binary detection map, and is defined as

$$f_{\text{TLocal}}(\mathbf{x}) = \begin{cases} 1 & I(\mathbf{x}) \geq t(\mathbf{x}) \\ 0 & \text{otherwise} \end{cases} \tag{7.5}$$

where the local threshold for each pixel, $t(\mathbf{x})$, is calculated as a parametrised offset from the mean intensity within a 3×3 pixel window

$$t(x_0, x_1) = \frac{1}{9}\left(\sum_{i=x_0-1}^{x_0+1}\sum_{j=x_1-1}^{x_1+1} I(i,j)\right) - p_{\text{local}}. \tag{7.6}$$

Here the parameter p_{local} defines the offset. Note that defining the local threshold over a 3×3 pixel window means that the central pixel is only compared to the intensity of a 3×3 patch centred on that location. This means that when an object extends larger than this size, the central pixel will be compared to other pixels also contained in the vehicle and as such vehicles larger than this size will either not be detected, or will only partially be detected.

Motion

The motion HG aims to detect vehicles which are moving through the scene. As the camera is also assumed to move through the scene, methods such as background subtraction (which assume a constant background) are unlikely to work. To overcome this, the motion HG tracks a number of feature points through the scene, determines which tracks can be attributed to the global motion of the camera, and assesses whether any further secondary motion (such as that from an object moving separately to the background) can be detected from the remaining tracks. This HG is essentially a simple implementation of the multi-body structure from motion problem, which is discussed extensively in the literature [114–119]. The algorithm is described below, and summarised in Algorithm 7.1.

In the first image of a sequence, the motion HG is initialised by detecting FAST (Features from Accelerated Segment Test [120]) feature points, which are then tracked using the Pyramidal implementation of the Lucas Kanade Optical Flow algorithm [121, 122], where correspondences are sought within a 10×10 pixel window over four pyramidal scale levels. As the sequence progresses, not every point will be successfully tracked and so the HG is reinitialised by detecting new FAST features points every n frames, where currently $n = 10$.

The points which are successfully tracked are used in a RANSAC (RANdom SAmple Consensus) [123] framework to obtain the fundamental matrix [124] which describes the camera motion required (in terms of rotation and translation) to explain a certain set of tracks on a rigid object. The RANSAC method makes the assumption that the data is noisy, and so aims to find the fundamental matrix which agrees with the majority of the tracked points. It is further assumed that the majority of tracked points will arise from background feature points and thus this first estimation of the fundamental matrix will relate to the motion of the camera

Algorithm 7.1 Motion Detection

INPUT: I_i current frame, q radius to mark as output
OUTPUT: f_{Motion} detection map for current frame

if first frame OR re-initialise **then**
  $\mathbf{X}_i \leftarrow$ detect FAST feature points from I_i
else
  $\mathbf{X}_i \leftarrow$ track feature points $\mathbf{X}_{i-1}$ from previous frame
  $\mathbf{s}^{(1)} \leftarrow$ find fundamental matrix between $\mathbf{X}_i$ and $\mathbf{X}_{i-1}$ ▷ $\mathbf{s}$ is a binary vector indicating whether the corresponding track is an inlier or outlier to the fundamental matrix obtained
  $\mathbf{X}_i^{(2)} \leftarrow \left\{\mathbf{X}_i | \mathbf{s}^{(1)} = 0\right\}$ ▷ select tracks marked as outliers
  $\mathbf{X}_{i-1}^{(2)} \leftarrow \left\{\mathbf{X}_{i-1} | \mathbf{s}^{(1)} = 0\right\}$ ▷ select tracks marked as outliers
  $\mathbf{s}^{(2)} \leftarrow$ find fundamental matrix between $\mathbf{X}_i^{(2)}$ and $\mathbf{X}_{i-1}^{(2)}$
  $\mathbf{X}^{(\text{o})} \leftarrow \left\{\mathbf{X}_i^{(2)} | \mathbf{s}^{(2)} = 1\right\}$ ▷ Output points are marked as inliers
  $f_{\text{Motion}}(\mathbf{x}) = \begin{cases} 1 & \min\limits_i ||\mathbf{x} - \mathbf{x}_{\text{o}i}||_2 \leq q \\ 0 & \text{otherwise} \end{cases}$ ▷ q is radius to output
end if
$\mathbf{X}_{i-1} \leftarrow \mathbf{X}_i$ ▷ Mark current points as previous points

through the scene.

Having determined which tracks belong to the global motion of the camera through the background, any points which are termed as outliers to this motion are reassessed with a second RANSAC fundamental matrix estimation. This aims to identify any secondary consistent motion from these additional tracked points which may arise from a rigid object, such as a vehicle, moving through the scene.

If any secondary motion is identified, the output detection map produced by the Motion HG is formed by marking detection regions around the current location of the set of points which are consistent with this secondary motion. The locations are marked onto the detection map using a top-hat function with radius q. As the Motion HG only tracks a sparse set of feature points, this final step is required to enlarge the tracked points in an attempt to span any vehicles detected.

7.2.2 Registration

Image registration is the process of applying a transformation to a set of images in order to align them to a common coordinate frame. In homogeneous coordinates the two dimensional point (x_0, x_1) is represented by the three vector $(x_0, x_1, 1)^T$. Using homogeneous coordinates, the perspective transform, which maps coordinates

$\mathbf{x}' \leftarrow \mathbf{x}$, may be described as

$$\begin{pmatrix} x'_0 \\ x'_1 \\ 1 \end{pmatrix} = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ 1 \end{pmatrix} \tag{7.7}$$

or alternatively through the relations

$$x'_0 = t_x(x_0, x_1) = \frac{a_{00}x_0 + a_{01}x_1 + a_{02}}{a_{20}x_0 + a_{21}x_1 + 1}, \tag{7.8a}$$

$$x'_1 = t_y(x_0, x_1) = \frac{a_{10}x_0 + a_{11}x_1 + a_{12}}{a_{20}x_0 + a_{21}x_1 + 1}. \tag{7.8b}$$

The parameters of the perspective transform can be obtained with a robust estimation process (such as RANSAC) given at least four known corresponding points from each image. Twelve pairs of corresponding points were chosen from each of the visible and TI images, and these were used within the MATLAB `cp2tform` function to obtain the perspective transformation matrix required to transform the TI image into the coordinate system of the Visible image.

Figure 7.2 shows a pair of Visible and TI images before registration, and after applying the resulting transform,

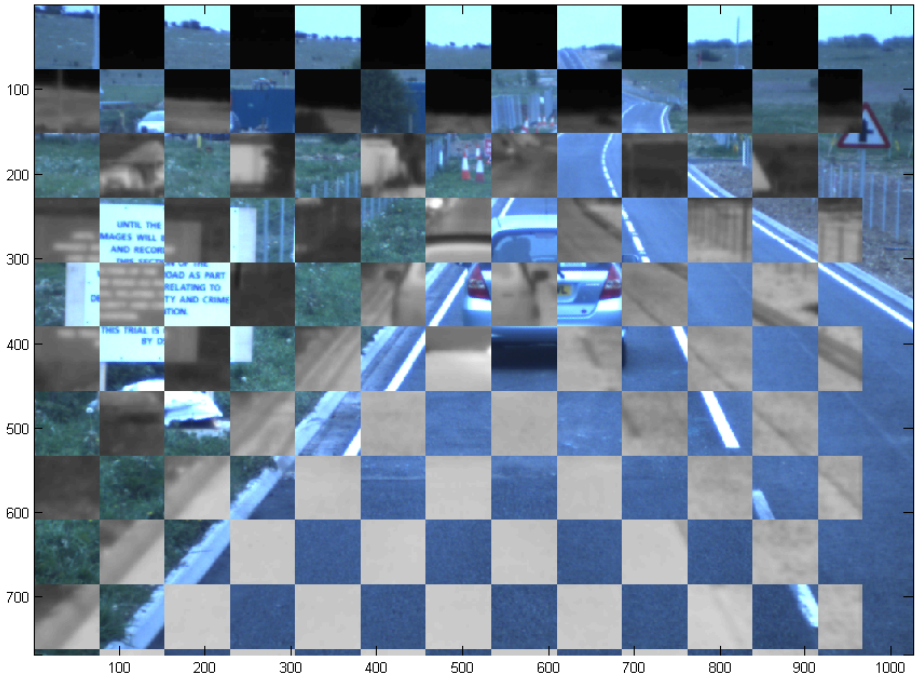
$$t_{\text{TI2Vis}} = \begin{pmatrix} 3.35833 & -0.0750325 & 48.9214 \\ -0.0578246 & 3.73689 & -90.4587 \\ 0.0 & 0.0 & 1.0 \end{pmatrix}. \tag{7.9}$$

One limitation of the perspective transform is that the transform is two dimensional – it registers one plane to another and so doesn't account for non-planar structures. From Figure 7.2 the ground can be considered approximately planar, and indeed the the road and horizon seem to be registered fairly well. However, objects such as the vehicle in the foreground and the signs on either side of the road, deviate from this plane and so there are greater registration errors around these objects. This will mean that there are later problems with fusion for large foreground objects, but the fusion of detections from more distant objects, which can be considered part of the ground plane, will be more successful.

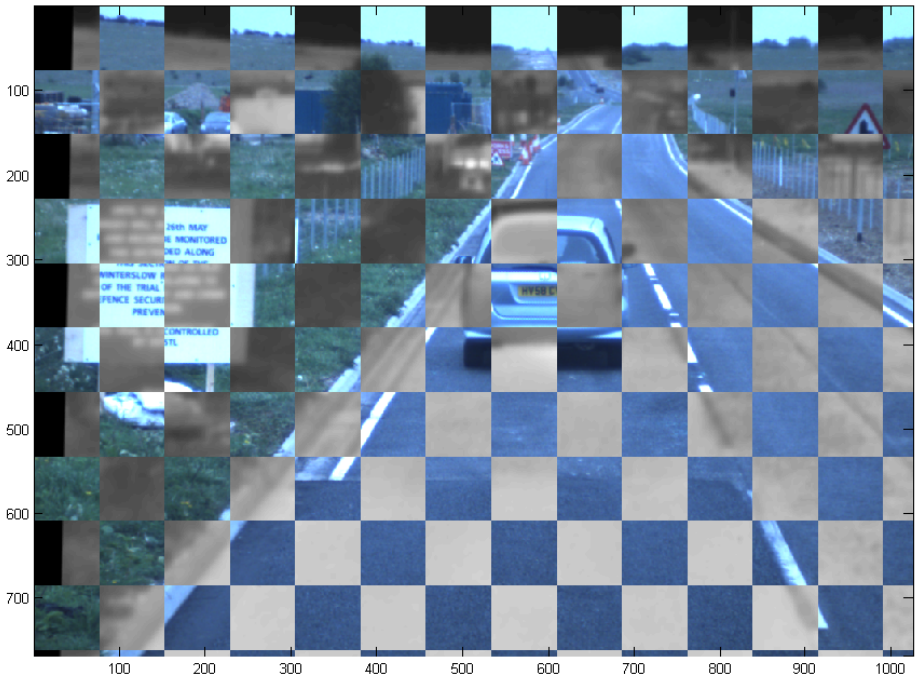
7.2.3 Hypothesis Combination

At this stage the set of detection maps generated at the HG stage, $\mathbf{y}'$, are combined into a single detection map, z

$$z = g(\mathbf{y}') \tag{7.10}$$



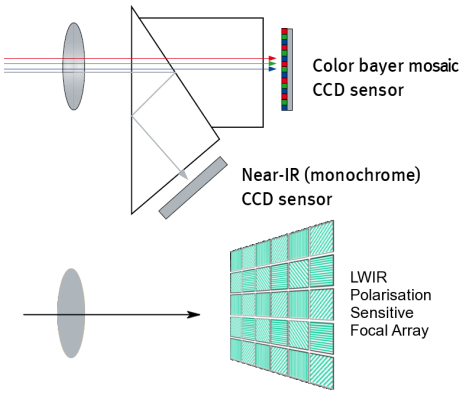
(a) Before registration.



(b) After registration.

Figure 7.2: An example TI and Visible image showing (a) resized but not registered and (b) TI registered to Visible with perspective transformation.

(a) Jai's AD-080 GE (upper) and Thales' Catherine MP (lower) mounted on a pan & tilt head.



(b) Optical diagram showing two different camera systems.

Figure 7.3: Demonstration of sensor setup on the trials van: the Visible and NIR sensors come through a single optical system to two CCD arrays, while the LWIR polarimeter is separate, using a polarisation sensitive focal array.

The hypothesis combination (HC) method, $g(.)$, is required to determine locations where the different HGs agree, while eliminating false detections from any individual HG. The current work uses the relatively simple voting method to combine detection maps. With this method, binary detection maps are combined additively and a threshold, t, is then applied which allows the user to select how many of the n HGs must agree for each pixel location to constitute a positive output pixel.

$$g(\mathbf{y}'(\mathbf{x})) = \begin{cases} 1 & \sum_{i=1}^{n} y_i'(\mathbf{x}) \geq t \\ 0 & \text{otherwise} \end{cases} \tag{7.11}$$

7.3 Methodology

The experimental goal is to determine the detection performance of the system, specifically how each HG performs independently, and how effective the HC stage is at eliminating false positive detections. The system was tested using a previously collected dataset where both sensors were securely mounted on top of the trials van and positioned to look at the road ahead. The sensor setup is demonstrated in Figure 7.3. The transformation matrices for the registration process were obtained offline as described in Section 7.2.2.

The sequence used extends for a total of 700 image frames, during which a number of vehicles are present at a variety of distances from the sensors. Ground truth data was created for the sequence using VATIC (the Video Annotation Tool

(a) Frame 100 (b) Frame 300

(c) Frame 500 (d) Frame 700

Figure 7.4: Example frames from test data with ground truth annotations, the captions are frame number. Note the presence of vehicles at a range of distances.

from Irvine, California [44]) in offline mode. Example frames from the dataset with ground truth annotations are shown in Figure 7.4, the ground truth rectangles outline areas of the image which contain vehicle pixels. Note the combination of a close range vehicle on the road, some medium distance vehicles parked to one side, and some distant vehicles on the road ahead.

Experiments were performed to evaluate each HG independently, and then test how different combinations of HGs altered the individual detection rates. The sequence of 700 frames was subsampled at five frame intervals to select a total of 141 frames over which the methods were evaluated. Performance was assessed using receiver operating characteristic (ROC) curves which plot the true positive rate (tpR) against the false positive rate (fpR) over a number of sensitivity levels.

The rates were defined based on the number of vehicle and non-vehicle pixels correctly identified

$$\text{tpR} = \frac{tp}{tp + fn} \tag{7.12a}$$

$$\text{fpR} = \frac{fp}{fp + tn} \tag{7.12b}$$

where tp, fp, tn, and fn are the number of true positive, false positive, true negative, and false negative pixels respectively. The ROC curves were generated by comparing detection maps produced at various sensitivities for each of the test frames with the ground truth detection maps. An ideal curve will pass through the point $(0, 1)$, and so a representative metric was also calculated as the shortest distance between this point and the ROC curve.

$$\text{metric} = \sqrt{(\text{fpR})^2 + (\text{tpR} - 1)^2} \tag{7.13}$$

7.4 Discussion

The detection results for each HG are presented in isolation in the following sections. Following this, an overall comparison of each HG is presented, and the performance benefits gained by fusing HGs are discussed.

7.4.1 Thermal HG

The thermal hypothesis generator is formed by using the LWIR I image as input in combination with a threshold HG. Results for both the local and global threshold HGs are shown alongside example output in Figure 7.5.

When comparing the ROC curves shown, recall that the axes of these plots are assessed on pixel-wise statistics. For both methods there are a large number of positive pixels covering the large vehicle in the foreground which have not been detected. Due to the scale of this vehicle, this significantly suppresses the true positive rate. Additionally, the extensive detection of the foreground road, blue container, and traffic sign in the global method increases its false positive rate.

As described in Section 7.2.1, the local threshold calculates the threshold applied to each pixel based on the mean intensity within a 3×3 pixel window. As such, regions of similar intensity which are larger than this scale will not be detected, regardless of their absolute intensity. Comparing the results of the global and local thresholds in Figure 7.5 the effect of this can be seen to suppress the detection of the high intensity road region towards the bottom of the image. However, it can also have the effect of fragmenting larger objects, or detecting anything of the appropriate scale. To successfully detect the large vehicle in the foreground, the HG may require to process the image at different scales, or an alternative approach which considers the detection of objects rather that pixels (such as those outlined in Chapters 5 and 6) may be required. However, both local and global methods have correctly identified the more distant vehicles to the left of the road.

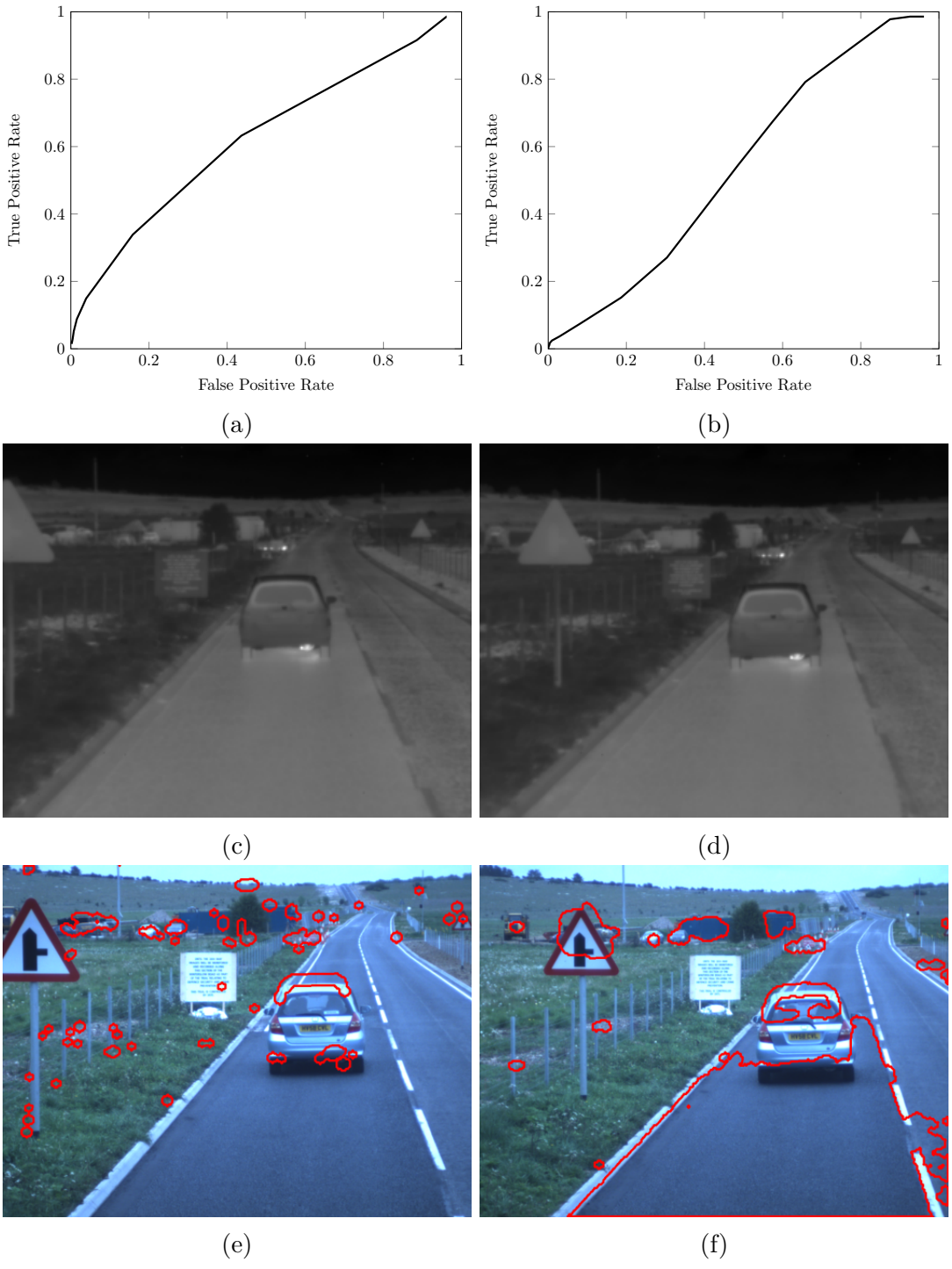


Figure 7.5: ROC curves for Thermal Intensity HG when using (a) local and (b) global thresholds with corresponding example input (c), (d) and output (e), (f) respectively.

7.4.2 Polarimetric HG

The polarimetric hypothesis generator is formed by using LWIR Q imagery with both the local and global threshold HGs. Results are shown alongside example output in Figure 7.6. The Stokes Q image used for this HG can be seen to be locally much more noisy than the thermal intensity image resulting in many false detections with the local threshold HG around the road region. However, comparing the ROCs curves for global and local threshold HGs, overall the local method displays a partially improved performance. This is likely to be due to the local threshold identifying a greater number of small false detections, while the global threshold will highlight the entire road region at a certain sensitivity setting.

The strong polarimetric signatures of vehicles highlighted in this chapter were the main considerations behind the subsequent development of a specific vehicle detection algorithm based on these signatures, as presented in Chapters 5 and 6.

7.4.3 Motion HG

The motion hypothesis generator uses visible (RGB) imagery as input and extracts motion signatures in an attempt to identify moving vehicles. The results from the motion HG are shown in Figure 7.7 alongside example tracked points and resultant detections. The tracks, denoted by white lines in Figure 7.7b, are terminated with either greed, red or blue markers. Tracks terminating with green dots are those considered to be consistent with the global camera motion as determined by the initial fundamental matrix estimation. Those terminating in red circles are outliers to global motion but are consistent with the second fundamental matrix and so are marked as secondary motion, the termination points form the detection output by marking a radius around the point as detailed in Algorithm 7.1, Operation 10. All tracks which are not consistent with either of the fundamental matrix estimations are considered outliers to both global and secondary motion and are terminated with blue dots.

These results show that this HG identifies consistent secondary motion for a number of points on the large moving vehicle in the foreground. Additionally a number of tracks that are inconsistent with both global and secondary motion (those terminating in a blue dot) are identified and discarded — these will typically correspond to tracking errors. However, a number of points which are part of the background are marked as having motion consistent with that of the vehicle, and additionally a number of pixels on the vehicle are not identified.

Note that the ROC curve presented in Figure 7.7a is not a typical shape in that it does not connect the points $(0, 0)$ and $(1, 1)$, instead only showing an increase in

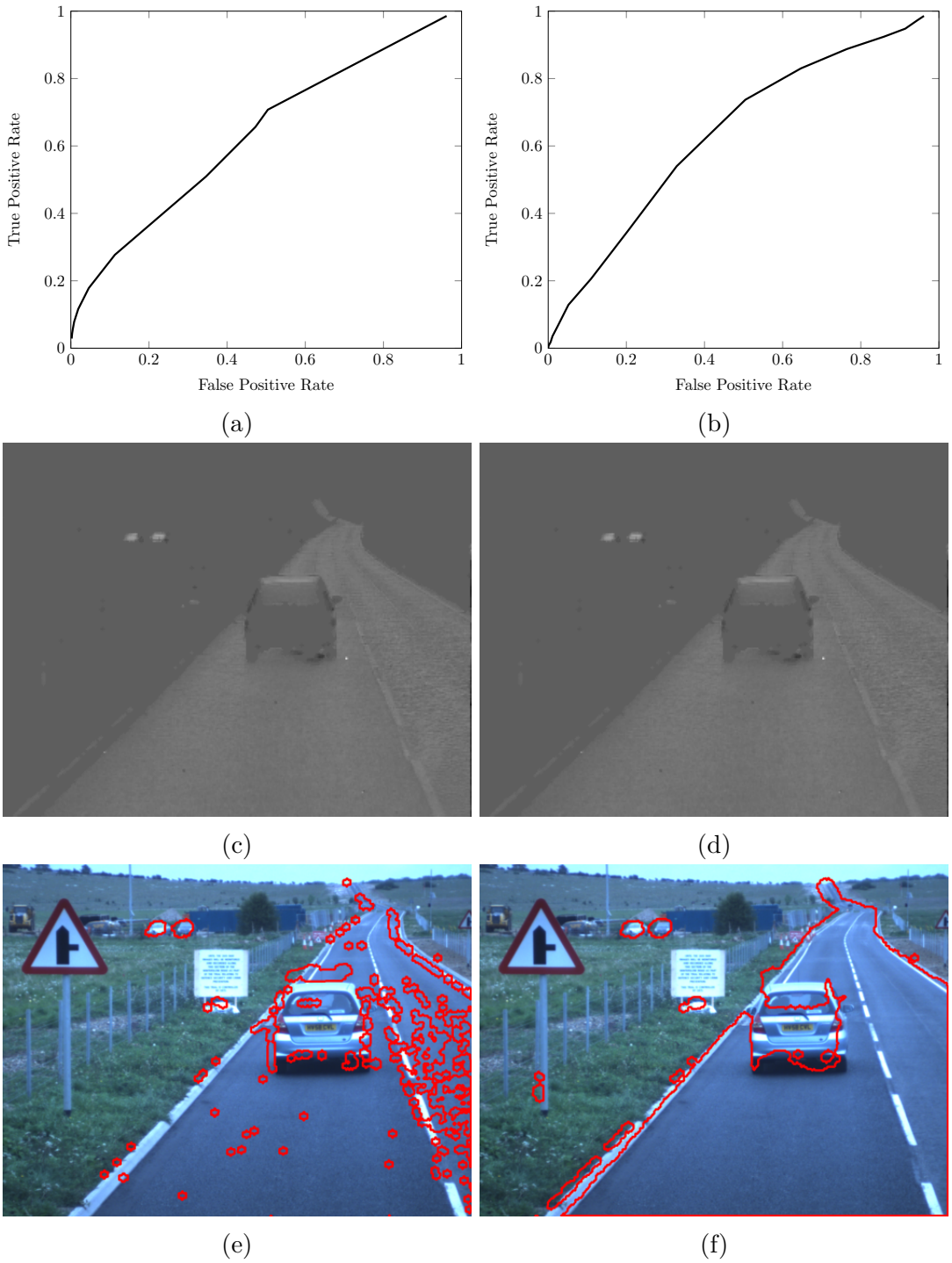
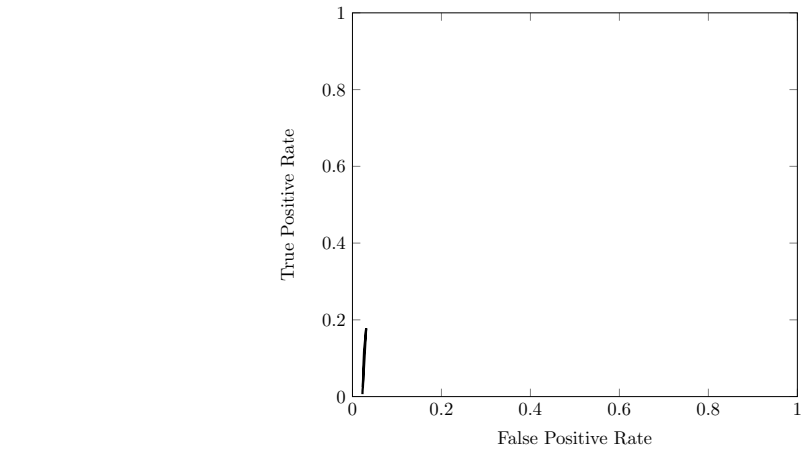


Figure 7.6: ROC curves for Polarimetric HG when using (a) local and (b) global thresholds with corresponding example input (c), (d) and output (e), (f) respectively.



(a)

(b) (c)

Figure 7.7: The ROC curve generated using the Motion HG is shown in (a), alongside an example (b) showing the tracks from which the output was generated. Tracked feature points are indicated by white lines terminated by large red circles if they are outliers to global motion and have consistent motion. Figure (c) then shows the output produced from this HG.

true positive rate at an approximately constant false positive rate. An ROC curve is normally obtained by varying a parameter which controls the number of positive and negative detections output by the algorithm. For the motion HG the parameter chosen determines the sensitivity of the function which finds the fundamental matrices. With a fundamental matrix, the end point of tracks can be estimated, and the sensitivity parameter chosen alters the maximum distance between the estimated and actual location of each end point allowed in order to consider the point an inlier for the estimated fundamental matrix. This parameter does have an effect on the true positive/false positive ratio but doesn't span the full ROC curve, and so perhaps an alternative parameter would be preferable.

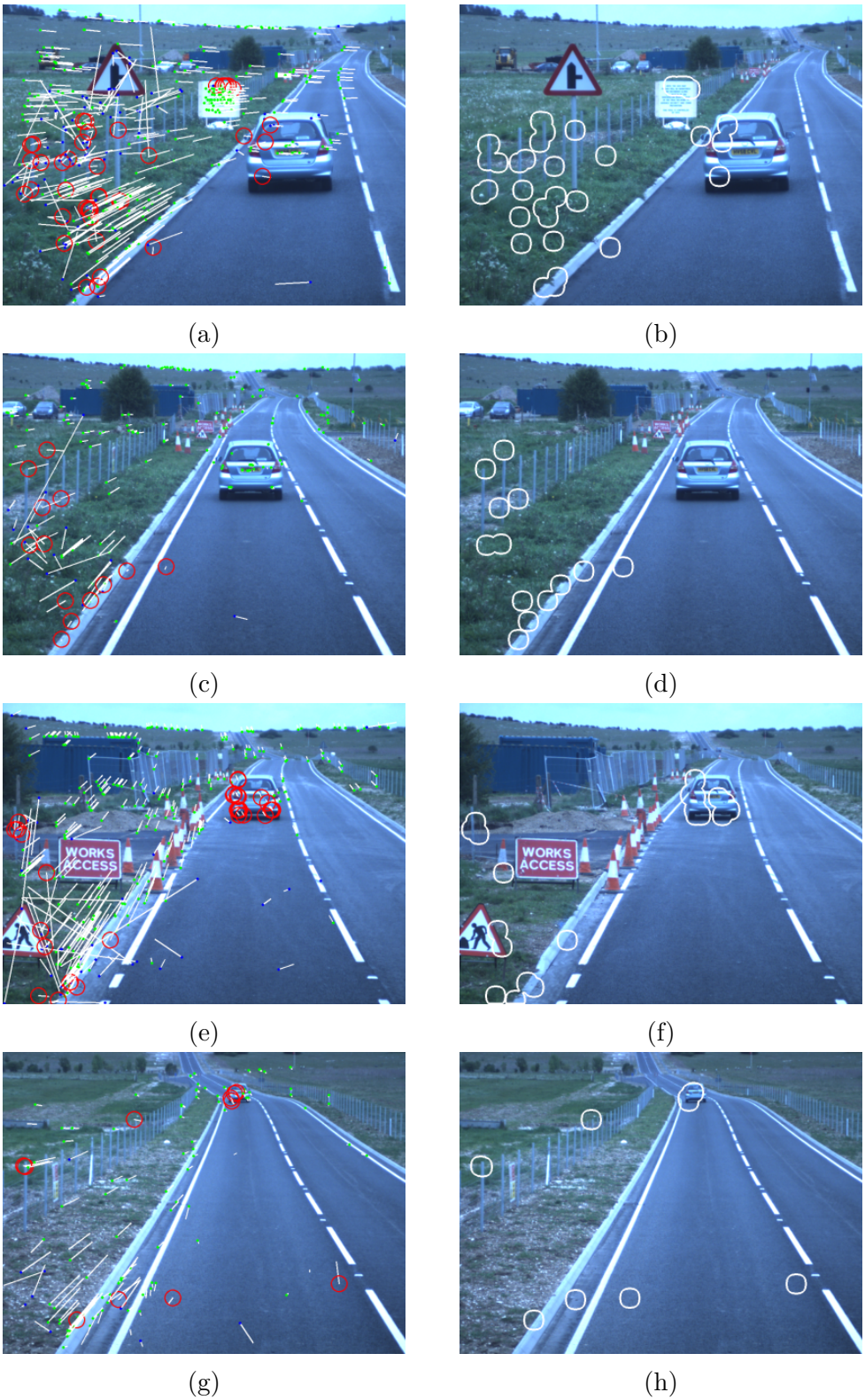
To help further qualitatively assess the performance of this HG, four additional examples of tracked points and their associated output detections are shown in Figure 7.8.

The first point to note is that the overall method is reliant on a good distribution of a number of tracks spanning both the background and any moving object. The first fundamental matrix estimation to separate the background motion from everything else is fairly reliable, all tracks terminating in green dots in Figure 7.8 are qualitatively consistent with the motion of the camera. The difficulty arises in identifying any secondary motion amongst the noise created by tracking errors. If a number of tracking errors exist, such as the case in Figure 7.8a, any tracks arising from secondary motion can become obscured by the noise.

As a second consideration, note that in Figure 7.8c tracks of points on the vehicle have been associated with the background motion of the scene, perhaps due to the small difference in relative motion. This leads to a number of frames where the vehicle blends in with the background in terms of the motion signature. These errors could be overcome by tracking points for a greater number of frames – by allowing the moving object to move a greater distance, any differences between the background and object locations will be accentuated.

In Figure 7.8g the vehicle is more distant, and in this case tracks on the vehicle are successfully identified as consistent secondary motion. Additionally, as the vehicle is present at a small scale, the fixed size detection superimposed on each track covers most vehicle pixels. This is in contrast with, for example, Figure 7.8e where although the vehicle centred tracks are identified as a secondary motion in the scene, the output detections do not span the entire vehicle.

Additionally, in all examples of Figure 7.8 there are background points which have been associated with the secondary motion within the scene and thus marked as detections. This suggests an extension to this HG could consider an additional stage for this algorithm, to dismiss isolated points which are detected but most likely due to tracking errors, and associate points which are spatially close and have



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 7.8: Example output from the motion HG. Images on the left show the full set of tracks from the sequences, with the end point of each track coloured green for background motion, red for consistent non-background motion, and blue for noise. The images on the right are the output detections marked around the current location of tracks determined to arise from secondary motion in the scene.

Table 7.1: This table shows the true and false positive rates, alongside the metric, for system runs consisting of different combination of detectors, and combination thresholds. PLocal is the Polarisation HG with local threshold; TLocal is the Thermal HG with local threshold; Motion is the Motion HG; t defines how many HGs must agree for a final detection to be marked; tpR, fpR, and metric are the true positive rate, false positive rate, and metric as defined in (7.12a), (7.12b), and (7.13) respectively. The ideal metric is 0.0.

PLocal	TLocal	Motion	t	tpR	fpR	metric
y	-	-	1	0.7077	0.5043	0.5829
-	y	-	1	0.6323	0.4363	0.5706
-	-	y	1	0.1778	0.0308	0.8227
y	y	y	2	0.1380	0.0081	0.8621
y	y	-	2	0.0828	0.0038	0.9173
y	y	y	3	0.0200	0.0002	0.9800

a motion which is not consistent with global motion.

7.4.4 Overall Comparison

A comparison of the optimum detection rates for each of the three HGs, along with different combinations is provided in Table 7.1. These results are displayed graphically in Figure 7.9.

Comparing the ROC curves of the individual HGs, the Thermal HG and Polarimetric HG have similar performance. As discussed above, a low true positive rate is observed as there are a large number of pixels, primarily on the large foreground vehicle, which are not detected. These HGs detect the smaller vehicles more distant in the scene, and smaller components (such as the roof, tires and exhaust) of the larger vehicle.

Note that in Figure 7.6 many false positive detections are made on the road region, which exhibits a significant polarimetric signature; while in Figure 7.5, the false positive detections arise from different objects in the scene which exhibit a thermal signature.

When the thermal and polarimetric HGs are combined, a detection is only marked when both HGs individually mark detections and so a large reduction in both true positive and false positive rates is observed (the result in row four in Table 7.1 and the $I + Q$ result in Figure 7.9). By qualitatively examining the individual and fused outputs from these HGs, as shown in Figure 7.10, it is clear that a significant quantity of the false positive results have been eliminated, while vehicle pixels remain as detections. The exception to this being the central region of the large vehicle in the foreground. Note that in Figure 7.10 there are two vehicles to the left of the traffic sign, and two to the right, all four of which are marked as detec-

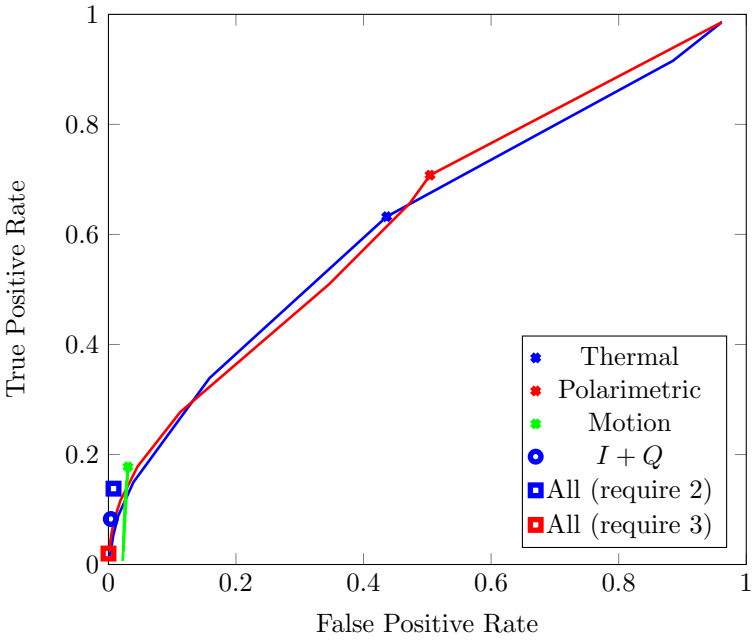


Figure 7.9: The ROC curves for each HG are reproduced here with the lowest metric points highlighted alongside some combined results of the proposed system. Note that the $x-$axis is plotted on a logarithmic scale for clarity. The data plotted is from Table 7.1.

tions in the final output, while much of the false positive detections arising from the cluttered grass and road regions have been removed. From this it can be suggested that this approach of pixel based vehicle detection through the fusion of thermal and polarimetric HGs is successful but only for smaller more distant vehicles in the scene. To detect larger vehicles either this method needs to alter the scale of the image, or an alternate approach is required.

The motion HG has been discussed above in terms of its individual performance. When used in combination with the thermal and polarimetric HGs (referring to the square points in Figure 7.9) a reduction in both true and false positive rates is observed. One issue with the motion HG is that, even if perfect, it would only

(a) (b) (c)

Figure 7.10: Maps of detected pixels from (a) thermal and (b) polarimetric HGs and the combined system output (c).

detect vehicles which are in motion.

7.5 Conclusion

This chapter has presented a case study investigating how multiple types of imagery could be processed by different detection algorithms in a fusion framework such that the overall vehicle detection performance is improved. In this case study, each HG provided output in terms of a binary detection map, highlighting pixels which were considered to be part of a vehicle, and these detection maps were naively combined in a pixel-level fusion process. Given its simplicity, it has been shown that this fusion process does significantly reduce the overall false positive rate, but at the same time the true positive rate is also reduced.

The simple detectors derived from the thermal intensity and polarisation imagery were shown to be effective at identify areas of the image likely to contain vehicles. This is particularly evident in Figure 7.6 where the two vehicles in the region to the left of the road are very distinct from the background. By extending the thermal polarimetric detector to incorporate some of the more advanced descriptors developed in previous chapters it is anticipated that the performance will continue to improve.

The motion HG presented in this chapter attempted to incorporate some additional information by using the visible imagery to detect moving objects in the scene. It was acknowledged as being a fairly simple implementation of the more complex multi-body structure from motion algorithms in the literature, with two major limitations including the sparse set of points which are analysed, and only two motions being identified (the global motion, and one additional motion). By only considering the motion of the set of points originally detected, areas of the image are not analysed, and only a sparse set of points on regions of interest are covered. The use of a dense optical flow algorithm could improve this as the motion of the whole scene could be considered, not just the set of feature points detected and tracked. Additionally, by identifying regions of locally consistent motion, rather than a set of global points corresponding to similar motions, it should be possible to remove the sparse set of erroneous detections which are currently being detected in addition to the vehicle.

In terms of quantitative results, the current approach is penalised by its simplicity - the polarimetric threshold HG only detects pixels on the roof of the vehicle, while the thermal HG only detects the exhaust and tires. Grouping these detections, along with the motion estimates of a region of pixels should improve the results further. Additionally, it is clear that some detectors have different performance to

others, and yet this has not been captured in the fusion. As such, adding some form of trustworthiness assessment is also expected to improve the results further. However, even the simple combination of HGs presented in this chapter has been shown to significantly reduce the number of false positive detections and so the concept of fusing the output of different detection algorithms operating on different forms of imagery is thought to be worthwhile pursuing further.

Chapter 8
Conclusion

8.1 Summary

In considering methods of object detection using different types of sensor, it was noted that a number of previous publications had highlighted the utility of the polarisation properties of light for a variety of applications including segmentation [5, 91, 92], anomaly detection [95–98], and target identification [102, 103]. Given this, the two questions considered in this thesis were:

1. Could a successful vehicle detection algorithm be developed using infrared polarimetric information, and would this show improved performance above algorithms processing infrared intensity imagery alone?

2. Could the polarimetric content of a scene be combined with other information, such as from infrared intensity, or visible imagery, to reduce the false alarm rate of detections?

Much of the content of this thesis revolved around the use of long wave infrared polarimetric imagery as obtained from the Thales Catherine MP LWIR Polarimeter. This device was introduced in Chapter 3 and a number of experiments designed to characterise and understand its output were presented. The detector response was shown to be linear with incident irradiance. The focal plane array was explained to be made up of a series of individual linear polarising elements oriented in one of four directions. It was found that the polarisation contrast of these elements was different for different orientations of polarisation sensitivity and so a correction was required in order to measure the true polarisation state of the scene. Finally, the theory of polarisation upon reflection and emission, as introduced in Chapter 2, was investigated with a number of simple experiments. These experiments consisted of observing different materials under different ratios of emission and reflection as controlled by the relative temperatures of the object and the optical background.

An overview of the object detection literature, in terms of training classifiers using features computed from the imagery, was presented in in Chapter 4. Following this, Chapter 5 described the first major novel content of this thesis. In this Chapter, an understanding of the polarimetric signatures of vehicles was developed, leading to a series of histogram-based descriptors for vehicle detection. The histogram-based descriptors were found to be most successful when used without spatial cells spanning the window. However by removing the cells, the spatial information contained in the vehicle signature was also removed. To overcome this an alternative novel cluster-based descriptor was developed in Chapter 6.

The cluster-based approach, which was shown to be the more accurate than the histogram-approach, identified vehicles by detecting clusters of pixels which were close in space and angle of polarisation with a higher degree of polarisation than the surrounding area. Once these clusters were detected, a descriptor based on their characteristics was proposed and was found to improve the discrimination of vehicles from the background beyond other methods considering only the thermal intensity.

Finally, Chapter 7 presented an initial investigation into how different sensors and different detection algorithms could be combined to reduce the overall false alarm rate. Three relatively simple detectors were introduced, based on a local threshold on the thermal intensity and degree of polarisation, and on motion cues derived from visible imagery. It was shown that each detection pipeline revealed different false positive detections, and so their combination did reduce the overall false positive rate. However, due to the simplicity of the detectors, not all vehicles were detected in their entirety and so some improvements were suggested.

8.2 Contributions

The principal contributions presented in this thesis are as follows:

Polarimetric Histogram During the initial study of histogram-based approaches for vehicle detection, it is shown in Section 5.4 that the use of a two dimensional histogram of the degree and angle of polarisation across a window allows more accurate detection of vehicles in thermal polarimetric imagery when compared to descriptors which process intensity imagery alone. Results from this work were presented at the IET Intelligent Signal Processing Conference 2013 [104] and at the International Symposium on Optronics in Defence and Security (Optro) 2014 [105].

Cluster-based descriptor The main problem with the Polarimetric Histogram was the lack of spatial information retained. The cluster-based approach, presented in Chapter 6, addresses this by locating clusters of pixels with high

degree of polarisation which are spatially close and have similar angles of polarisation. These clusters were shown to naturally form on flat vehicle surfaces, and by training a descriptor based on the patterns of these clusters, a reliable vehicle detector was demonstrated.

Combination of sensors and detectors In Chapter 7 a system is presented which passes input from multiple types of imaging sensor through different detection pipelines to produce a set of detection maps which are then combined and presented as output. Two methods based on detecting local hotspots in the thermal intensity and polarimetric imagery were shown to be relatively successful individually but when combined significantly reduce the false positive rate. A third method, based on detecting moving objects from a moving camera, was shown to be able to detect a foreground vehicle relatively successfully but also marked a number of false positives due to tracking errors. This work was presented at the SPIE Security & Defence Conference 2012 [125].

8.3 Future Research Directions

This project naturally leads to a number of interesting research opportunities which may be explored in extension of this work.

8.3.1 Extensions to the Polarimetric Vehicle Detector

The cluster-based polarimetric descriptor presented in Chapter 6 offers a number of possible opportunities for extension.

Different object classes The descriptor was developed to detect civilian vehicles. However, the detection of a number of other object classes may benefit from the use of a polarimetric descriptor. Given a suitable dataset any object class observed to be significantly polarising could be incorporated, potentially including different types of road vehicles, trains, marine vessels, or potentially aircraft.

Different camera locations The current work focussed on improving the detection performance for a camera mounted on a moving land-based vehicle. However, such a polarimetric camera could be mounted on a static road-side location for traffic monitoring, or on-board an aircraft. The effect these different camera locations would have on the accuracy of the detector could be investigated, and the features extracted could be adapted accordingly.

8.3.2 Extensions to the Multi-modal Detection System

The multi sensor-multi detector framework presented in Chapter 7 could also be extended in a number of directions.

Three dimensional registration The current system registers all detection maps as planar images onto the image plane of the visible sensor for fusion. This allows for simple transformation and display to the user, but does introduce errors arising from a non-planar scene. An extension could be to instead project detections from each sensor onto a three dimensional map of the scene and perform fusion in this three dimensional coordinate system. A map of the local three dimensional topography could be obtained using structure from motion algorithms and would more naturally allow the integration of detections from different non-imaging sensors.

Addition of different sensors Leading on from the previous point, the system as it was presented made use of a relatively small number of sensors. Extending this by adding additional sensor types such as lidar or radar, or duplicating existing sensors to create a stereo pair would provide more information available to detectors.

Additional detection algorithms The hypothesis generation algorithms (detectors) presented were fairly simplistic. Extending the complexity of detector, or introducing new detectors to process one or more input stream could allow additional redundancy and could potentially allow different types of object to be detected. However, this may require a more advanced, probabilistic method of combining detections.

8.3.3 Combination of Components

As a final suggestion for a future direction of research, it may be possible to combine elements of both the multi-modal detection system and the polarimetric vehicle detector.

Hypothesis generation and verification A two-stage approach could consist of initially identifying locations of interest within the scene using the multi-modal detector, and then applying the polarimetric detector to confirm the identify of any identified regions.

References

[1] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sensing of Environment*, vol. 8, no. 2, pp. 127–150, 1979. 1

[2] J. Letham, N. M. Robertson, and B. Connor, "Contextual smoothing of image segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. IEEE Comput. Soc. Press, 2010, pp. 7–12. 1, 124

[3] C. Fredembach and S. Süsstrunk, "Colouring the Near-Infrared," in *Proceedings of the IS&T/SID 16th Color Imaging Conference*, 2008, pp. 176–182. 1

[4] F. Somma, P. Aloe, and G. S. Spagnolo, "Defects in UV-vis-NIR reflectance spectra as method for forgery detections in writing documents," *Journal of Physics: Conference Series*, vol. 249, p. 012060, 2010. 1

[5] P. Terrier, V. Devlaminck, and J. M. Charbois, "Segmentation of rough surfaces using a polarization imaging system." *J. Opt. Soc. Am. A*, vol. 25, no. 2, pp. 423–30, Feb. 2008. 2, 65, 146

[6] G. A. Atkinson and E. R. Hancock, "Recovery of surface orientation from diffuse polarization," *IEEE Trans. Image Proc.*, vol. 15, no. 2006, pp. 1653–1664, 2006. 2

[7] ——, "Shape estimation using polarization and shading from two views." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 2001–17, Nov. 2007. 2, 36

[8] J. C. Maxwell, P. Trans, and R. S. Lond, "A Dynamical Theory of the Electromagnetic Field," *Phil. Trans. R. Soc. Lond.*, vol. 155, pp. 459–512, 1865. 4

[9] E. Hecht, *Optics*, 3rd ed. Addison Wesley Longman, Inc., 1998. 4, 13, 15

[10] M. Born and E. Wolf, *Principles of Optics*, 6th ed. Pergamon Press Ltd., 1980. 4, 9, 15

[11] R. M. A. Azzam and N. M. Bashara, *Ellipsometry and polarised light*. North Holland Publishing Company, 1977. 5, 9, 11, 13

[12] M. G. Gartley, "Polarimetric Modeling of Remotely Sensed Scenes in the Thermal Infrared," Ph. D., Rochester Institute of Technology, 2007. 11, 20

[13] D. L. Jordan and G. Lewis, "Measurements of the effect of surface roughness on the polarization state of thermally emitted radiation." *Opt. Lett.*, vol. 19, no. 10, pp. 692–694, 1994. 11, 20

[14] G. W. Kattawar, "A Search for Circular Polarization in Nature," *Opt. Photonics News*, vol. 5, no. 9, pp. 42–43, 1994. 11

[15] B. J. DeBoo, J. M. Sasian, and R. A. Chipman, "Depolarization of diffusely reflecting man-made objects." *Appl. Opt.*, vol. 44, no. 26, pp. 5434–5445, Sep. 2005. 11

[16] R. Walraven, "Polarization imagery," *Optical Engineering*, vol. 20, no. 1, pp. 14–18, 1981. 11

[17] M. Planck, "Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum," *Verh. Dtsch. phys. Ges.*, vol. 2, no. 237-245, 1900. 14

[18] ——, "On the Law of Distribution of Energy in the Normal Spectrum," *Annalen der Physik*, vol. 4, pp. 553–563, 1901. 14

[19] R. A. Millikan, "A study of the polarization of the light emitted by incandescent solid and liquid surfaces I," *Phys. Rev.*, vol. 3, pp. 81–99, 1895. 15

[20] ——, "A study of the polarization of the light emitted by incandescent solid and liquid surfaces II," *Phys. Rev.*, no. 3, pp. 177–192, 1895. 15

[21] O. Sandus, "A Review of Emission Polarization," *Appl. Opt.*, vol. 4, no. 12, p. 1634, Dec. 1965. 15

[22] D. C. Bertilone, "Stokes parameters and partial polarization of far-field radiation emitted by hot bodies," *J. Opt. Soc. Am. A*, vol. 11, no. 8, pp. 2298–2304, 1994. 15

[23] L. B. Wolff, A. J. Lundberg, and R. Tang, "Image understanding from thermal emission polarization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1998, pp. 625–631. 15

[24] A. Resnick, C. Persons, and G. Lindquist, "Polarized Emissivity and Kirchhoff's Law," *Applied Optics*, vol. 38, no. 8, p. 1384, Mar. 1999. 15

[25] P. J. Leonard, "Refractive indices, Verdet constants, and Polarizabilities of the inert gases," *Atomic Data and Nuclear Data Tables*, vol. 14, no. 1, pp. 21–37, 1974. 18

[26] J. Zhang, Z. H. Lu, and L. J. Wang, "Precision refractive index measurements of air, N_2, O_2, Ar, and CO_2 with a frequency comb," *Applied Optics*, vol. 47, no. 17, pp. 3141–3151, 2008. 18

[27] P. Křen, "Comment on "Precision refractive index measurements of air, N_2, O_2, Ar, and CO_2 with a frequency comb"," *Applied Optics*, vol. 50, no. 35, pp. 6484–6485, 2011. 18

[28] G. M. Hale and M. R. Querry, "Optical Constants of Water in the 200 nm to 200 µm Wavelength Region," *Applied Optics*, vol. 12, no. 3, pp. 555–563, 1973. 18

[29] H. H. Li, "Refractive index of alkali halides and its wavelength and temperature derivatives," *J. Phys. Chem. Ref. Data*, vol. 5, no. 2, pp. 329–528, 1976. 18

[30] P. Frintz, "Über Brechungsindizes und Absorptionskonstanten des Diamanten zwischen 644 und 226 µm," *Zeitschrift für Physik*, vol. 15, no. 1-2, pp. 358–368, 1923. 18

[31] A. D. Rakic, A. B. Djurišic, J. M. Elazar, and M. L. Majewski, "Optical Properties of Metallic Films for Vertical-Cavity Optoelectronic Devices," *Applied Optics*, vol. 37, no. 22, pp. 5271–5283, 1998. 18

[32] J. S. Tyo, B. M. Ratliff, J. K. Boger, W. T. Black, D. L. Bowers, and M. P. Fetrow, "The effects of thermal equilibrium and contrast in LWIR polarimetric images," *Opt. Express*, vol. 15, no. 23, pp. 15 161–15 167, Nov. 2007. 18

[33] K. E. Torrance and E. M. Sparrow, "Theory for off-specular reflection from roughened surfaces," *J. Opt. Soc. Am.*, vol. 57, no. 9, pp. 1105–1114, 1967. 20

[34] J. Maxwell, J. Beard, S. Weiner, D. Ladd, and S. Ladd, "Bidirectional Reflectance Model Validation and Utilization," Tech. Rep., 1973. 20

[35] R. G. Priest and T. A. Germer, "Polarimetric BRDF in the Microfacet Model: Theory and Measurements," in *Proc. Military Sensing Symposia Specialty Group on Passive Sensors*, vol. 1, no. August, 2000, pp. 169–181. 20

[36] K. P. Gurton, R. Dahmani, and G. Videen, "Measured Degree of Infrared Polarization for a Variety of Thermal Emitting Surfaces," U.S. Army Research Laboratory, Adelphi, MD, Tech. Rep. June, 2004. 20, 33

[37] M. G. Gartley, S. D. Brown, A. D. Goodenough, N. J. Sanders, and J. R. Schott, "Polarimetric scene modeling in the thermal infrared," *Proc. SPIE*, vol. 6682, p. 66820C, 2007. 20, 33

[38] S. Crawford, R. Craig, A. Haining, J. Parsons, E. Costart, P. Bois, F.-H. Gauthier, and O. Cocle, "Thales long-wave advanced IR QWIP cameras," in *Proc. SPIE*, 2006, p. 62060H. 22

[39] M. Z. Tidrow, "Device physics and state-of-the-art of quantum well infrared photodetectors and arrays," *Materials Science and Engineering: B*, vol. 74, no. 1-3, pp. 45–51, May 2000. 22

[40] J. Parsons and R. Craig, "A LWIR polarimetric imager," in *Proc. EMRS DTC Conf.*, Edinburgh, 2008. 23, 24, 27, 29

[41] I. D. Carrie and B. Connor, "Polarimetric imaging on the move," in *Proc. Int. Symp. Optronics in Defence and Security*, Paris, France, 2010. 23

[42] M. Ferraton, C. Stolz, and F. Meriaudeau, "Surface Reconstruction of Transparent Objects by Polarization Imaging," in *Proc. IEEE Int. Conf. Signal Image Technology and Internet Based Systems*, no. 3, Nov. 2008, pp. 474–479. 36

[43] C. P. Huynh, A. Robles-Kelly, and E. Hancock, "Shape and Refractive Index Recovery from Single-View Polarisation Images," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010, pp. 1229–1236. 36

[44] C. Vondrick, D. Patterson, and D. Ramanan, "Efficiently Scaling Up Crowdsourced Video Annotation," *Int. J. Comput. Vision*, vol. 101, no. 1, pp. 184—-204, 2012. 38, 134

[45] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A Local Contrast Method for Small Infrared Target Detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 574–581, Jan. 2014. 42

[46] B. Tian, Y. Li, B. Li, and D. Wen, "Rear-View Vehicle Detection and Tracking by Combining Multiple Parts for Complex Urban Surveillance," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 597–606, 2014. 42

[47] M. J. Leotta and J. L. Mundy, "Vehicle Surveillance with a Generic, Adaptive, 3-D Vehicle Model." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1457–1469, Nov. 2010. 42, 53, 118

[48] B. Southall, M. Bansal, and J. Eledath, "Real-time Vehicle Detection for Highway Driving," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 541–548. 42, 48

[49] V. Milanés, D. F. Llorca, J. Villagrá, J. Pérez, C. Fernández, I. Parra, C. González, and M. a. Sotelo, "Intelligent automatic overtaking system using vision for vehicle detection," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3362–3373, Feb. 2012. 42

[50] C. Conaire, N. O'Connor, E. Cooke, and A. Smeaton, "Multispectral Object Segmentation and Retrieval in Surveillance Video," *Proc. Int. Conf. Image Processing*, pp. 2381–2384, Oct. 2006. 42

[51] J. Philbin, M. Isard, and J. Sivic, "Descriptor learning for efficient retrieval," in *Proc. Euro. Conf. Computer Vision*, 2010. 42

[52] S. Theodoridis and K. Koutroumbas, *Pattern recognition*. Academic Press, 2006. 42

[53] S. Sivaraman and M. M. Trivedi, "Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, 2013. 43

[54] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: an evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–61, Apr. 2012. 43, 47, 59

[55] P. Viola and M. Jones, "Robust Real-Time Face Detection," *Int. J. Comput. Vision*, vol. 57, no. 2, pp. 137–154, May 2004. xvii, 43, 44, 45, 55, 56

[56] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005, pp. 886–893. 43, 47, 69, 71

[57] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. IEEE Conf. Computer Vision*

and Pattern Recognition, vol. 2. Citeseer, 2003, pp. II–264–II–271. 43, 48, 49, 50

[58] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997. 44, 55

[59] S. Matzka, "Efficient Resource Allocation For Automotive Active Vision Systems," PhD, Heriot-Watt University, 2009. 45

[60] T. P. Breckon, S. E. Barnes, M. L. Eichner, and K. Wahren, "Autonomous Real-time Vehicle Detection from a Medium-Level UAV," in *Proc. Int. Conf. Unmanned Air Vehicle Systems*, 2009, pp. 29.1–29.9. 45, 46

[61] A. Gąszczak, T. Breckon, and J. Han, "Real-time People and Vehicle Detection from UAV Imagery," in *Proc. SPIE Conference Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*, 2011, p. 78780B. 45

[62] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *Int. J. Comput. Vision*, vol. 38, no. 1, pp. 15–33, 2000. 47

[63] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral Channel Features," *Proc. British Machine Vision Conference*, pp. 91.1–91.11, 2009. 47

[64] N. Buch, K. Thames, and S. A. Velastin, "3D Extended Histogram of Oriented Gradients (3DHOG) for Classification of Road Users in Urban Scenes," in *Proc. British Machine Vision Conference*, 2009. 48, 51

[65] J. Arróspide, L. Salgado, and M. Nieto, "Video analysis-based vehicle detection and tracking using an MCMC sampling framework," *EURASIP Journal on Advances in Signal Processing*, pp. 1–20, 2012. 48

[66] I. Creusen, R. Wijnhoven, E. Herbschleb, and P. de With, "Color exploitation in hog-based traffic sign detection," in *Proc. IEEE Conf. Image Processing*, 2010, pp. 2669–2672. 48

[67] M. Villamizar, J. Scandaliaris, A. Sanfeliu, and J. Andrade-cetto, "Combining Color-Based Invariant Gradient Detector with HoG Descriptors for Robust Image Detection in Scenes under Cast Shadows," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2009, pp. 1997–2002. 48

[68] T. Kadir and M. Brady, "Saliency, scale and image description," *Int. J. Comput. Vision*, vol. 45, no. 2, pp. 83–105, 2001. 48

[69] I. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002. 48, 62

[70] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *Proc. Euro. Conf. Computer Vision.* Springer, 2004, pp. 228–241. 48

[71] Y. Xiang, R. Mottaghi, and S. Savarese, "Benchmark for 3D Object Detection in the Wild," in *Proc. IEEE Winter Conference on Applications of Computer Vision*, Steamboat Springs, CO, USA, 2014, pp. 75–82. 51

[72] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets Robotics: The KITTI Dataset," *Int. J. Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013. 51

[73] N. Buch, J. Orwell, and S. A. Velastin, "Urban Road User Detection and Classification using 3D Wire Frame Models," *IET Computer Vision*, vol. 4, no. 2, pp. 105–116, 2010. 51

[74] N. E. Buch, "Classification of Vehicles for Urban Traffic Scenes," Doctor of Philosophy, Kingston University London, 2010. 52

[75] T. K. Marks, J. R. Hershey, and J. R. Movellan, "Tracking motion, deformation, and texture using conditionally gaussian processes." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 348–63, Feb. 2010. 53

[76] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," in *Proc. Annual ACM Workshop on Computational Learning Theory.* ACM Press, 1992, pp. 144–152. 55, 56

[77] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995. 55, 56

[78] C. M. Bishop, *Neural Networks for Pattern Recognition.* New York, NY, USA: Oxford University Press, Inc., 1995. 55, 59

[79] T. Mita, T. Kaneko, B. Stenger, and O. Hori, "Discriminative feature co-occurrence selection for object detection." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1257–69, Jul. 2008. 55

[80] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning.* Springer New York Inc., 2001. 58, 97

[81] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proc. Int. Joint Conf. Neural Networks*, Washington, DC, USA, 1989, pp. 593–605. 59

[82] M. Everingham, L. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *Int. J. Comput. Vision*, vol. 88, no. 2, pp. 303–338, Sep. 2009. 61

[83] L. B. Wolff, "Polarization vision: a new sensory approach to image understanding," *Image Vision Comput.*, vol. 15, no. 2, pp. 81–93, Feb. 1997. 62, 63

[84] B. Connor, I. Carrie, R. Craig, and J. Parsons, "Discriminative imaging using a LWIR polarimeter," in *Proc. SPIE*. Spie, 2008, p. 71130K. 62

[85] I. Carrie and B. Connor, "Detection of Difficult Targets Using a QWIP Based Polarimeter," in *Proc. SEAS DTC Technical Conf.*, Edinburgh, 2009, p. B18. 62

[86] B. Barbour, S. Kordella, M. Dorsett, and B. Kerstiens, "Mine detection using a polarimetric IR sensor," in *Proc. EUREL Int. Conf. The Detection of Abandoned Land Mines: A Humanitarian Imperative Seeking a Technical Solution*, no. 431, 1996, pp. 78–82. 62

[87] W. D. Jong, F. Cremer, K. Schutte, and J. Storm, "Usage of polarisation features of landmines for improved automatic detection," in *Proc. SPIE: Det. and Rem. Techn. for Mines and Minelike Targets V 4038*, 2000, pp. 1–12. 62

[88] D. A. Lavigne, M. Breton, M. Pichette, V. Larochelle, and J.-R. Simard, "Enhanced Military Target Discrimination using Active and Passive Polarimetric Imagery," *Proc. IEEE Int. Symp. Geoscience and Remote Sensing*, pp. V–354 – V–357, Jul. 2008. 64

[89] D. A. Lavigne, M. Breton, G. Fournier, J.-F. Charette, M. Pichette, V. Rivet, and A.-P. Bernier, "Target discrimination of man-made objects using passive polarimetric signatures acquired in the visible and infrared spectral bands," in *Proc. SPIE*, vol. 8160, no. 1, 2011, p. 816007. 64

[90] B. Connor, J. Letham, N. Robertson, and I. Carrie, "Scene understanding and task optimisation using multimodal imaging sensors and context: a real-time implementation," in *Proc. SPIE*, 2011, p. 80120A. 64

[91] H. U. A. Chen and L. B. Wolff, "Polarization Phase-Based Method For Material Classification In Computer Vision," *Int. J. Comput. Vision*, vol. 28, no. 1, pp. 73–83, 1998. 64, 146

[92] L. B. Wolff, "Polarization-based material classification from specular reflection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 11, pp. 1059–1071, 1990. 64, 146

[93] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, 1990. 65

[94] B. Basener, E. J. Ientilucci, and D. W. Messinger, "Anomaly detection using topology," in *Proc. SPIE*, vol. 6565. SPIE, Apr. 2007, p. 65650J. 65

[95] M. G. Gartley and W. Basener, "Topological anomaly detection performance with multispectral polarimetric imagery," in *Proc. SPIE*, vol. 7334, 2009, p. 73341O. 66, 146

[96] B. D. Bartlett, A. Schlamm, C. Salvaggio, and D. W. Messinger, "Anomaly detection of man-made objects using spectropolarimetric imagery," in *Proc. SPIE*, vol. 8048, 2011, p. 80480B. 66, 146

[97] J. M. Romano, D. Rosario, and J. Mccarthy, "Day/Night Polarimetric Anomaly Detection Using SPICE Imagery," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–10, 2012. 66, 67, 146

[98] S. Tan, J. Stoker, and S. Greenlee, "Detection of foliage-obscured vehicle using a multiwavelength polarimetric lidar," *Proc. IEEE Int. Symp. Geoscience and Remote Sensing*, pp. 2503–2506, 2007. 66, 146

[99] A. Parsa, "Surface roughness observation of arctic ice using low grazing angle radar," in *Proc. IEEE Radar Conf.*, Apr. 2013, pp. 1–5. 66

[100] R. K. Dutta, "Theoretical Determination of Monostatic Co- & Cross- Linearly Polarized Radar Cross Section of a Missile at Each point of Missile Trajectory - Part I: Theory," in *Proc. Int. Conf. Recent Advances in Microwave Theory and Applications*, no. 2, Jaipur, 2008, pp. 614–617. 66

[101] A. Voronovich and V. Zavorotny, "Depolarization of microwave backscattering from a rough sea surface: modeling with small-slope approximation," in *Proc. IEEE Int. Geoscience and Remote Sensing*, 2011, pp. 2033–2036. 66

[102] C. S. Chun and F. A. Sadjadi, "Polarimetric imaging system for automatic target detection and recognition," Physics Innovations Inc., Tech. Rep. March, 2000. 68, 146

[103] F. Sadjadi and F. Sadjadi, "Passive Polarimetric Information Processing for Target Classification," in *Augmented Vision Perception in Infrared: Algorithms and Applied Systems.* Springer, 2008, ch. 2, pp. 37–61. 68, 146

[104] C. N. Dickson, A. M. Wallace, M. Kitchin, and B. Connor, "Improving infrared vehicle detection with polarisation," in *Proc. IET Conf. Intelligent Signal Processing*, 2013, pp. 1–6. 69, 147

[105] ——, "Polarimetric signatures for vehicle detection in the long wave infrared," in *Proc. Int. Symp. Optronics in Defence and Security*, 2014. 69, 147

[106] J. Almeida, L. Barbosa, A. Pais, and S. Formosinho, "Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering," *Chemometrics and Intelligent Laboratory Systems*, vol. 87, no. 2, pp. 208–217, Jun. 2007. 98

[107] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis." *J. Neurosci. Meth.*, vol. 134, no. 1, pp. 9–21, Mar. 2004. 101

[108] A. Steinberg and C. Bowman, "Revisions to the JDL Data Fusion Model," in *Handbook of Multisensor Data Fusion*, D. L. Hall and J. Llinas, Eds. CRC Press, 2001, ch. 2. 124

[109] N. Salamati and S. Süsstrunk, "Material-Based Object Segmentation Using Near-Infrared Information," in *Proc. IS&T/SID Color Imaging Conference*, 2010, pp. 196–201. 125

[110] N. Salamati, D. Larlus, and G. Csurka, "Combining Visible and Near-Infrared Cues for Image Categorisation," in *Proc. British Machine Vision Conference*, 2011. 125

[111] S. Süsstrunk, C. Fredembach, and D. Tamburrino, "Automatic skin enhancement with visible and near-infrared image fusion," in *Proc. Int. Conf. ACM Multimedia.* ACM, 2010, pp. 1693–1696. 125

[112] L. Schaul, C. Fredembach, and S. Süsstrunk, "Color image dehazing using the near-infrared," in *Proc. IEEE Int. Conf. Image Processing.* IEEE, 2009, pp. 1629–1632. 125

[113] E. D'Arca, N. Robertson, and J. Hopgood, "Person tracking via audio and video fusion," in *Proc. IET Data Fusion & Target Tracking Conference: Algorithms & Applications*, 2012. 125

[114] K. E. Ozden, K. Schindler, and L. Van Gool, "Multibody structure-from-motion in practice." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1134–1141, Jun. 2010. 129

[115] M. Farenzena, A. Fusiello, and R. Gherardi, "Structure-and-motion pipeline on a hierarchical cluster tree," in *Proc. IEEE Int. Conf. Computer Vision Workshops*. IEEE, Sep. 2009, pp. 1489–1496. 129

[116] P. Sturgess, K. Alahari, L. Ladicky, and P. Torr, "Combining appearance and structure from motion features for road scene understanding," *Proc. British Machine Vision Conference*, pp. 1–11, 2009. 129

[117] K. Schindler and D. Suter, "Two-view multibody structure-and-motion with outliers," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 676–683. 129

[118] R. Vidal, Y. Ma, S. Soatto, and S. Sastry, "Two-View Multibody Structure from Motion," *Int. J. Comput. Vision*, vol. 68, no. 1, pp. 7–25, Apr. 2006. 129

[119] O. Kahler and J. Denzler, "Tracking and Reconstruction in a Combined Optimization Approach." *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–14, Jul. 2011. 129

[120] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Euro. Conf. Computer Vision*, 2006, pp. 430–443. 129

[121] J.-Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm," Tech. Rep., 1999. 129

[122] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artificial Intelligence*, 1981, pp. 674–679. 129

[123] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981. 129

[124] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision, 2nd ed.* Cambridge University Press, 2004. 129

[125] C. N. Dickson, A. M. Wallace, M. Kitchin, and B. Connor, "Vehicle detection using multimodal imaging sensors from a moving platform," in *Proc. SPIE*

8541, Electro-Optical and Infrared Systems: Technology and Applications IX, vol. 44, Edinburgh, 2012, p. 854112. 148