UNIVERSITY
OF TECHNOLOGY
SYDNEY

Subsurface Material Estimation Using Hyperspectral Imaging

by Jasprabhjit Mehami

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of
Dr Alen Alempijevic
Prof Robert Fitch
Dr Raphael Falque

University of Technology Sydney
Faculty of Engineering and Information Technology

December 2022

Certificate of Original Authorship

I, Jasprabhjit Mehami, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signed: Production Note:
Signature removed prior to publication.

Date: 01/01/2023

Subsurface Material Estimation Using Hyperspectral Imaging

by

Jasprabhjit Mehami

A thesis submitted in partial fulfilment of the requirements for the
degree of Doctor of Philosophy

Abstract

Traditional vision is focused on the perception of visible light but limits us to only detecting surface features. In the area of food analysis, features such as composition, defects and contamination cannot be typically measured using visible light. Near-infrared light can penetrate into certain materials, so to capture this light, hyperspectral cameras can be employed. However, the images from these cameras lack spatial understanding. This work centres on a multi-modal vision approach, that leverages the capabilities of modalities such as colour, depth, and computed tomography (CT) in order to improve estimation of material properties, which can be used to detect features within objects.

To remove geometric discrepancies in images and estimate camera poses relative to some world coordinate system, geometric camera calibration is required. The process generally involves an optimisation that uses images of a known planar pattern in different poses. For a hyperspectral camera this is a challenging task as they generally capture images in a line-scan manner, where there is only a single spatial dimension, which makes it difficult to find common pixel features across different images. To aid with calibration, a multi-modal camera-system is deployed by combing the hyperspectral with a traditional frame camera, and an active calibration algorithm is devised which uses observability to selectively choose images that will improve the calibration. Experiments show lower uncertainty and error when the algorithm is compared to a naive approach utilising all images.

Illumination from external light sources is an important element of image formation. Hyperspectral cameras rely on sufficient illumination within their measured spectral bands. Real light sources have an asymmetric distribution in radiant intensity but are generally modelled as being symmetric, which affects photometric estimation techniques that rely on lighting information. Therefore, the second contribution of this thesis is spatially modelling the distribution of a real light source. In order to capture the asymmetry, a data-driven model using a Gaussian process (GP) with an unique mean function is proposed. The distributions of simulated and real light sources are modelled using the proposed GP model, where it shows less error when compared to similar models whilst successfully capturing the inherit asymmetries.

Cameras measure radiance which is dependent on factors such as illumination, shape and the respective material. Reflectance, an intrinsic property of a material, is independent of these factors. The reflectance can be estimated from the radiance by assuming the interaction of light for the materials can be described using the dichromatic reflectance model (DRM). The third contribution of this thesis involved improvements to an optimisation approach for reflectance estimation, by introducing terms that exploit information from prior light source modelling and surface shape. Evaluation on rendered and real images shows improvements in estimated reflectance, whilst also having less variation across areas that contained the same material.

In order to demonstrate the advantages of the three contributions of this thesis, a case study is investigated to estimate subcutaneous fat depth on lamb carcasses using hyperspectral imaging. Ground truth fat depth was determined using a computed tomography (CT) scanner, and the previous contributions were used to estimate the reflectance of the cuts. Various regression models were fit using different reflectance estimation methods, which included the proposed method from the third contribution. It was found that fat depth was best modelled using deep learning-based regression methods with the proposed reflectance estimation method.

Acknowledgements

I want to thank my three supervisors for giving me the opportunity and guiding me along my research journey. To my principal supervisor, Dr Alen Alempijevic, I would like to thank you for being patient with me, supporting me immensely during my research, giving me opportunities to develop in the academic environment, and organising and liaising with many of the practical components of my work. You would always manage to find the positives during adverse times and kept me pushing forward. Being one of your students has truly been a joy and an honour.

To my co-supervisor, Prof Robert Fitch, thank you for taking me on and giving me this opportunity to conduct postgraduate research. You believed more in my potential than I ever could, and I admire your view of focusing on the bigger picture. Your firm and meaningful words during our one-on-one discussions reminded me that I was working on something significant and that I should always remain curious to explore new topics and ideas.

To my second co-supervisor, Dr Raphael Falque, thank you for your vast technical and research support and for always being the one to provide concise, constructive feedback. I was impressed at your understanding of so many topics and your ability to explain them in simple terms, which changed my own approach to learning. I aim to possess the skill of asking clear and thought-provoking questions like you do.

A special thank you to A/Prof Teresa Vidal Calleja, who was not one of my official supervisors but, during many times of my degree, provided as much support and guidance as an actual supervisor. You offered numerous innovative ideas that facilitated the discovery of novelty in my work and showed genuine interest in my thoughts.

I want to thank Dr Liang Zhao and A/Prof Jaime Valls Miro, my chair and examiner, respectively, during my candidature assessments. You provided great constructive feedback that helped me better convey my ideas on paper and in presentation.

To the OTE group members, specifically Dr Lan Wu, Dr Behnam Maleki, and Mohammad Okour, I enjoyed doing my research and developing academically as a group. You were all eager, passionate and committed to your work, and you created a positive environment that made me push myself to not lag behind. Watching your research projects develop over the years was a great pleasure. A special thank you to one other OTE member, Dr Phillip Quin, who initially taught me everything about hyperspectral cameras and caught me up with the progress made before I arrived.

To Malcolm McPhee, the OTE group's connection to the industry, thank you for helping and organising data collection for my research in Armidale at the University of New England. I enjoyed your positive in-person interactions and admired the respect you gave to everyone you spoke to.

To the CDMRG group and each individual member, thank you for the many opportunities to present my research and topics of interest during the meetings. This was such a valuable component during my research that pushed me to improve my presentation skills. Thank you for the many occasions to do mock presentations before candidature assessments and providing quality feedback.

I'm extremely grateful to my family for their love and support. In particular, my parents, Paramjit Mehami and Parvinder Mehami, for pushing me to reach high feats in my studies, instilling the importance of education throughout my life and working tirelessly to give me such amazing opportunities. Their confidence in my skills vastly exceeded my own, and they would never miss the chance to let me know how proud they were of my accomplishments. I would not have the capability to complete my degree without them. To my brothers, Abheyjit Mehami and Nikhanjit Mehami, thank you for your admiration towards my work and for stepping up in the family while I completed my degree. I'm incredibly proud of you both.

I am also grateful to my partner, Janelle Bugauisan, for the love, support, and assistance throughout my degree. You boosted my confidence when times got tough and cheered me on during each milestone. Your presence made me feel at ease, and I would not have been able to push through the many hurdles without you. Thank you for always being there for me.

I want to acknowledge the various entities for their financial support during my degree, which includes the Australian Government as a part of their Research Training Program (RTP), the University of Technology Sydney for the Research Excellence scholarship and the Australian Government Department of Agriculture and Water Resources as part of its Rural R&D for Profit program under MLA Grant V.RDP.2005.

Finally, I express my gratitude to everyone at the Robotics Institute (formerly known as the Centre for Autonomous Systems), including students, staff, and faculty members, who have assisted me during my degree. While I haven't mentioned specific names, I value the time and assistance each of you provided. I am very fortunate to have worked alongside you all.

Contents

List of Figures

List of Tables

List of Acronyms

AUC	normalised area under the curve **xv, 131, 134, 136, 137**
BCS	body condition score **120**
BRDF	bidirectional reflectance distribution function **34–36, 38, 39, 76, 83, 98, 99**
cd	candela **79**
CNN	convolutional neural network **122, 132, 133, 139**
CRF	camera response function **84, 85**
CT	computed tomography **xiv, 7, 115, 117, 119, 120, 122, 127–130, 139, 144**
DEXA	dual-energy X-Ray absorptiometry **120**
DLT	direct linear transformation **56**
DRM	dichromatic reflectance model **6, 39, 96–101, 104, 114, 122–124, 126, 139, 144, 146**
EM	electromagnetic spectrum **xiii, 28**
FIM	Fisher information matrix **63**
GP	Gaussian process **6, 72, 77, 86–94, 98, 100, 111, 118, 123, 124, 126, 142–146**
GR	girth rib **120**

IMU	inertial measurement unit **44**
IR	infrared **25**
LED	light emitting diode **24, 25, 119**
MAE	mean absolute error **xiv, xv, 131, 136, 137**
MLE	maximum likelihood estimation **53, 58**
MLP	multilayer perceptron **132, 136–139**
MRI	magnetic resonance imaging **120**
NIR	near-infrared **2, 4, 118, 119, 121**
PCA	principal component analysis **121**
PCM	pinhole camera model **xiii, 9, 10, 12, 14–20, 46, 47, 49, 50, 52**
PLS	partial least squares regression **121**
PLS-DA	partial least squares discriminant analysis **121**
PnP	Perspective-n-Point **55, 60, 61**
ReLU	rectified linear unit **132**
RGB	red, green, and blue colour channels **1, 2, 23, 46, 78, 99, 113, 114, 119, 121–123, 125, 127, 132, 135, 142**
RGB-D	colour depth channels **122, 128, 142**
RID	radiant intensity distribution **xiv, 6, 74–76, 79–81, 85–87, 94, 142, 143**
RMS	root mean square **110, 112, 113**
RP	rough plastic **110, 112, 113**
SAM	spectral angle mapper **102, 111, 112, 114**
SE	squared exponential **77, 86**

SGD	stochastic gradient descent **136**
SI	International System of Units **32**
SLAM	simultaneous localisation and mapping **41, 45**
SSE	sum of squared error **111, 112**
STD	standard deviation **63, 64, 66–68, 104, 106, 108**
SVM	support vector machine **121**
UT	unscented transform **59–62**
UV	ultraviolet **25**

List of Symbols

Spaces, Geometry, Algebra

$\mathbb{R}$	The set of Real numbers
$\mathbb{P}^n$	The n-dimensional projective space
$\mathbb{E}^n$	The n-dimensional Euclidean space
$\boldsymbol{x}$	A vector
$\mathbf{X}$	A matrix
$\mathcal{A}$	A coordinate system
$\mathcal{W}$	The world coordinate system
P	A point in the world coordinate system
$\boldsymbol{P}$	The vector of P in the world coordinate system
${}^{\mathcal{F}}\boldsymbol{P}$	The vector $\boldsymbol{P}$ defined in the coordinate system $\mathcal{F}$
$\boldsymbol{x}$	A measurement vector, where the components of one or many Euclidean points are stacked together
$\boldsymbol{x}^\top, \mathbf{X}^\top$	The transpose of $\boldsymbol{x}$ and $\mathbf{X}$
$\mathbf{X}^{-1}$	The inverse of $\mathbf{X}$

${}^{\mathcal{F}}\mathbf{T}_{\mathcal{G}}$	The homogeneous transformation of $\mathcal{G}$ with respect to $\mathcal{F}$
$\mathbf{T}_{\mathcal{G}}$	The homogeneous transformation of $\mathcal{G}$ with respect to world coordinate system
$\mathbf{R}$	A rotation matrix
$\boldsymbol{t}$	A translation vector
r_x, r_y, r_z	Euler angles measured in radians
$f(x)$	A function defined in terms of x
$\mathbf{J}_f, \mathbf{J}_f(\boldsymbol{x})$	The Jacobian matrix of a function $f(\boldsymbol{x})$ with respect to $\boldsymbol{x}$

Vision

u, v	Image coordinates in pixels
x, y	2D normalised image coordinates.
X, Y, Z	3D Cartesian coordinates in the world coordinate system.
$\tilde{p}, \tilde{\boldsymbol{p}}$	Homogeneous coordinates of p and $\boldsymbol{p}$
$\tilde{w}$	The appended dimension to create a homogeneous coordinate
$\mathbf{E}$	The extrinsic matrix
$\mathbf{K}$	The intrinsic matrix
$\mathcal{I}$	The pixel image coordinate system
$\mathcal{F}$	The frame camera optical coordinate system
$\mathcal{S}$	The 2D Cartestian imaging sensor coordinate system
$\mathcal{H}$	The hyperspectral camera coordinate system
$\mathcal{P}$	The pattern coordinate system
F	Focal length of a pinhole camera measured in world units

u_0, v_0	Principal point of the image plane in pixels
α	Skew angle of a pixel
s_α	Skew parameter of the pinhole camera model
f_x, f_y	Focal lengths in the x and y directions of a pinhole camera image measured in pixels
$\mathbf{M}$	Camera projection matrix of a pinhole camera
K_n	Radial distortion parameter n of an optical lens
T_n	Tangential distortion parameter n of an optical lens
l_x, l_y	Length of pixels in the x and y directions measured in world units
$\boldsymbol{\vartheta}$	A vector of camera parameters
$\boldsymbol{u}$	Measurement vector of pixel features where the u and v are stacked together
f_{crf}	The camera response function

Statistics, Probability and Optimisation

μ_x	The mean of a random variable x
σ_x	The standard deviation (STD) of a random variable x
σ_x^2	The variance (VAR) of a random variable x
$\mathbf{\Sigma}_{\boldsymbol{x}}$	The covariance matrix of a random vector $\boldsymbol{x}$
$\mathcal{N}(\mu, \sigma^2)$	A normal or Gaussian distribution with mean μ and variance σ^2
$\widehat{x}$	The optimised parameter solution of x

Radiometry and Light

$\mathcal{L}$	The light source coordinate system

ν	The wavelength of light from the electromagnetic spectrum in world units
λ	The frequency of light from the electromagnetic spectrum in hertz
c	The speed of light in a vacuum
ω	A solid angle measured in steradians
$\hat{\boldsymbol{\omega}}$	The solid angle direction vector of a light ray
Q	The radiant energy of given radiation measured in joules
Φ	The radiant flux of given radiation measured in watts
E	The irradiance of given radiation measured in watts per square meter
s	The radiant intensity of given radiation measured in watts per steradian
L	The radiance of given radiation measured in watts per square meter per steradian
f_{brdf}	A BRDF function of a surface for material
$\hat{\boldsymbol{n}}$	The direction vector surface normal
(r, θ, φ)	The 3D spherical coordinate system defined using the euclidean distance and the zenith and azimuth angles
δ	The Dirac delta function
ρ	The diffuse reflectance of a surface
ρ_k	The specular reflectance of a surface
k	The specular coefficient of a surface
g	The shading factor of a surface due to lighting and surfae shape

Chapter 1

Introduction

Vision sensors play a crucial role in understanding the physical world, due to their ability to gather rich information from incoming light in a way that captures its spatial dimensions. Vision itself is a powerful sense that gives the ability to distinguish objects based on characteristics such as colour, size, and distance. Historically, camera sensors have been designed to mimic the behaviour of human vision. Our eyes have always been the vision sensors of our body that collect light from the environment on photoreceptive regions called the retina. If we ignore the underlying chemistry that occurs in the retina, the image formation process can be simplified to that of converting collected light energy into electrical signals that can be processed by the brain. The critical component to perceiving colour is that there are different types of retina cells that each have distinct responses to the wavelengths in the visible spectrum that closely match the colours red, green, and blue colour channels (RGB).

Visual information from the environment is carried by the medium of light (or radiation) that is encoded in its continuous wavelengths in the form of radiant energy (or luminous energy when working in the visible spectrum). We observe the resulting light after it has interacted with materials in the world. The type of interaction and its location depends on numerous factors such as the intrinsic reflective properties of the material, the wavelength of the light, and the direction of the incident light with respect to the material's surface. The interactions of light can occur at the surface for highly reflective mirror-like materials, and within the material. There is energy balance that must be maintained with the incoming, outgoing and absorbed radiant energy. Therefore, if a material is not self-illuminating, light will be reflected in a given direction or directions with total outgoing radiant energy that must be less than or equal to the total incoming radiant energy, and the difference is the energy that is absorbed by the material.

When we place a vision sensor, such as a camera, in the world, it receives all the light from the environment that is directed towards its lens. The lens then focuses the light onto the photosensitive image sensor. For traditional 2D frame colour cameras, the image sensor operates in the visible spectrum between the wavelengths of 400nm and 700nm in three isolated bands that represent the RGB colours. The spectra of the measured light is continuous, and even in the visible domain, our traditional cameras ignore much of the information to form discrete images of the world. Spectroscopy is a field that is interested in capturing the entire spectra produced by light interactions, and tools within this field are used in applications relating to astronomy and chemical analysis. They operate through the effect of dispersion using optical equipment such as prisms. Analysing the spectra allows for identification of chemicals, atomic bonds, and individual elements. The main drawback of spectroscopy techniques is that they only capture point measurements. This makes it difficult to use such equipment for photometry purposes. If we were to combine the spatial capabilities of the image sensor from the traditional frame camera and light dispersing prisms used in spectroscopy, you would get the vision sensor known as a hyperspectral camera.

Hyperspectral cameras (also referred to as imaging spectroscopy) capture imagery data using a number of contiguous bands typically in the near-infrared (NIR) wavelengths of the electromagnetic spectrum. They capture data in a non-evasive and remote manner, making them ideal for space-borne applications, aerial mapping, agriculture monitoring and food analysis. Depending on the observed material and the measured wavelengths of light by the camera, interactions may occur at sub-surface levels revealing internal features that can be exploited to determine some desired attribute. An example of such a purpose is in the meat industry, where knowing the fat depth of a carcass in specific regions is used to estimate the overall fat composition in order to grade the meat. Traditional methods for acquiring overall carcass compositions are either carried out before the animal is slaughtered by an expert assessor or by probing the carcass with a physical device. These methods are generally slow, subjective, and could introduce contaminants due to physical handling. Hyperspectral cameras could be used as tool to capture data in a controlled manner, and once processed would produce objective grading results.

1.1 Motivation

The imagery data that is captured by hyperspectral cameras contains a large number of dimensions. In order to reliably record data at a high frame-rate, hyperspectral cameras typically operate in a line-scan manner with a single spatial dimension. Given that a hyperspectral behaves like a traditional colour frame camera, spatial camera calibration is an essential step before using the hyperspectral for any application.

The illumination for hyperspectral imagery comes from a source that produces illumination in the measured spectral bands of the camera. The intensity and direction of the incident illumination on an object will depend of the location, geometry, and distribution properties of the source. Hyperspectral cameras are generally operated in outdoor applications that assume constant spatial illumination and no variation in ray directions, by assuming the light source is far away such as the sun. In near-field experiments, the distance between the light source and surface is small, and spatial variations in both intensity and direction are present. These variations will effect the incident light energy arriving at a surface point.

Capturing data with hyperspectral cameras requires understanding the underlying physics involved with light propagation and its interaction with the environment. The hyperspectral camera measure the light that has reflected off a material which is referred to as radiance. This property is dependent on many factors such as the viewing direction, incident illumination, surface shape, and material properties. If there are any variations across these factors, when measuring the same material, there will be differences in the captured hyperspectral radiance image. It is more appropriate to estimate an intrinsic material property that is invariant to these variables, such as the diffuse reflectance.

The hyperspectral imaging could be used to measure some desired attribute (or feature) of a material that varies spatially across an object of that material. Examples include the thickness of material layers, locating defects, and identifying the composition of a material. Machine learning models could be used to build appropriate models that relate the desired attribute to the estimated reflectance. Hyperspectral cameras have numerous spectral bands, many of which will not show significant variations, so it is essential to reduce dimensionality. Once built, these models will be used to locally infer the desired attribute of an object given reflectance values at that point.

1.2 Scope

The hyperspectral cameras capture images by collecting the light (or radiation) that is reflected or emitted by surfaces, which is in a specific wavelength range (typically NIR) of the electromagnetic spectrum. Therefore, only interactions of light with materials are of interest which are measured by the hyperspectral camera of interest. The cameras have a complex arrangement of lenses, prisms and filters to capture the wavelengths in a spatial manner similar to that of colour frame cameras. Therefore, hyperspectral cameras are assumed to behaviour like regular colour frame cameras, especially in regards to the image formation process which follows central perspective projection.

In this work, light behaviour follows the properties described by ray optics, where light is treated as infinitesimally thin rays travelling in the direction of energy propagation at the speed of light c. The incident light illumination can either be reflected, absorbed or transmitted through a material. It is assumed that materials which are measured by the hyperspectral camera are non-luminous and show no fluorescent effects. Therefore, the hyperspectral camera only measures reflected light from a material. The light passing through the object is assumed to be insignificant due to the thickness of the materials. Incident illumination on a surface can come from multiple sources such as light rays reflecting off another surface, which is also referred to as indirect lighting. It is assumed that the incident illumination on the surfaces being captured only come from known light sources (or direct light).

This thesis focuses on hyperspectral imaging in a near-field setup, where the light source is located close to the object being captured. The absorbed light may cause heating effects in materials that may alter the way in which light frequencies behave, but this behaviour is assumed to be negligible. While imaging with the hyperspectral camera, materials are considered not to deform and all their properties will remain in a steady-state. In this research, non-metallic and solid heterogeneous materials will be considered. This encompasses many groups of materials such as:

- Consumable goods: fruit, vegetables, meats, nuts
- Organic matter: humans, animals, plants
- Woods: trees, timber, paper
- Plastics: Polystyrene, Polyethylene
- Composites: concrete, plywood, fibreglass

For these materials, determining locally varying attributes relating to composition and quality may be required. These attributes generally come from assessments of the internal structure. In many cases destructive and evasive techniques, such as chemical analysis, can be harmful, impractical and slow. In this thesis, we will focus on consumable animal products. The desired attribute is the thickness of subcutaneous fat layers, or fat depth, but the overarching approach using hyperspectral imaging should be applicable to any other attributes.

1.3 Contributions

There are four main contributions of this thesis:

1. **A novel probabilistic calibration algorithm for line-scan hyperspectral cameras**

   The first contribution of this thesis describes a probabilistic calibration of a line-scan hyperspectral and frame camera-system using a newly developed active calibration algorithm, that uses a metric of observability to selectively choose measurements that will improve estimation of calibration parameters. For the line-scan camera, uncertainty propagation is a crucial part of this work as noisy pixel measurements are used to determine pattern feature points in world coordinates. Measurements of similar poses with insufficient parallax are removed from the optimisation to avoid degrading the calibration parameters. The proposed algorithm is introduced in Chapter 3 and is shown to better estimate calibration parameters with less uncertainty compared to a naive approach that uses all images. The calibrated camera system is used in all later contributions.

2. **A probabilistic data-driven light source model to capture asymmetric radiant distribution**

   The second contribution focuses on modelling the irradiance of a light source that is used to illuminate the objects for hyperspectral imaging in a near-field setup. The light source is assumed to behave as a non-isotropic disk-source with an asymmetric radiant intensity distribution (RID). In order to capture the asymmetry, a data-driven model is devised using Gaussian process (GP) with a non-zero mean function that captures properties such as the attenuation. The light source model is built for a single spectral band using discrete measurements of a known planar Lambertian surface. The irradiance distribution of the GP model is shown to be more accurate than a symmetrical parametric model, by capturing the asymmetries that exist in the light source distribution. The proposed light source model is defined in Chapter 4.

3. **A reflectance estimation formulation that incorporates known shape and lighting information**

   In the third contribution, the prior modelling of the light source and known surface shape information is used to improve the estimation in the material property of diffuse reflectance. Light interactions of the material are assumed to behaviour according to the dichromatic reflectance model (DRM). The incident irradiance on the surface of a material is estimated from the GP light source model. An existing quadratic optimisation is reformulated to include terms for the surface normal and light source direction vectors, which aid in constraining the estimated reflectance.

The proposed method is compared with existing methods using a spectral renderer as discussed in Chapter 5, and is shown to best estimate the reflectance when compared to ground truth reflectance.

4. **A case study to estimate subcutaneous fat depth of lamb cuts using hyperspectral imaging**

   A case study is presented, aimed at building a regression model of subcutaneous fat depth on the surface of lamb leg cuts using hyperspectral imaging. This incorporates all the previous contributions by estimating the reflectance of the lamb cuts from their radiance measurements. Ground truth fat depth is acquired from computed tomography (CT) scans by ray casting 3D reconstructed CT models. Three different reflectance methods and four different regression models are trained using one lamb cut and tested with a different cut. The results in Chapter 6 show that the best fits come from using the proposed reflectance estimation from Chapter 5, and a deep learning-based regression model.

1.4 Publications

Directly Related

J. Mehami, T. Vidal-Calleja, and A. Alempijevic. Observability driven Multi-modal Line-scan Camera Calibration. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2020-Septe:285–290, 2020. doi: 10.1109/MFI49285.2020.92352
Contributions covered in Chapter 3

J. Mehami, R. Falque, T. Vidal-Calleja, and A. Alempijevic. Multi-modal Non-Isotropic Light Source Modelling for Reflectance Estimation in Hyperspectral Imaging. *IEEE Robotics and Automation Letters*, 2022. ISSN 2377-3766.
Contributions covered in Chapter 4 and Chapter 5

F. Peña, J. Mehami, R. Guenot-Falque, T. Patten, A. Alempijevic, and T. Vidal Calleja. Subcutaneous fat depth regression using hyperspectral and depth imaging. In *Australasian Conference on Robotics and Automation*. ARAA, 2022.
Contributions covered in Chapter 6.

Indirect

M. Usayiwevu, C. Le Gentil, J. Mehami, C. Yoo, R. Fitch, and T. Vidal-Calleja. Information driven self-calibration for lidar-inertial systems. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9961–9967. IEEE, 2020

1.5 Thesis Outline

The thesis is structured into six chapters, where chapters 3 to 6 contain the content relating to the contributions. The next chapter, Chapter 2, is an overview of background topics which are common amongst all works and help to introduce foundational concepts. In Chapter 3 the calibration of the line-scan camera-system with the active algorithm is discussed. Then Chapter 4, contains the probabilistic light source modelling using the calibrated camera-system. Chapter 5 covers reflectance estimation, where an existing optimisation method is built upon to include shape and light information. Chapter 6 examines a case study of estimating subcutaneous fat depth on lamb leg cuts using the proposed reflectance estimation method from Chapter 5. Finally, Chapter 7 summaries the contributions and discusses their significance to the thesis, and ends with future research problems.

Chapter 2

Background

2.1 Perspective Camera

Cameras are essentially a device that capture light from a physical 3D world and map the light information onto a 2D image plane through a projective transformation. If the projection is one of perspective, certain geometric properties are preserved where the cameras are classed as being *perspective cameras* and the resulting images are seen to exist in the *projective space*. An example of such an image is shown in Figure 2.1 where a perspective camera in the Gazebo simulator [5] captures a scene of cubes and spheres. The 3D view of the scene shows that size and distance are not preserved in the image. There is a surprising relationship between the projective space and the more familiar Euclidean space which needs to be discussed as it will lead us towards the necessary introduction of *homogeneous coordinates*.

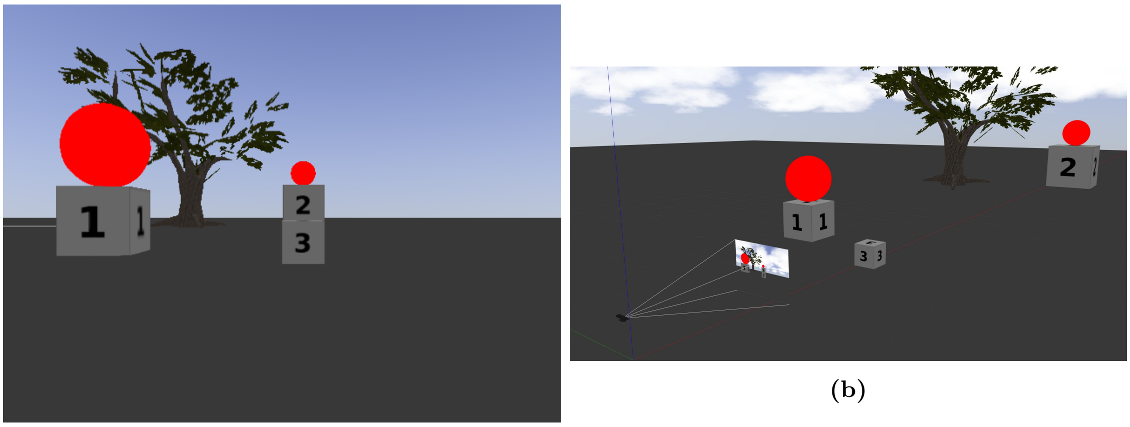
The entire process of forming an image from a perspective camera is creatively called *image formation*. This is a complex process in the physical camera that involves elements of photometry, optics, and electronics. Image formation can be simplified to just consider the geometric effects in the image due to the projective transformation where ideal camera models, such as the pinhole camera model (PCM), mathematically describe the transformation that is applied to the 3D world. The definition of the PCM is not realistic, its practical flaw is the assumption of acquiring sufficient light which is the reason why physical cameras all require a lens. But sometimes fixing one issue introduces another, with lens distortion as a side effect in images. Fortunately, appropriate camera models beyond the PCM can be used to counter these effects.

With a focus on robotics and computer vision, a perspective camera is thought of as a visual sensor that provides rich spatial information of the physical 3D world, with some inherit limitations due to its hardware. Similar to other sensors, images from a perspective camera are considered to be noisy measurements which will affect algorithms and decision-making tasks that use those measurements. There are also many practical elements to cameras which are dependent on the scene being captured such as the choice of lens, the size of the imaging sensor, and the amount of time the sensor is exposed (known as the exposure time). There is a dedicated field of photography that study how these elements and features can be optimised (or even exploited) to produce appealing images that have artistic or philosophical value. A good introduction to photography can be found in the guide by Carroll [6]. This thesis will not focus on any photography topics, practical camera features (e.g. aperture, flash, ISO, white balance), optical components, and imaging sensors and their corresponding electronics. This section will give an overview of fundamental topics that are discussed in depth by Hartley and Zisserman [7] but will use updated notation and illustrations similar to that of Barfoot [8] and Sola [9].

2.1.1 Projective Space

The projective space $\mathbb{P}^n$ is where images exist that are captured by a perspective camera. To understand its significance we need to compare it to the Euclidean space $\mathbb{E}^n$ that typically represents the physical spatial dimensions of reality. Euclidean space is homogeneous and its transformation (which is also referred to as a rigid transformation) preserves straightness, shape, angles, lengths, and ratio of lengths. Whereas, the projective transformation only preserves **straightness**, which turns out to be the only property required to geometrically comprehend a scene from an image. The effects of both transformations to 2D space can be seen in Figure 2.2.

Euclidean space is not perfect as it requires exceptions to intersections of N-D geometry in N-D Euclidean space (e.g. 2D lines in 2D space, 3D planes in 3D space, etc). The exception



(a)

(b)

Figure 2.1: A scene captured by a perspective camera in the Gazebo simulator [5] highlighting the geometric effects of the projective transformation. (a) image from the camera of three numbered cubes of different size and two red spheres of equal size. Cube 1 appears to be largest and cubes 2 and 3 appear to be the same size and stacked on top of each other. (b) is a 3D view of the scene showing the camera and its virtual image plane, the actual size of the cubes and there distances from the camera. It can be clearly seen that size and distance are not preserved in the projective space.

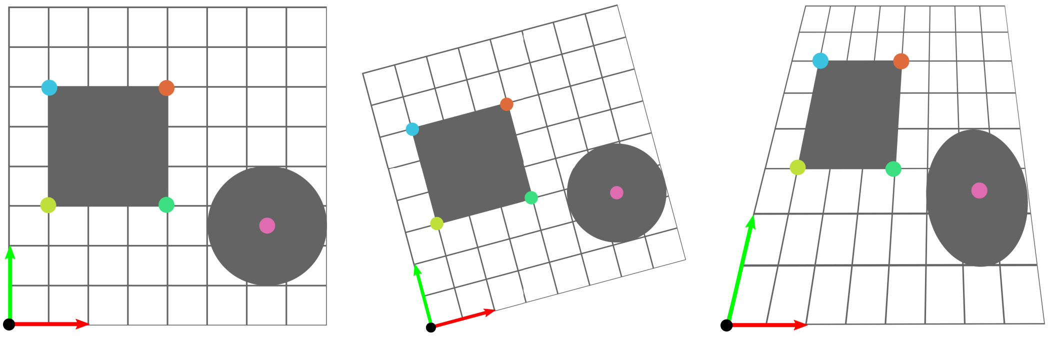


Figure 2.2: An Euclidean and a perspective projective transformation being applied to a 2D space. On the left is the original Euclidean space, the middle is the resulting Euclidean space, and the right is the projective space. Notice how the Euclidean transformation preserves all angles and lengths of the grey square, and the circle remains as a circle. In the projective transformation both shapes have been altered.

occurs when you have two parallel geometry which are said to ”meet at infinity”. The notion of infinity in Euclidean space is for mathematical completion but it does not exist in the space. The role of projective space is to extend Euclidean space by including points at infinity. Therefore, intersections of all N-D geometry in N-D projective space are possible at finite coordinates (e.g. 2D parallel lines in 2D projective space meet at a finite point). The coordinates in projective space are known as homogeneous coordinates.

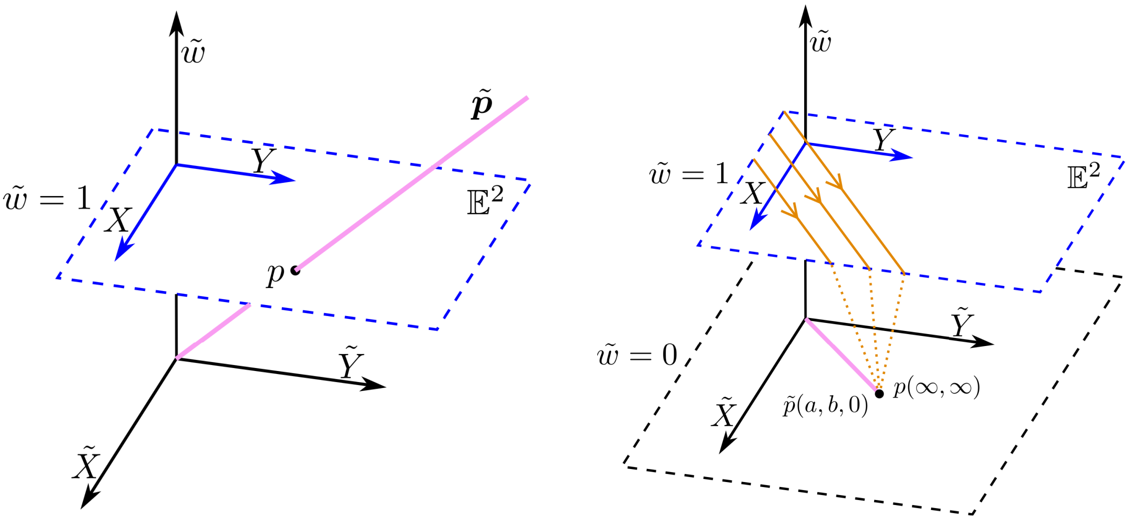
2.1.2 Homogeneous Coordinates

Similar to how projective space extends Euclidean space, homogeneous coordinates extend Euclidean coordinates to able to uniquely define a point in projective space. In Euclidean space a point is uniquely represented by a Cartesian coordinate, which itself is an ordered list of numbers from $\mathbb{R}$. These coordinates can be defined in 2D, 3D and higher N-D dimensions as $P(X, Y)$, $P_1(X, Y, Z)$, and $P_2(X, Y, Z, ...)$ respectively. An N-D Cartesian coordinate can be converted to a (N+1)-D homogeneous coordinate by first scaling the coordinate by $\tilde{w} \in \mathbb{R}$, and then appending $\tilde{w}$ to the coordinate. Therefore, homogeneous coordinates for P, P_1, and P_2 are defined as $\tilde{P}(\tilde{w}X, \tilde{w}Y, \tilde{w})$, $\tilde{P}_1(\tilde{w}X, \tilde{w}Y, \tilde{w}Z, \tilde{w})$, and $\tilde{P}_2(\tilde{w}X, \tilde{w}Y, \tilde{w}Z, ..., \tilde{w})$ respectively. The conversion back to Cartesian coordinates is the opposite of this operation.

Let us focus on the single 2D Cartesian coordinate P. There exists a family of 3D homogeneous coordinates $\tilde{P}$ that lie on a 3D line in projective space. By setting $\tilde{w} = 1$, the 2D Euclidean plane $\mathbb{E}^2$ can be embedded in the 3D projective space $\mathbb{P}^3$ as illustrated in Figure 2.3, and the conversion to $\tilde{P}$ is simplified to just appending and truncating a 1. Any value for $\tilde{w}$ will still represent the same P, so if we only look to change $\tilde{w}$, as it approaches 0 P that are further away from the origin can be represented. Therefore, by setting $\tilde{w} = 0$, a P at infinity can be represented by finite homogeneous coordinates as shown in Figure 2.3. These same ideas also apply to higher dimensions.

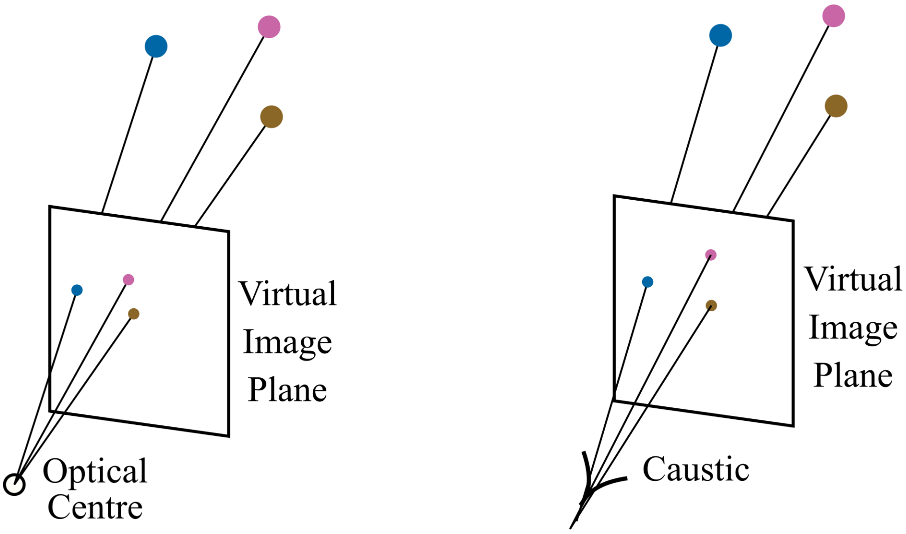
2.1.3 Camera Models

It is necessary to have well established mathematical models that can accurately describe the projective transformation of perspective cameras. All cameras in this work are assumed to follow *central projection* where all light rays will pass through a single point called the *optical centre*. The difference between a central and non-central camera is shown in Figure 2.4. I start with defining the most basic camera known as the PCM and build upon its foundations to be able to replicate the behaviour of real cameras.



(a) **(b)**

Figure 2.3: In (a) The 2D Euclidean plane $\mathbb{E}^2$ embedded in the 3D projective space $\mathbb{P}^3$ at $\tilde{w} = 1$. The Cartesian point P exists as a 3D line $\tilde{\boldsymbol{P}}$ (shown in purple) in projective space. In (b) a family of parallel lines in $\mathbb{E}^2$ (shown in orange) intersect at $P(\infty, \infty)$ which is located at infinity. This same coordinate can be represented in finite homogeneous coordinates $P(a, b, 0)$ which exists on the plane at $\tilde{w} = 0$ in $\mathbb{P}^3$.



(a) Central projection **(b)** Non-central projection

Figure 2.4: Central projection (a) vs non-central projection (b) of a perspective projection camera. The cameras both have a virtual image plane that shows the respective image. The central camera has an optical centre where all rays meet. The non-central camera has a caustic [10] caused by adding an optical lens.

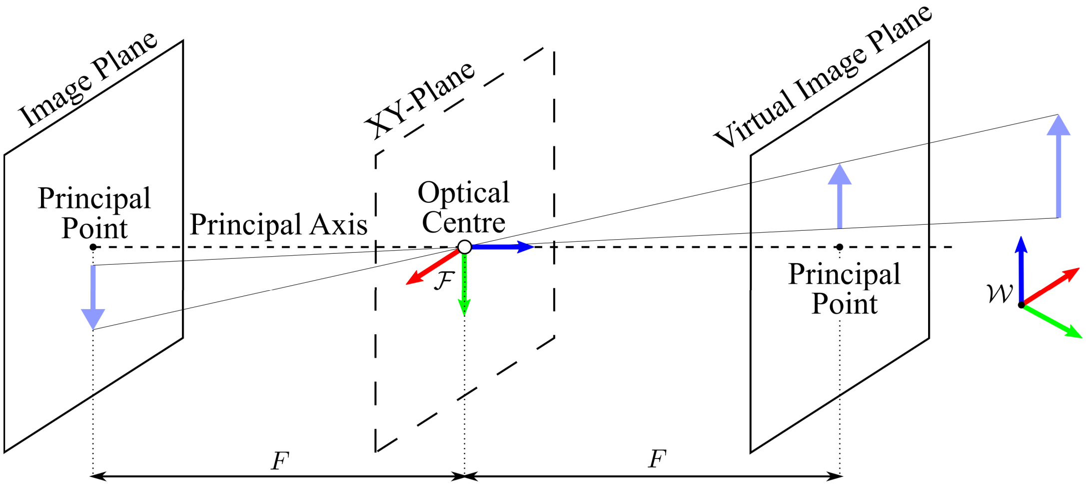


Figure 2.5: The pinhole camera model (PCM) showing the 3D coordinate system $\mathcal{F}$ at the optical centre where the infinitely small aperture exists. The image plane is parallel and behind the XY-plane of $\mathcal{F}$ at the focal length F. An object defined in $\mathcal{W}$ is projected onto the image plane inverted. There exists a virtual image plane in front of the XY-plane which has the same features as the image plane, but the projected image is not inverted.

2.1.3.1 Pinhole Camera

The PCM is shown in Figure 2.5. It consists of a 3D optical coordinate system $\mathcal{F}$ where the Z-direction points in the viewing direction of the camera and is referred to as the *principal axis* (or *optical axis*). At the origin exists an infinitely small aperture that allows light rays to all pass through a single point called the *optical centre*. The light rays hit a plane that is parallel and behind the XY-plane known as the *image plane* which is at a distance F called the *focal length* measured in world units (e.g. millimetres). The principal axis intersects the image plane at the *principal point*. An object defined $\mathcal{F}$ is projected to the image plane through the optical centre but appears inverted. There also exists a virtual image plane, which is a distance F in front of the XY-plane of $\mathcal{F}$, and has the exact same features as the image plane but the projected object is not inverted.

It is common to see a simplified version of the PCM as shown in Figure 2.6 which only contains the positive-Z quadrants of $\mathcal{F}$. In this form, the image plane is ignored and we only focus on the virtual image plane which has the image coordinate system $\mathcal{I}$ at its top-left corner and a 2D Cartesian coordinate system $\mathcal{S}$ attached to the origin of the imaging sensor[1]. A 3D point ${}^{\mathcal{F}}\boldsymbol{P}[X, Y, Z]$ which is defined in $\mathcal{F}$ describes a line which intersects

---

[1] The physical imaging sensor would be located at the real imaging plane but it would have the same 2D Cartesian coordinate system $\mathcal{S}$.

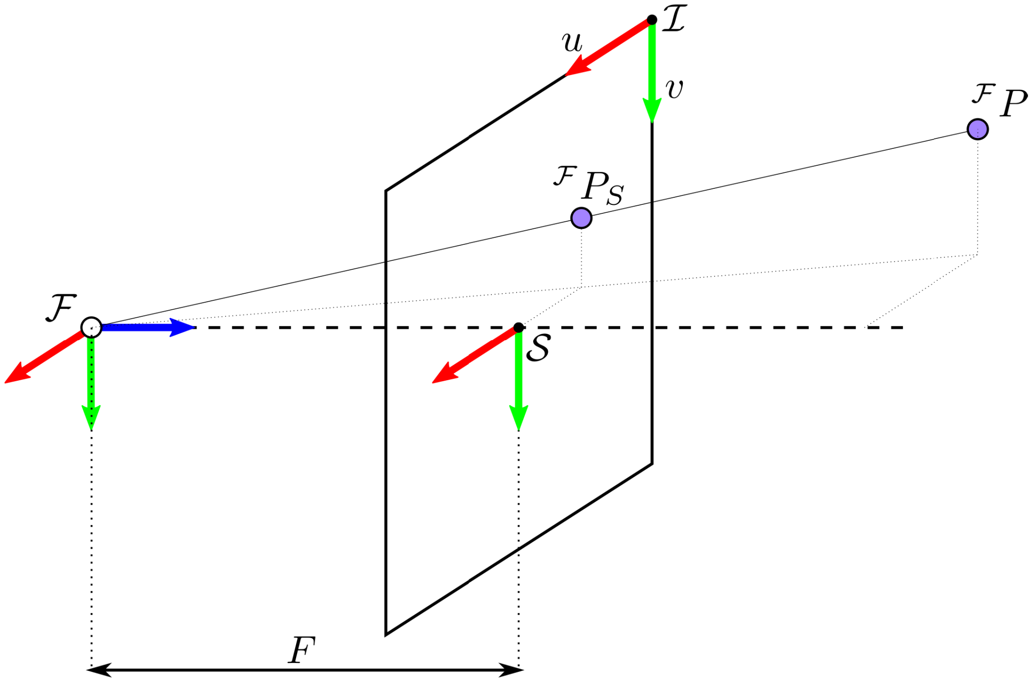


Figure 2.6: A simplified version of the pinhole camera model (PCM) which only shows the positive-Z quadrants of the 3D optical coordinate system $\mathcal{F}$. Only the virtual image plane is shown at the focal length F with the image coordinate system $\mathcal{I}$ situated at the top-left corner and the imaging sensor coordinate system $\mathcal{S}$ at the principal point. A 3D point ${}^{\mathcal{F}}\boldsymbol{P}[X, Y, Z]$ is projected onto the virtual image plane at the 3D point ${}^{\mathcal{F}}\boldsymbol{P}_S[X_S, Y_S, F]$ which are both related by a similar triangle relationship on the XZ and YZ planes.

the virtual image plane at the 3D point ${}^{\mathcal{F}}\boldsymbol{P}_S[X_S, Y_S, F]$. ${}^{\mathcal{F}}\boldsymbol{P}$ and ${}^{\mathcal{F}}\boldsymbol{P}_S$ are related through similar triangle relationships on the XZ and YZ planes. If we apply the similar triangle relationship on the XZ plane we can conclude that $X_S/F = X/Z$. Following the same logic for the YZ plane, an initial form of the PCM is shown as follows:

$$\begin{bmatrix} X_S \\ Y_S \end{bmatrix}_{\mathcal{F}} = \frac{F}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}_{\mathcal{F}} . \tag{2.1}$$

This initial projection equation determines the unknown components of ${}^{\mathcal{F}}\boldsymbol{P}_S$ but note that X_S and Y_S are still measured in world units and no conversion to pixels has been done yet. If we take Equation (2.1) and convert to homogeneous coordinates it will look as follows:

$$\begin{bmatrix} X_S \\ Y_S \\ 1 \end{bmatrix}_{\mathcal{F}} = \begin{bmatrix} F & 0 & 0 & 0 \\ 0 & F & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{\mathcal{F}} . \tag{2.2}$$

In this form the homogeneous coordinate of ${}^{\mathcal{F}}\boldsymbol{P}_S$ is on the left-hand side and the homogeneous coordinate of ${}^{\mathcal{F}}\boldsymbol{P}$ is on the right-hand side. The PCM works with 3D points in the world coordinate system $\mathcal{W}$ and so far the points have been defined in the camera's coordinate system $\mathcal{F}$. A rigid transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{W}} \in \mathbb{R}^{4\times4}$ is required that transforms a 3D point from $\mathcal{W}$ to $\mathcal{F}$. This transformation is known as the *extrinsic matrix* specifically denoted as $\mathbf{E} \in \mathbb{R}^{4\times4}$ and it is comprised of a 3D rotation matrix from the SO_3 group $\mathbf{R} \in \mathbb{R}^{3\times3}$ and a translation vector $\boldsymbol{t} \in \mathbb{R}^3$ shown as follows:

$$\mathbf{E} = \begin{bmatrix} \mathbf{R} & \boldsymbol{t} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} , \tag{2.3}$$

where $(r_{11}, r_{12}, r_{13}, ...)$ are elements of $\mathbf{R}$ and (t_1, t_2, t_3) are elements of $\boldsymbol{t}$. Using Equation (2.2), a point $\boldsymbol{P}[X, Y, Z]$ defined in $\mathcal{W}$ is projected as follows:

$$\begin{bmatrix} X_S \\ Y_S \\ 1 \end{bmatrix}_{\mathcal{F}} = \begin{bmatrix} F & 0 & 0 & 0 \\ 0 & F & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{E} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} . \tag{2.4}$$

The final element of the PCM is the conversion from world units to pixels in the resulting image. If we focus on the virtual image plane, as shown in Figure 2.7, a grid of pixels can be seen with the $\mathcal{I}$ located at $(0, 0)$ which is the the top-left corner and $\mathcal{S}$ is located at the pixel (u_0, v_0) which is the principal point. The pixels are assumed to be rectangular and evenly spaced. A single pixel is denoted as p and its dimensions in the x and y directions

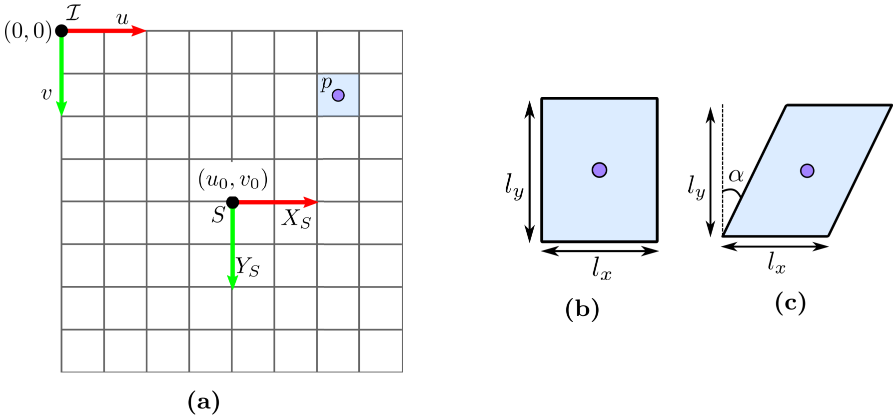


Figure 2.7: In (a) the virtual image plane of a camera with its rectangular pixels distributed as a grid. In the top-left corner is the image coordinate system $\mathcal{A}$. In the middle is the principal point located at pixel (u_0, v_0) which also houses the 2D Cartesian imaging sensor coordinate system $\mathcal{S}$. A rectangular pixel p is shown in (b) with its dimensions l_u and l_v measured in world units. In (c) a skewed pixel can be seen where its skew angle is α.

are l_x and l_y. Figure 2.7 also shows a skewed pixel with a skew angle α. The physical explanation for this phenomena is that the x and y axes of the pixels on the image sensor are not perpendicular [2]. This will shift the expected u component of the pixel by an offset $\tan(\alpha)$. The principal point, pixel dimensions and skew are used to convert the 2D Cartesian point $[X_S, Y_S]$ to a pixel (u, v) in $\mathcal{I}$ as follows:

$$\begin{aligned} u &= \frac{1}{l_x} X_S + \frac{1}{l_x} \tan(\alpha) Y_S + u_0 \\ v &= \frac{1}{l_y} Y_S + v_0 \,. \end{aligned} \tag{2.5}$$

If $[X_S, Y_S]$ and (u, v) are converted to homogeneous coordinates, then Equation (2.5) can be written as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/l_x & \tan(\alpha)/l_x & u_0 \\ 0 & 1/l_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_S \\ Y_S \\ 1 \end{bmatrix} . \tag{2.6}$$

The PCM in Equation (2.4) can then be updated to output pixels by substituting into Equation (2.6) as follows:

---

[2] Skew in pixels is rare and in most cases $\alpha = 0$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/l_x & \tan(\alpha)/l_x & u_0 \\ 0 & 1/l_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} F & 0 & 0 & 0 \\ 0 & F & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{E} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} . \tag{2.7}$$

The form can be simplified by performing the matrix multiplication as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s_\alpha & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{E} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} , \tag{2.8}$$

Where $f_x = F/l_x$ and $f_y = F/l_y$ are the focal lengths in the x and y direction respectively measured in pixels, and $s_\alpha = f_x \tan\alpha$ is the skew parameter measured in world units. If the 3×4 matrix in Equation (2.8) is extracted and only the non-zero columns are considered, we are left with the *intrinsic matrix* $\mathbf{K} \in \mathbb{R}^{3\times3}$ as follows:

$$\mathbf{K} = \begin{bmatrix} f_x & s_\alpha & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.9}$$

The intrinsic matrix fully captures the internal parameters of the PCM. The complete PCM can then defined as:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{M} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} , \tag{2.10}$$

Where $\mathbf{M} \in \mathbb{R}^{3\times4}$ is the camera projection matrix which is defined as $\mathbf{M} = \begin{bmatrix} \mathbf{K} & \mathbf{0} \end{bmatrix} \mathbf{E}$, and is valid to an unknown arbitrary scale factor s which accounts for the fact that a family of 3D Cartesian points can be projected to the same pixel.

2.1.3.2 Thin Lens Approximation

The main issue with the PCM is assumption of the aperture size. An infinitely small aperture would capture no light to produce an image. Even if the aperture was given a small finite size, this would require a long exposure time to let in enough light to produce an adequate image. Fortunately, converging lenses from optics can be fitted to an enlarged aperture of a camera to allow more light rays to enter whilst still obeying central projection.

The thickness of the lens will effect how significantly the light rays are bend in order to reach the image plane. When the thickness of the lens is much smaller than the radii of curvature on the lens surfaces, the thickness is said to be negligible and the thin lens approximation can be applied as shown in Figure 2.8. Multiple light rays from the 3D point ${}^{\mathcal{F}}\boldsymbol{P}$ are shown to arrive at a pixel p on the image plane. In order for ${}^{\mathcal{F}}\boldsymbol{P}$ to appear sharp in the image, all light rays must be bend towards p, which is determined by the following relationship:

$$\frac{1}{F_l} = \frac{1}{F} + \frac{1}{d}, \tag{2.11}$$

Where F_l is the focal length of the lens, F is the focal length of the PCM, and d is the Euclidean distance of ${}^{\mathcal{F}}\boldsymbol{P}$. All distances are in world units and the lens does not cause any form of distortion to the image.

2.1.3.3 Real Lens

Real lens do have significant thickness which causes geometric effects in the pixel location observed as distortion. This is caused by violating central projection in the PCM as all rays do not enter via the optical centre. The further from the optical centre the greater the effects of the distortion. This is commonly referred to as *radial distortion* which can cause three types of aberrations in images known as barrel, pincushion, and mustache. There is a second type of distortion known as tangential and this occurs when the lens and the camera sensor are not parallel. These distortions are all summarised in Figure 2.9.

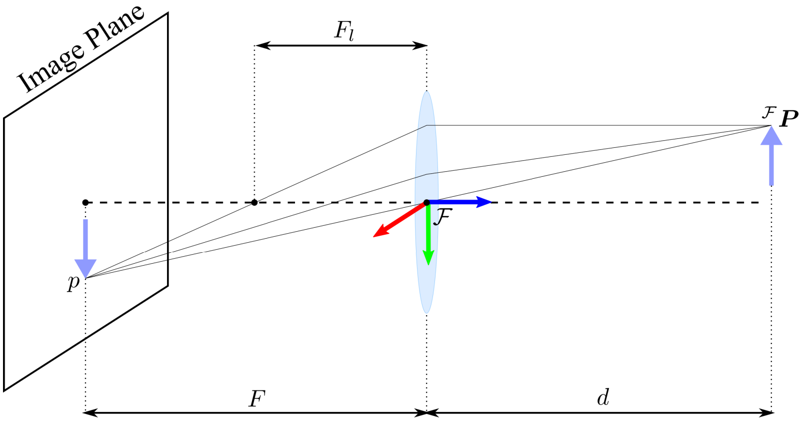


Figure 2.8: A thin lens combined with a pinhole camera model (PCM) showing its image plane and 3D optical coordinate system $\mathcal{F}$. A 3D point ${}^{\mathcal{F}}\boldsymbol{P}$ on an object is seen to be projected onto the image plane at pixel p. Multiple light rays originating from ${}^{\mathcal{F}}\boldsymbol{P}$ pass through the lens and are bend towards p. The focal length of the camera and lens are F and F_l respectively. d is the Euclidean distance of ${}^{\mathcal{F}}\boldsymbol{P}$.

The distortions caused by the lens are independent to the pixels and focal length of the PCM, and the 3D Cartesian points. Therefore, distortion is considered using new coordinates that are independent to the camera and 3D world known as *normalised image coordinates*. These are a set of 2D coordinates with no units. For a 3D Cartesian coordinate ${}^{\mathcal{F}}\boldsymbol{P}[X, Y, Z]$ and its corresponding homogeneous pixel coordinate $\tilde{\boldsymbol{p}}[u, v, 1]$ the normalised image coordinate (x, y) can be calculated as follows:

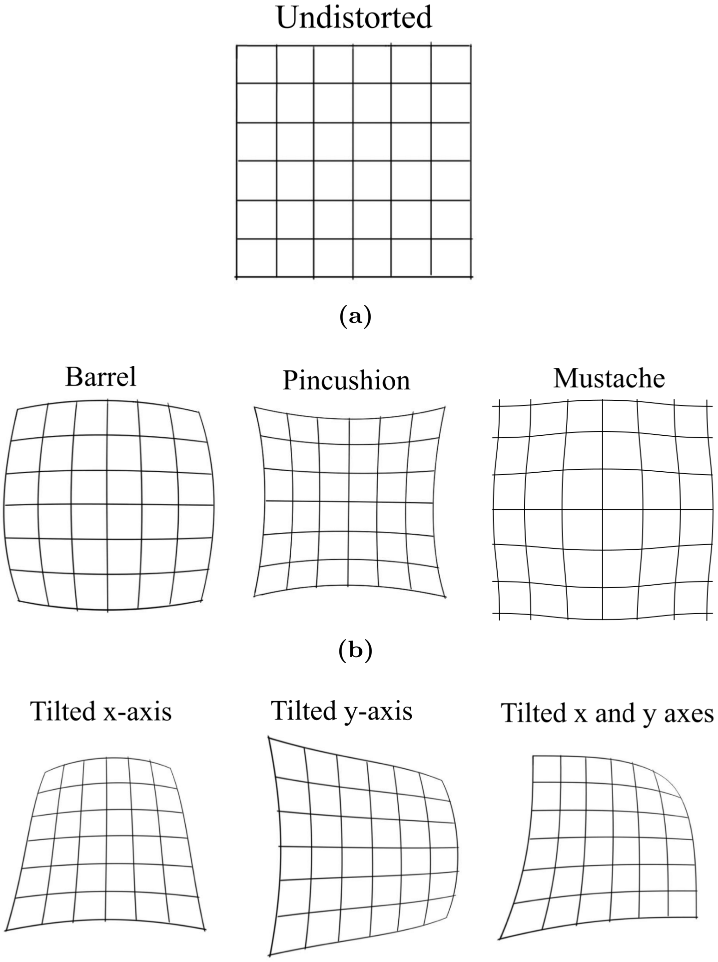
$$\begin{bmatrix} \mathrm{x} \\ \mathrm{y} \\ 1 \end{bmatrix} = \mathbf{K}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\mathcal{F}} . \tag{2.12}$$

Both ${}^{\mathcal{F}}\boldsymbol{P}$ and $\tilde{\boldsymbol{p}}$ will calculate the same normalised image coordinate. A well known model that describes both radial and tangential distortions is the Brown-Conrady model[3] [11] (or Plumb-Bob model) as follows:

$$\begin{aligned} \mathrm{x}_d &= \mathrm{x}(1 + K_1 r^2 + K_2 r^4 + K_3 r^6) + (2T_1 \mathrm{xy} + T_2(r^2 + 2\mathrm{x}^2)) \\ \mathrm{y}_d &= \mathrm{y}(1 + K_1 r^2 + K_2 r^4 + K_3 r^6) + (T_1(r^2 + 2\mathrm{y}^2) + 2T_2 \mathrm{xy}) . \end{aligned} \tag{2.13}$$

---

[3] For any distortion model there exists a correction model which uses the exact same radial and tangential parameters to convert distorted to undistorted pixels [7]. The values of the parameters cannot be used interchangeably.



(a)

(b)

(c)

Figure 2.9: Visualisation of distortions for real lens. In (a) is the original undistorted image. In (b) are the radial distortions of barrel, pincushion, and mustache. In (c) are the tangential distortions tilted along the x-axis, y-axis, and both x and y axes. Radial distortion does not move the centre pixel, whilst tangential distortion does move the centre pixel.

Where (x, y) and $(\mathrm{x}_d, \mathrm{y}_d)$ are the normalised image coordinates for the undistorted and distorted respectively. The Euclidean distance from the origin is $r = \sqrt{\mathrm{x}^2 + \mathrm{y}^2}$. In this model there are three radial coefficients $\{K_1, K_2, K_3\}$ and two tangential coefficients $\{T_1, T_2\}$. All distortion parameters have no units.

2.2 Hyperspectral Cameras

The typical 2D frame cameras have become a ubiquitous piece of portable hardware which can be found embedded in many devices. They capture images in the visible range with spatial dimensions like how the human eye would observe. The issue with these cameras is they do not add any more information than what is already seen. Spectroscopy is a technique which interprets how light interacts with matter by getting the response of a specimen to a vast range of light frequencies [12]. They have shown potential for determining the quality of food products in a non-destructive manner [13][14]. This technique is limiting in terms of the type and size of samples it can handle. Its main drawback is the lack of spatial dimensions, meaning that either it captures data for a single point or as an overall result for the entire sample. Hyperspectral cameras simultaneously obtain both spectral and spatial information to predict physical and chemical properties of an object [15].

2.2.1 Spectral Imaging

Spectral imaging sensors capture a scene at different spectral wavelength ranges called bands. The images that are captured are referred to as hypercubes. This data is three-dimensional: two spatial dimensions and one spectral dimension. There are three main types of spectral imaging classes: multispectral, hyperspectral, and ultraspectral, where the leading differences between them is how they manage their spectral dimension as shown in Figure 2.10. Multispectral imaging captures numerous discontinuous spectral bands which is similar to colour imaging. These cameras tend to have fewer number of spectral bands, typically around ten, and band widths tend to be larger (typical band width is 20 nm). In hyperspectral imaging the spectral bands are contiguous, which means a wavelength range is split up into equally spaced spectral bands that share borders. When a hypercube is captured is a hyperspectral camera, each spatial pixel is a continuous spectrum comprised of around 100+ spectral bands each with a bandwidth of 5-10nm. Ultraspectral imaging is similar in nature to hyperspectral but has even finer bandwidths of ¡1nm with around 1000 spectral bands. For the spectral imaging classes the cost of

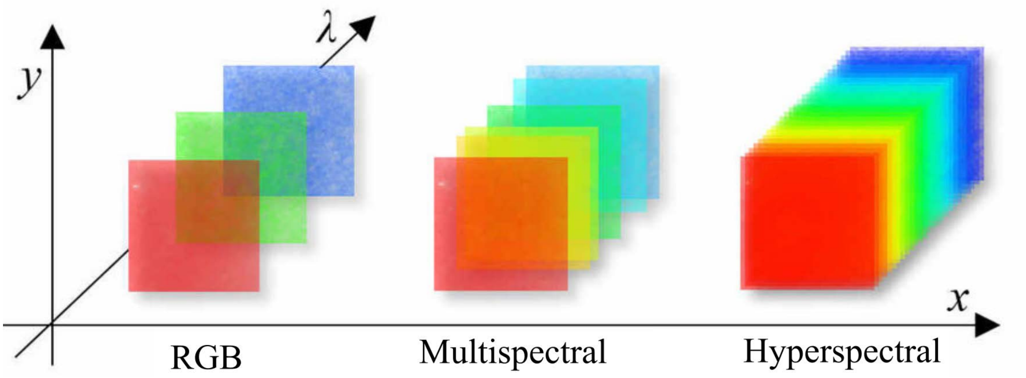


Figure 2.10: A comparison of the spectral bands in colour (red, green, and blue colour channels (RGB)), multispectral and hyperspectral imaging [16]. In colour and multispectral imaging there are discontinuous bands, while hyperspectral has contiguous bands forming a spectrum.

hardware, hypercube acquisition time, hardware complexity and data throughput, can be ordered from lowest to highest as multispectral, hyperspectral and ultraspectral.

2.2.2 Scanning Methods

The cameras have specific modes of scanning such as point, line, area and single-shot. Point scan records data for all frequency bands at a specific spatial point location, while line-scan (also referred to as push-broom), records data for a line of pixels. Area scan records a 2D area of pixels for a single frequency band. Single-shot attempts at recording data for all pixels and bands. The line-scan hyperspectral cameras are typically used for food quality purposes [17][18][19], but either the camera or the sample must be moved, and multiple images must be taken to get the full hypercube of the specimen. This is useful when combined with automated systems with conveyor belts [18]. The issue with the line-scan method is the likely underexposure of certain bands as the exposure time is short and fixed, to ensure other bands do not saturate [14][20].

2.2.3 Sensing Modes

The sensing mode of a hyperspectral camera is another crucial factor which depends on the application and objects it is capturing. It alters the position of the light source and the camera so that the radiation is captured in different ways. The most common modes are reflectance, transmittance and interactance [15, 21, 22]. In reflectance mode, the camera and the light source are looking directly at the specimen with the light source at an angle. Incident radiation from the light source reflects off the surface of the object and gets captured by the camera. The angle helps to reduce specular reflection but surface properties such as the shape, texture, defects, and roughness affect the measured reflectance [15]. With transmittance, the light source and the camera are facing each other with the specimen placed between them. The camera collects all the radiation that gets transmitted through the specimen. Although this radiation is relatively weak, internal information such as defects can be determined with little to no effects from surface properties of the specimen [23]. However, the thickness of the specimen will heavily impact the quality of the captured image as the amount of attenuation is proportional to the thickness. Environmental factors will also affect the captured image as a room with high ambient lights would make it difficult to measure the low levels of transmitted radiation through a specimen [21]. In interactance mode, the specimen is captured using the camera by ensuring that the light rays of the illumination source are parallel to the viewing direction. This combines the advantages of reflectance and transmittance by capturing the light that has penetrated specimen and scattered towards the camera's direction. Deeper effects can be seen in the specimen with little to no light from the surface [24]. However, interactance involves a more cumbersome setup as the camera must be blocked from both light that comes directly from the source and from the surface being illuminated. This maybe involve a setup where the light source is much closer to the specimen than the camera and would require a sophisticated light source that may include lenses and mirrors in order to control the ray directions.

2.2.4 Illumination

Hyperspectral cameras rely on sufficient illumination of objects within their spectral bands. This must provided by an external light source. There are four types of light sources that are commonly used with these cameras: halogen, light emitting diodes (LEDs), lasers and tunable sources. Halogen lamps are the most common illumination used with hyperspectral imaging as they offer continuous spectra over the visible and IR wavelengths, whilst being relative cheap. The filament is made of tungsten wire inside of a glass bulb filled with an inert gas with a small amount of the halogen gases iodine or bromine. The lamps typically operate at a low voltage but have a short lifespan, and high heat output which alters their illumination spectra.

A LED is a semiconductor based light source, which is a common element of any electronic device due to its small size, low cost, long lifespan, low heat, high efficiency and high frequency response. They are a solid-state sources without any filament which means they are shock resistant and more durable. LEDs can produce narrow bandwidths of radiation in the ultraviolet (UV), visible and infrared (IR) wavelengths, and can also produce broadband white light. They have a major advantage of being highly directional, when compared to halogen bulbs, and multiple LEDs can be arranged to produce a larger source of any geometry. The major disadvantages of LEDs is their performance depends on the ambient temperature and proper thermal management needs to be considered, highly sensitivity to input voltage and currents due to their semiconductor nature, and generally for the same package, LEDs produce lower power of illumination when compared to halogen lamps.

Lasers are directional monochromatic light sources that work via stimulated emission, where input photons of specific frequencies interact with excited electrons to release even more photons. The high energy beams of radiation can produce fluorescent effects when illuminating a specimen which has been captured with hyperspectral cameras [25]. Lasers have a much narrower bandwidth than LEDs and produce higher illumination intensity.

Lastly, tunable light sources are comprised of broadband illumination source, such as a halogen lamb, with a wavelength dispersion device, such as a filter wheel with band-pass filters. This type of source illuminates a specimen with a narrow bandwidth of radiation which can be changed through the wavelength dispersion device (i.e. different band-pass filter). The source is advantageous for hyperspectral cameras that scan using area-scan method, where there is only a single spectral band being captures, but is impractical for the other scanning methods.

2.2.5 Types of Calibrations

Calibration of the hyperspectral camera is essential to ensure data can be captured in a consistent manner and valid comparison between different hypercubes can be made. The calibration procedure involves correcting the spatial parameters and standardising the spectral measurements [18][14]. There are four major types of calibrations which need to be addressed: reflectance, wavelength, spatial, and curvature. The basis behind the reflectance calibration is that the measured spectral values can fluctuate due to changes in experimental parameters, variations of the environment, or due to equipment. An example of a likely change could result from the light source. It may have been moved or shifted so the amount of light radiation that hits the specimen may have changed. The outcome would likely give inconsistent results if the experiments were repeated, which would affect the analysis of the spectral data. This calibration is performed by correcting raw measured intensity values using black and white reference images. The black reference image is taken when the camera is blocked from capturing any radiation from the environment, and any measured intensities are due to heating effects from the dark current inside the camera sensor [12]. The white reference image should be taken alongside the raw data using a high reflectance standard white surface such as Teflon [14]. Wavelength calibration is used to correct the pixels in the spectral dimension that are used to measure the intensity of specific frequency bands. The calibration procedure is carried out using calibration lamps that emit a constant narrow band of radiation. Hyperspectral cameras which apply known band-pass filters do not require this calibration. Finally, the curvature calibration corrects the reflection effects due to curved surfaces [17]. The amount of reflection is due to the angle between the incident radiation and the normal to the surface. These reflections can be removed by utilising models that approximate the surface properties by incorporating information from the incident radiation [26][14]. A simplified approach to avoid modelling of reflections is to assume the surface is perfectly matte or made of a Lambertian material [27][26]. Spatial calibration determines the fixed internal parameters of the camera sensor and the camera's pose in the real world.

2.3 Light Radiation Behaviour

The physics of light stems from an argument of whether it is best modelled as particle or a wave. Evidence of its wave behaviour has been seen in the single-slit diffraction and double-slit experiments, whilst the introduction of the *photon* from quantum mechanics explains the photoelectric effect. Therefore, the current model for light combines the particle nature of photons with an electromagnetic wave that is comprised of oscillating electric and magnetic fields that are perpendicular to each other. This concept is known as the *wave-particle duality*.

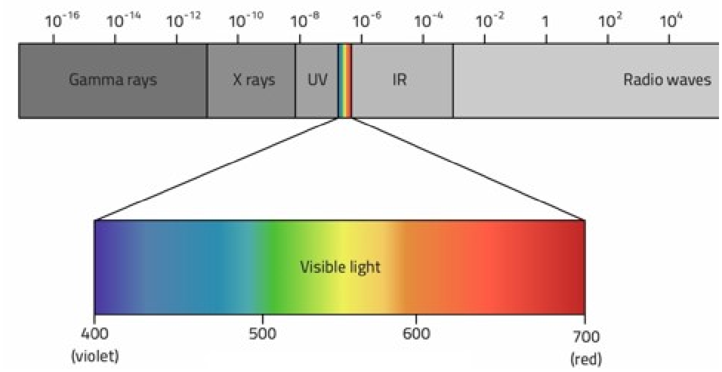
From the perspective of computer vision, using quantum models to describe light is impractical for both the scale in image scenes and the lack of sensitivity in camera hardware. In computer graphics, photo-realistic effects can be captured without the need for modelling quantum-level effects. In optics, properties such as reflection, refraction, diffraction and polarisation can be explained using simpler models.

This section will introduce the electromagnetic spectrum which characterises the various wavelengths (or frequencies) of light into different forms. The various light models will be introduced discussing their increasing complexity and practical applications. The physical quantity of light and will be thoroughly explained using units from *radiometry*. Properties such as refraction, diffraction and polarisation will not be discussed in this work.

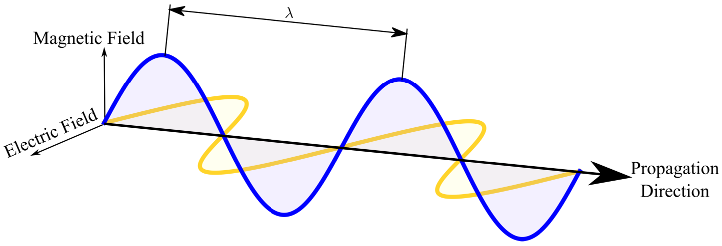
2.3.1 Electromagnetic Spectrum

The term *radiation* refers to electromagnetic waves released by a source travelling through space that carries energy. Electromagnetic waves are characterised by their frequency ν or wavelength λ as described in the electromagnetic spectrum which is shown in Figure 2.11. All forms of radiation travel at a constant speed known as the speed of light, which depends on the medium of travel. In a vacuum this speed is $c = 299,792,458ms^{-1} \approx 3 \times 10^8 ms^{-1}$ and it relates ν and λ as follows:

$$\lambda = \frac{c}{\nu} \,. \tag{2.14}$$



(a)



(b)

Figure 2.11: The electromagnetic spectrum (EM) in (a) showing the six forms of radiation [28]. The top axis is the wavelength of the radiation in metres. The visible light is highlighted showing the spectrum of colours between violet and red of wavelengths between 400nm to 700nm. In (b) an electromagnetic wave is shown, comprised of an electric field and a magnetic field that are perpendicular to each other moving in the propagation direction where energy flows. The wavelength λ is shown as the distance between two peaks.

In any other medium, the speed of light would be slower and reduce λ of the radiation. The term *light* typically refers to the visible radiation of the electromagnetic spectrum which can be seen by the human eye, but light and radiation can be used interchangeably

2.3.2 Light Models for Optics

Optics is a field of physics that focuses on the propagation of light (or radiation) in 3D space and its properties, which includes its interactions with matter which cause other related phenomena. There are models that are used to describe the behaviour and interactions of light which can be seen as a hierarchy shown in Figure 2.12. As you move up this hierarchy, the complexity of the models increases by capturing finer details and properties

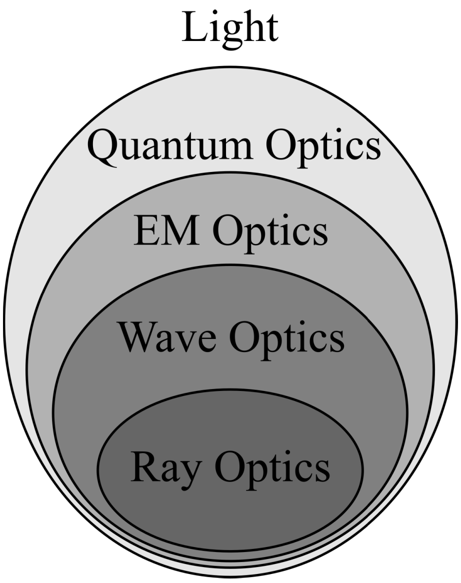


Figure 2.12: A hierarchy of light models that can be used in optics. There is an increasing complexity from ray to quantum optics that are able to capture greater interactions, properties and phenomena of light.

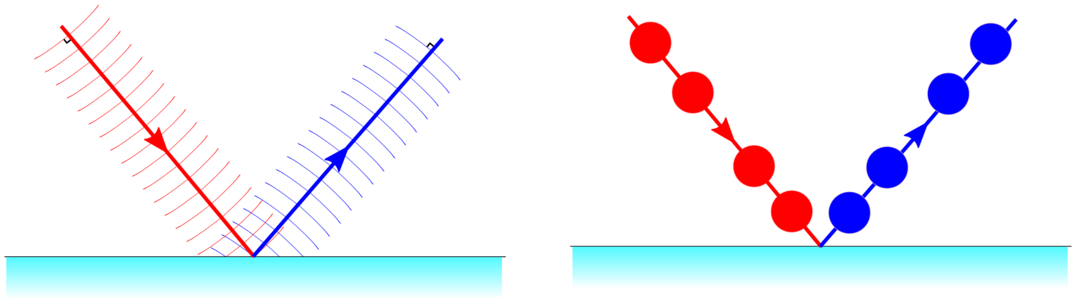
of the particles involved, in order to capture more phenomena that previous models cannot explain.

2.3.2.1 Ray Optics

Ray optics (or geometric optics), models light as propagating rays in the direction of energy flow moving at c for the medium. The rays are agnostic to either views of light consisting of particles or waves as shown in Figure 2.13. When light is considered a wave the ray direction is perpendicular to the wavefronts. When light is considered a particle[4], its ray coincides with the direction of motion. The space occupied by the light can be discretised into a number of rays which makes it advantageous for computers. Propagation of light in an environment can be modelled using techniques such as *ray tracing*, which can be considered a form of numerical integration of Maxwell's equations. The assumption with ray optics is that the dimensions of the environment is much greater than the wavelength of the light. This is able to capture light behaviours of reflection and refraction, and is used heavily to describe optical components and in image formation (refer to Section 2.1.3).

---

[4] This does not refer to photons. In this circumstance interactions with matter are not at the quantum scale where photons are absorbed and re-emitted. The light particles have interactions such as elastic collisions with surfaces.



(a)

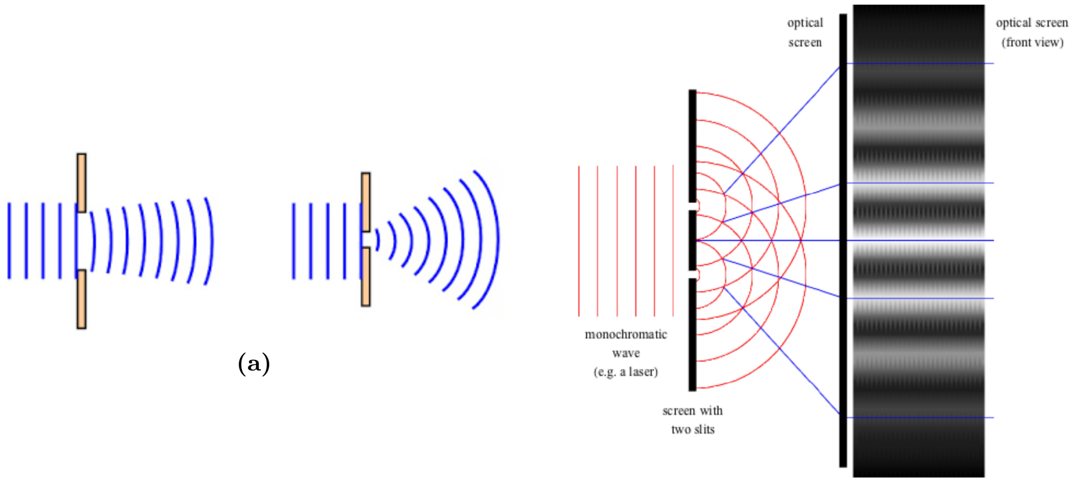
(b)

Figure 2.13: The ray optics model uses rays that point in the direction of energy flow. If the light is treated as waves, as shown in (a), the ray is perpendicular to the wavefronts. If the light is treated as particles, as shown in (b), the ray coincides with the direction of motion.

2.3.2.2 Wave Optics

There are two important behaviours that light exhibits which the ray optics model cannot explain. These are *diffraction* and *interference*. Wave optics is able to capture these two behaviours by assuming that light consists of planar waves, which can treated to be a planar solution of Maxwell's equations that only consider the electric field of an electromagnetic wave. Diffraction refers to the spreading of waves around an obstacle, where planar waves change into circular expanding waves due to a local change of propagation direction in the wavefronts. Interference first involves the diffusion of two coherent waves, which results in two circular expanding waves that experience superposition. The interference of light was first discovered through Young's double slit experiment [29], where light from a single source can be seen passing through two narrow slits, and resulted in the unexpected outcome of multiple dim and bright bands on a screen. This experiment can only be explained if light is considered to be a wave, where the bright bands are the result of constructive interference caused by the sum of either two crests or two troughs meeting, and the dark bands are the result of destructive interference caused by a crest and trough cancelling each other out.

There is a third behaviour that can be captured by wave optics but it requires an understanding in the direction of the oscillations in the wave. This is known as polarisation which extends the scalar planar wave into one of a vector. With the capability of modelling diffraction and interference, wave optics is able to better capture the interaction of light at



(a)

(b)

Figure 2.14: Wave optics is able to explain both diffraction and interference which are both behaviours of light. In (a) diffusion can be seen where a planar wave is passing through narrow and wide slits, which result in circular expanding waves. In (b), interference of light can be seen in Young's double slit experiment [29]. A single monochromatic source can be seen passing through two slits which results in two circular expanding waves that experience superposition. The effects of the superposition can be seen as dim and bright bands on a screen.

edges or boundaries of an object when compared to ray optics for rendering applications. Wave optics is used heavily in applications that involve radio waves, such as communication and broadcasting, as the large wavelength low frequency radiation commonly experiences diffraction with geographical objects.

2.3.2.3 Electromagnetic and Quantum Optics

Electromagnetic and quantum optics have been combined together as they represent the current pinnacle of light models for interactions at the atomic and quantum scale. In electromagnetic optics, light is modelled as per the definition of electromagnetic waves where the wave is comprised of an electric field and a magnetic field oscillating perpendicular to each other. There is no simplification made to the electromagnetic wave. The propagation of the wave is governed by Maxwell's equations. Beyond electromagnetic optics is quantum optics, where light is treated as discrete packets of energy, known as photons which are an elementary particle with zero mass. The interactions of photons with matter behave as both like a particle and like a wave which depends on the detection of the photon. In many cases, the location of individual photons are described as a probability distribution that are governed by Maxwell's equations.

These models can describe interactions with external magnetic and electric fields, all forms of scattering such as Rayleigh scattering, and with quantum optics, the photon's particle like behaviour accurately describes the observations in the photoelectric effect, where the emission of electrons is only affected by the frequency of the incoming radiation. The behaviours of light that are captured with such models are typically not visible by the human eye or traditional cameras, as such, these models are impractical for applications in computer graphics and computer vision.

2.4 Radiometry

The field that quantifies the energy of radiation is referred to as radiometry. It is interested in the physical quantity of energy that is transported by the electromagnetic waves (or photons) from a source to a surface, imaging sensor, or observer. New terms and units are introduced to normalise energy per a direction, which is described using the *solid angle*, and normalise energy per surface area. There are also spectral variants for the new units, that normalise per wavelength (or frequency) to account for a difference in sensitivity of imaging sensors to radiation of different wavelengths. The spectral variants will have an added $\mathrm{d}\lambda$ in the denominator of each definition to normalise per wavelength band. The International System of Units (SI) units for each definition will be included to aid in understanding the quantity.

2.4.1 Solid Angle

In radiometry the directional component for light energy is described using a solid angle ω with the units of steradian $[sr]$. A solid angle can be thought of as the 3D version of a 2D angle. If a 2D angle is subtended by an arc, than a solid angle is subtended by an area A of a sphere with radius r as follows:

$$\omega = \frac{A}{r^2}. \quad [sr] \tag{2.15}$$

In radiometry we work with light rays $\hat{\boldsymbol{\omega}}$ that illuminate small patches ∂A which subtend a solid angle of $\partial\omega$.

2.4.2 Radiant Flux

The radiant flux Φ is the flow of radiant energy Q per unit time contained within some radiation (or light) of wavelength λ as follows

$$\Phi(\lambda) = \frac{\partial Q}{\partial t}. \quad [W] \tag{2.16}$$

2.4.3 Irradiance

The irradiance E is the radiant flux per unit area arriving at a surface point $\boldsymbol{P}$ (e.g. object or imaging sensor) for radiation of wavelength λ and is defined as follows:

$$E(\boldsymbol{P}, \lambda) = \frac{\partial \Phi}{\partial A}. \quad [Wm^{-2}] \tag{2.17}$$

2.4.4 Radiant Intensity

The radiant intensity s is the radiant flux per unit solid angle leaving a source in a given direction $\hat{\boldsymbol{\omega}}$ for radiation of wavelength λ and is defined as follows:

$$s(\hat{\boldsymbol{\omega}}, \lambda) = \frac{\partial \Phi}{\partial \omega}. \quad [Wsr^{-1}] \tag{2.18}$$

2.4.5 Radiance

The radiance L is the radiant flux that is either arriving or leaving a surface point $\boldsymbol{P}$ with surface normal $\hat{\boldsymbol{n}}$ along a given direction $\hat{\boldsymbol{\omega}}$ for radiation of wavelength λ. If we consider the viewer's direction, the radiance is the radiant flux that is reflected off a surface in the direction of the viewer. It is defined as the radiant flux per unit foreshortened area per unit solid angle as follows:

$$L(\boldsymbol{P}, \hat{\boldsymbol{\omega}}, \lambda) = \frac{\partial^2 \Phi}{\partial A_\perp \partial \omega} . \quad [W m^{-2} sr^{-1}] \tag{2.19}$$

The foreshortened area $A_\perp$ is a reduction in area due to the surface being tilted away from $\hat{\boldsymbol{\omega}}$ and can be calculated as follows $A_\perp = (\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{\omega}})A$.

Radiance can be used to describe the irradiance E (defined in Equation (2.17)) arriving at a surface as an integral of the incoming L_i over all incoming directions $\hat{\boldsymbol{\omega}}_i$ as follows:

$$E(\boldsymbol{P}, \lambda) = \int_\Omega L_i(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \lambda)(\hat{\boldsymbol{n}} \bullet \hat{\boldsymbol{\omega}}_i) \, \mathrm{d}\hat{\boldsymbol{\omega}}_i \; . \tag{2.20}$$

2.5 Reflectance Models

An observer captures a scene by collecting the resulting light, headed towards their direction after it has interacted with the matter in the scene. If we focus on a single object in the scene, the outgoing light from that object as seen by the observer depends on the incoming light from the source, the observer's direction, the shape of the object, and the surface and internal material properties for the matter of the object. The *reflectance* defines the ability of a surface to capture radiant flux and emit it in a given direction. This is considered a property of the material that varies with the light's wavelength . In a general sense, reflectance is directional, but assumptions applied to surfaces, such as *isotropy*, can remove its directional dependence.

Reflectance models have been designed that consider the illumination on a surface, the surface geometry, the internal scattering of light, and the viewer's direction. These models are defined using the bidirectional reflectance distribution function (BRDF). The surfaces geometry involves a discussion of roughness, which considers the structure and shape of micro-facets on the surface. This section will introduce the BRDF, micro-facet geometry, and some important BRDF models for this thesis. The BRDFs attempt to capture a complex behaviour of light with materials, therefore, certain assumptions are required. These models are defined using ray optics (see Section 2.3.2.1), we ignore any fluorescent effects of the material, generally light that is reflected from other objects is negligible, self-shadowing is ignored, and material properties are assumed to be at steady-state and are not affected by external factors (i.e. temperature, pressure, etc).

2.5.1 Bidirectional Reflectance Distribution Function

The BRDF is a function f_{brdf} defined for a surface at a point $\boldsymbol{P}$ with a given material as the fraction of irradiance E, from a source, arriving at the surface reflected as radiance L in a given direction. A spherical coordinate system is defined at a surface as shown in Figure 2.15. The incident and viewing directions are $\hat{\boldsymbol{\omega}}_i[\theta_i, \varphi_i]$ and $\hat{\boldsymbol{\omega}}_r[\theta_r, \varphi_r]$ respectively. The BRDF is then defined as follows:

$$f_{brdf}(\boldsymbol{P}, \theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) = f_{brdf}(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \hat{\boldsymbol{\omega}}_r, \lambda) = \frac{\mathrm{d}L\left(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_r, \lambda\right)}{\mathrm{d}E\left(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \lambda\right)} . \quad [sr^{-1}] \tag{2.21}$$

Where dL and dE represent small changes in the radiance and irradiance respectively caused by light arriving at $\boldsymbol{P}$ along the single ray $\hat{\boldsymbol{\omega}}_i$. Defining it in this manner ignore any light that does not arise from the source (i.e. global illumination). Notice that dE depends on $\hat{\boldsymbol{\omega}}_i$ which goes against its definition in Equation (2.17). The reason for this is that the irradiance dE arriving at the small area at $\boldsymbol{P}$, is effectively the radiance L_i from Equation (2.20) arriving at the same point along the single direction $\hat{\boldsymbol{\omega}}_i$. The BRDF is dependent on the wavelength λ of the incoming light, has units of sr^{-1}, and always results in a non-negative value. The BRDF also follows the Helmholtz reciprocity principle [30], which can summarised to say that if the incident and viewing directions are switched, the BRDF is equal as follows:

$$f_{brdf}(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \hat{\boldsymbol{\omega}}_r, \lambda) = f_{brdf}(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_r, \hat{\boldsymbol{\omega}}_i, \lambda) . \tag{2.22}$$

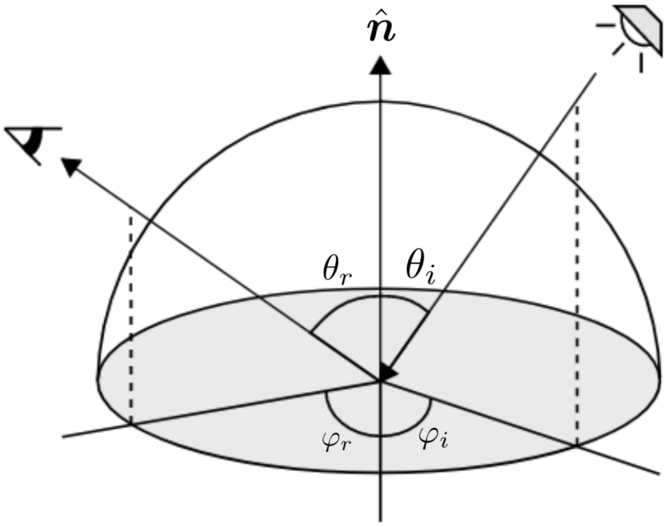


Figure 2.15: A coordinate system defined at a surface for a bidirectional reflectance distribution function (BRDF) with surface normal $\hat{\boldsymbol{n}}$. The directions are defined using spherical coordinates (θ, φ) which are the zenith and azimuth angles. The incident and viewing directions are (θ_i, φ_i) and (θ_r, φ_r) respectively.

A BRDF assumes no self-emission of light (i.e. fluorescent effects) and that radiant energy is conserved through the following principal:

$$\int_{\Omega} f_{brdf}(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \hat{\boldsymbol{\omega}}_r, \lambda) \cos(\theta_i) d\hat{\boldsymbol{\omega}}_i \leq 1 \,, \tag{2.23}$$

The BRDF is a function of five variables which can be simplified to four variables if we consider the material to be isotropic. This refers to an in-variance of material properties to rotations about the surface normal $\hat{\boldsymbol{n}}$. Therefore, the BRDF of an isotropic surface can be defined as $f_{brdf}(\boldsymbol{P}, \theta_i, \theta_r, |\varphi_i - \varphi_r|, \lambda)$.

2.5.2 Light Transport Equation

The light transport equation is a solution to the outgoing radiance at a point on a surface, by summing the reflected and emitted radiance components to ensure radiant energy is conserved. For a point $\boldsymbol{P}$ on a surface with a surface normal $\hat{\boldsymbol{n}}$ and a BRDF of $f_{brdf}(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \hat{\boldsymbol{\omega}}_r, \lambda)$, the outgoing radiance L_r in the direction $\hat{\boldsymbol{\omega}}_r$ is calculated using the light transport equation as follows:

$$L_r(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_r, \lambda) = L_e(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_r, \lambda) + \int_{\Omega} f_{brdf}(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \hat{\boldsymbol{\omega}}_r, \lambda) \cdot L_i(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \lambda) \cdot (\hat{\boldsymbol{n}} \bullet \hat{\boldsymbol{\omega}}_i) \,\mathrm{d}\hat{\boldsymbol{\omega}}_i \ . \tag{2.24}$$

Where L_i is the incoming radiance towards $\boldsymbol{P}$ along $\hat{\boldsymbol{\omega}}_i$ and the integral sums up contribution of L_i from all directions $\hat{\boldsymbol{\omega}}_i$ on the hemisphere Ω. L_e is the self-emitted radiance at $\boldsymbol{P}$ which for a non-emitting and non-fluorescent material $L_e = 0$.

2.5.3 Microfacet Geometry

Smoothness, shininess and roughness are words used to describe surface geometry of a material at the microscopic scale. The surface geometry is independent to material properties and is modelled by small perfect mirror segments called facets, which join together to form the surface. The facets are defined by their local surface normal, as shown in Figure 2.16 and the alignment of the surface normals with respect to light source will define the roughness or shininess of a surface. If all facets are aligned to a single orientation then a perfect mirror is formed. Otherwise, if an equal number of facets are aligned for any orientation on a hemisphere, the surface is perfectly diffuse. The orientation of the facet normals is treated to be a distribution.

2.5.4 Specular Model

The specular model represents one end of the extreme where only specular reflections occur at the surface. Such a material is known as a mirror that reflects incident irradiance in a single direction as follows:

$$f_{brdf}(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \hat{\boldsymbol{\omega}}_r) = \rho_k \frac{\delta(\theta_i - \theta_r)\delta(\varphi_i + \pi - \varphi_r)}{\cos(\theta_i)} \, . \tag{2.25}$$

where δ is the Dirac delta function used to mathematically describe the single infinitely thin light ray that follows the incident and viewing directions, and ρ_k is the specular reflectance, which describes the actual proportion of radiant flux that gets reflected in a

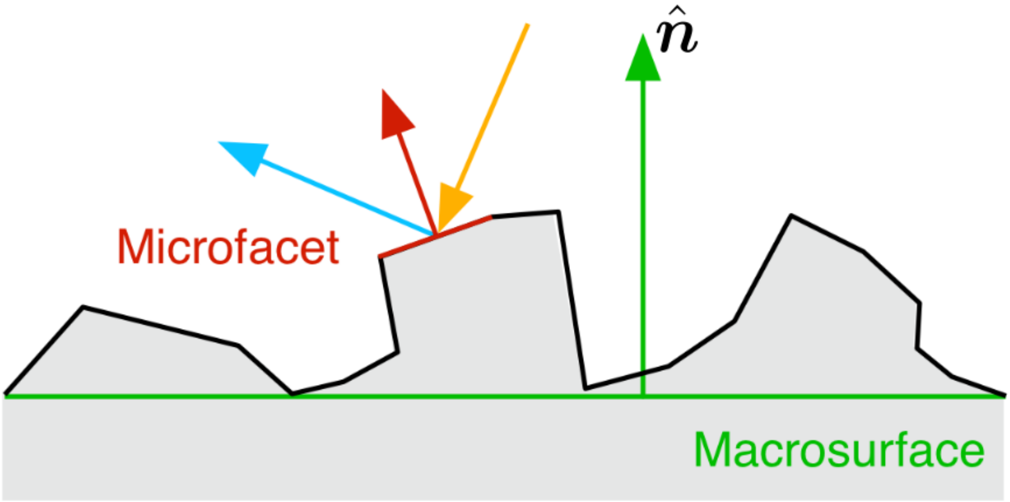


Figure 2.16: Microfacets of a surface which are defined by their local surface normal (red arrow), compared to the surface normal at the macrosurface scale (green arrow). The incoming irradiance (yellow arrow) is reflected in a specular manner as radiance (blue arrow) about the local surface normal.

specular manner as there is no such practical mirror that reflects 100% of the irradiance. The denominator of the model is used to ensure conservation of radiant energy. Notice that λ is not shown as the function applies to all wavelengths.

2.5.5 Lambertian Model

One of the most widely used models involves pure diffuse reflections which are described using the Lambertian model. Unlike specular reflections which occur at the surface, diffuse reflections occur within the material. A surface that is considered to be Lambertian, follows Lambert's cosine law where the effective irradiance arriving at the surface depends on the cosine of the angle between the light direction and surface normal, and the radiant flux is reflected equally in all directions. The BRDF for a Lambertian surface is:

$$f_{brdf}(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \hat{\boldsymbol{\omega}}_r, \lambda) = \frac{\rho(\lambda)}{\pi}\,, \tag{2.26}$$

where ρ is the wavelength dependent diffuse reflectance material property which represents the actual proportion of incoming radiant flux that is reflected outwards in a diffuse manner. The division by π accounts for the radiance being equally emitted in all directions over the hemisphere which follows Lambert's cosine law.

2.5.6 Dichromatic Reflectance Model

The dichromatic reflectance model (DRM) assumes that the reflected light at a point on a surface exhibits a linear combination of both specular and diffuse reflections, which occur at the surface and within the material respectively. Its BRDF is defined as:

$$f_{brdf}(\boldsymbol{P}, \hat{\boldsymbol{\omega}}_i, \hat{\boldsymbol{\omega}}_r, \lambda) = \frac{\rho(\lambda)}{\pi} + \frac{k(\hat{\boldsymbol{\omega}}_r)}{\cos(\theta_i)}\,, \tag{2.27}$$

where ρ_λ is the wavelength dependent reflectance material property. The specular coefficient k represents the proportion of radiant flux that is reflected in a specular manner at $\boldsymbol{P}$ and depends the viewing direction $\hat{\boldsymbol{\omega}}_r$. It is caused by the irregularities in the microfacets on the surface.

Chapter 3

Active Calibration of Line-scan and Frame Camera System

Multi-modal vision sensing systems have become increasingly common in robotics for applications in dynamic or featureless environments. Combining sensing modalities in a complementary manner overcomes the limitations of individual sensors when working in difficult scenarios with poor lighting, occlusions and texture-less surfaces. The ubiquitous frame camera, which provides dense 2D spatial information, is often used for multi-frame operations such as visual odometry [31] or simultaneous localisation and mapping (SLAM) [32], but is limited to the visible spectrum. Hyperspectral cameras can see beyond the visible spectrum, and are commonly found to capture data in a line-scan manner, with a single spatial dimension, that makes spatial calibration more difficult. Therefore, this chapter presents a spatial calibration procedure of a line-scan hyperspectral camera, by first combining the hyperspectral with a 2D frame camera into a single camera system, which is then used to capture images of a single planar calibration board. Measured noisy feature points from the images are used to calibrate the camera system. Therefore, a novel algorithm is devised that minimises the uncertainty in the calibration by considering the information gain from individual image measurements.

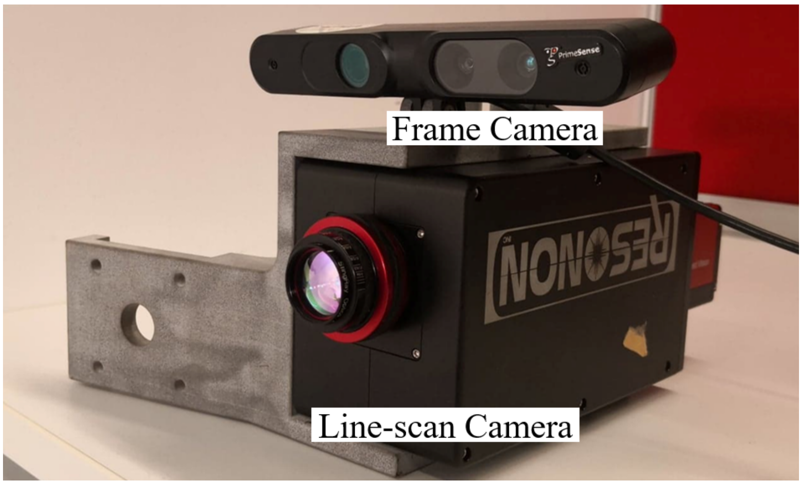


Figure 3.1: Rigidly mounted camera system comprised of a Resonon Pika NIR-320 hyperspectral line-scan camera and a Primesense Carmine 1.09 frame camera.

For hyperspectral cameras, the line-scan mode of operation is advantageous for high resolution and high frame rate imaging applications such as in conveyor belt vision systems [33][34][35]. The single spatial dimension makes it difficult to determine which 3D feature points are seen by the line-scan camera. To overcome this issue, special patterns have been designed to work with line-scan cameras, where feature points can be extracted through the knowledge of the pattern's dimensions and the line-scan images [34][35][36]. These generally calibrated a line-scan camera using a single image of non-planar calibration boards. The ideal calibration procedure would replicate Zhang's method [37] that uses a single planar checkerboard in multiple poses.

All camera pixel measurements exhibit uncertainty due to inherent random noise in the camera hardware. Therefore, uncertainty estimation plays a vital role as the measurements are not only used to perform the overall calibration through an optimisation, but prior to this, they aid in estimating feature points that are seen by the line-scan camera. These estimated feature points will now have an associated uncertainty which can be incorporated into the optimised parameters. The singular spatial dimension of the line-scan camera makes it difficult to ensure images of the calibration board are captured in different poses. Measurements of the board in similar poses will not have sufficient parallax which can degrade the optimisation. Therefore, such images should not be included in the calibration.

3.1 Literature Review

Hyperspectral cameras typically operate in a line-scan manner as the hardware is cheaper and more reliable compared to the other modes such as point-scan, area-scan and single-shot [14]. Further details of these modes can be found in Section 2.2.3. When it comes to calibration, a hyperspectral line-scan camera can be treated as an ordinary line-scan camera as the calibration only involves dealing with spatial information [36]. Compared to frame cameras, line-scan cameras present a challenge in calibration due to the single spatial dimension. The basis of camera calibration relies on point correspondences between images of a known target [33]. Frame cameras have well-established calibration methods, such as Zhang's method [37], where a checkerboard is captured by a camera in at least two different poses, and the corresponding points on the checkerboard are found in the images. Translating these methods to line-scan cameras are difficult for the sole reason of not being able to see all 2D points on the target. Therefore, the existing work can be split up into methods that calibrate a single line-scan camera and multi-modal calibrations, where measurements of a single calibration target in different poses can be used, if the poses can be acquired using external information or additional cameras. In order to consider uncertainty estimation in the line-scan calibration, existing work on uncertainty analysis of cameras need to be reviewed.

3.1.1 Single Line-Scan Calibrations

Feature points for camera calibration generally come from known patterns. For the line-scan camera, clever patterns have been devised which take advantage of the linear properties of the line-scan pixels. A zigzag pattern was popularised by [35], which comprised of vertical and diagonal lines drawn on two parallel offset planes. All relevant dimensions of the pattern were known. A single image of the pattern would be captured which contain points of the zigzag lines as seen by the line-scan camera. Using similar triangles, the ratio of the pixel distances between lines on the image is equal to the ratio of the physical distances between the lines on the pattern. Thus, the 3D points on the pattern, as seen by the camera, can be determined. Two assumptions are necessary: there must be reasonable distance between the planes to ensure some variation in depth (or parallax) is seen by the line-scan camera, and the pattern must be close to parallel to the image plane of the camera. Although, the error between a non-parallel calibration pattern was shown to be small in simulation, in practice such errors would magnify when dealing with the actual cameras. To address these assumptions, recent work has found the 3D points on the pattern by using a projective invariance property of lines known as the cross-ratio [34][38][36]. This relationship is valid if four corresponding points exist on two different lines that are related by a projective transformation. The use of the cross-ratio invariance helps to ease the assumption of the pattern needing to be parallel to the image plane. Modifications to the pattern have made the calibration less dependent on the distance between the planes by having the planes orthogonal to each other. In this case, the view-plane which describes the flat spatial plane seen by the line-scan camera, can be estimated to improve the calibration [34][36]. Most methods that have been mentioned can only calibrate using a single image, which offers little to no redundancy to noise caused by the camera or the environment [33]. The method outlined by [36] performed the calibration using multiple images of the orthogonal planed pattern using the cross-ratio. This calibration could only be used to optimise the intrinsic parameters of the line-scan camera.

3.1.2 Multi-modal Line-Scan Calibrations

The techniques that have been presented so far have solely focused on calibrating the line-scan camera by itself. Setting up the line-scan camera with another camera, such as a 2D frame camera, would provide the missing spatial information. These cameras could be treated as a single camera system, where they are set up in a rigid mount that will ensure the transformation between the cameras remain constant. If the intrinsic parameters of the frame camera are found prior using Zhang's method, the unknown calibration parameters are comprised of the intrinsic parameters of the line-scan camera and the rigid transformation between the cameras. The added frame camera helps to simplify the calibration process by only requiring the pattern to be comprised of a single plane [39][38][33]. The calibration can be further simplified if the pose of either the pattern or the camera is known, such as moving either of the two at a constant known speed [39][40], or through measuring the change in pose by an external device such as an inertial measurement unit (IMU) [41]. These calibration procedures are problematic as they require a complicated setup and errors may be introduced when estimating the poses by external means. Others have used this camera setup in a different manner by using the frame camera to transform all feature points into a single coordinate frame, and then using all feature points to calibrate the camera system. This has been demonstrated by [33][42][43], where it was shown that the error in the calibration parameters reduced with an increasing number of images.

3.1.3 Camera Uncertainty Estimation

An element of camera calibration that gets overlooked is the uncertainty. Most methods may qualitatively mention that uncertainty has decreased, but many do not take a quantitative approach to estimating it. If we consider the hardware of a traditional frame camera, there are many sources of noise present. In the sensor, the number of photons that have been collected by a pixel is determined by measuring a current from that pixel. Even when there is no light, current is recorded due to the thermal effects of the circuity. This is referred to as dark current. Thermal noise is also present in the circuity for measuring the current, and conversions from an analogue current to a digit value introduce their own noise [44]. These are all noises present during the capturing phase of an image, and now consider that we still need to perform some form of feature detection to carry out the calibration. It is important that this measurement uncertainty is addressed because the optimisations involved in the calibration will use these measurements to find the calibration parameters, and if the measurements have a high uncertainty then the calculated parameters may be unreliable. Uncertainty analysis of frame camera calibration was first formalised by [45], where measurement noise is propagated through each step of the calibration and the final optimised parameters have an estimated uncertainty. Others have extended this uncertainty estimation to applications such as determining the error in measuring mechanical parts [46], localisation and reconstruction using line features [47], and improving pose-graph visual SLAM [32]. In all cases, the uncertainty noise is approximated by a zero-mean Gaussian random variable, and uncertainty propagation is carried out using first-order covariance propagation. For line-scan cameras, uncertainty estimation of the intrinsic parameters has been studied for a moving pattern [40], as mentioned earlier, estimating either the pose of the pattern or the camera by external means, will introduce greater uncertainty. For handling multiple frame cameras, a complete uncertainty analysis of a stereo frame camera system was proposed by [32].

3.2 Cameras

In this calibration there are two different cameras that include the line-scan hyperspectral and red, green, and blue colour channels (RGB) frame. This section will introduce the cameras, and their respective models that are required to formulate the calibration which includes the calibration parameters. In order to quantify the noise in the optimised calibration parameters due to pixel measurements, the probabilistic nature of cameras needs to be discussed.

In this work, the perspective projection transformation of all cameras is defined through the pinhole camera model (PCM) as described in Section 2.1.3.1. The cameras are assumed to follow central projection, as described in Section 2.1.3, with a lens that causes distortion effects in the image as highlighted in Section 2.1.3.3.

3.2.1 Frame Camera

The frame camera refers to a traditional 2D imaging camera that can capture modalities of RGB, depth, or greyscale. In this chapter the camera is one that captures RGB. The PCM of a frame camera $\mathcal{F}$ projects a 3D point $\boldsymbol{P}[X, Y, Z]$ in the world coordinate system $\mathcal{W}$ to a 2D pixel $\boldsymbol{p}[u, v]$ on the image plane, through its projection matrix $\mathbf{M}_{\mathcal{F}} \in \mathbb{R}^{3\times 4}$ as follows:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{M}_{\mathcal{F}} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{3.1}$$

where $\boldsymbol{P}$ and p are defined in homogeneous coordinates with the appended dimension (which are introduced in Section 2.1.2), and s is an unknown arbitrary scale factor. The projection matrix can be defined as follows:

$$\mathbf{M}_{\mathcal{F}} = \begin{bmatrix} \mathbf{K}_{\mathcal{F}} & \mathbf{0} \end{bmatrix} \mathbf{E}_{\mathcal{F}} \,, \tag{3.2}$$

where $\mathbf{E}_{\mathcal{F}} \in \mathbb{R}^{4\times4}$ is a homogeneous transformation matrix known as the extrinsic matrix, which transforms a 3D point from $\mathcal{W}$ to $\mathcal{F}$. It is comprised of a 3D rotation matrix from the SO_3 group $\mathbf{R} \in \mathbb{R}^{3\times3}$ and a translation vector $\boldsymbol{t} \in \mathbb{R}^3$, which grouped together form the *extrinsic parameters*. The matrix $\mathbf{K}_{\mathcal{F}} \in \mathbb{R}^{3\times3}$ is the camera intrinsic matrix as shown:

$$\mathbf{K}_{\mathcal{F}} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \,, \tag{3.3}$$

where f_x and f_y are the focal lengths in the x and y direction respectively measured in pixels, and (u_0, v_0) is the principal point on the image plane. It is assumed there is no skew in the pixels for the frame camera . These parameters are grouped together and referred to as the *intrinsic parameters*. A greater in-depth explanation of both the intrinsic and extrinsic parameters can be found in Section 2.1.3.1.

The PCM for the frame camera , shown in Equation (3.1), can be written in non-homogeneous pixel coordinates as follows:

$$\begin{aligned} u &= u_0 + f_x \frac{r_{11}X + r_{12}Y + r_{13}Z + t_1}{r_{31}X + r_{32}Y + r_{33}Z + t_3} \\ v &= v_0 + f_y \frac{r_{21}X + r_{22}Y + r_{23}Z + t_2}{r_{31}X + r_{32}Y + r_{33}Z + t_3} \,, \end{aligned} \tag{3.4}$$

where $(r_{11}, r_{12}, r_{13}, ...)$ are elements of $\mathbf{R}$ and (t_1, t_2, t_3) are elements of $\boldsymbol{t}$. In reality , the frame camera has an optical lens in order to capture sufficient light which causes distortion on the image plane. The distortion for an optical lens is calculated using a new set of coordinates called the normalised image coordinates, that are independent to both the camera intrinsics and 3D world points. Further details of normalised coordinates can be found in Section 2.1.3.3. The normalised image coordinates (x, y) of (u,v) in Equation (3.4) are calculated as follows:

$$\begin{aligned} \mathrm{x} &= \frac{u-u_0}{f_x} = \frac{r_{11}X + r_{12}Y + r_{13}Z + t_1}{r_{31}X + r_{32}Y + r_{33}Z + t_3} \\ \mathrm{y} &= \frac{v-v_0}{f_y} = \frac{r_{21}X + r_{22}Y + r_{23}Z + t_2}{r_{31}X + r_{32}Y + r_{33}Z + t_3}\,. \end{aligned} \tag{3.5}$$

Pixel distortion can then be modelled using the Brown-Conrady model[11] that has parameters for radial and tangential distortions. For the frame camera , we include a three-order radial (K_1, K_2, K_3) and two-order tangential (T_1, T_2) distortion coefficients, and calculate the distorted normalised image coordinates ($\mathrm{x}_d, \mathrm{y}_d$) as follows:

$$\begin{aligned} \mathrm{x}_d &= \mathrm{x}(1 + K_1 r^2 + K_2 r^4 + K_3 r^6) + 2T_1\mathrm{xy} + T_2(r^2 + 2\mathrm{x}^2) \\ \mathrm{y}_d &= \mathrm{y}(1 + K_1 r^2 + K_2 r^4 + K_3 r^6) + T_1(r^2 + 2\mathrm{y}^2) + 2T_2\mathrm{xy}\,. \end{aligned} \tag{3.6}$$

Where $r = \sqrt{\mathrm{x}^2 + \mathrm{y}^2}$ is the Euclidean distance from the origin. The distortion coefficients are typically grouped together with the intrinsic parameters. The distorted pixel $\boldsymbol{p}_d[u_d, v_d]$ on the image plane can be found by converting ($\mathrm{x}_d, \mathrm{y}_d$) pixels as follows:

$$\begin{aligned} u_d &= f_x \mathrm{x}_d + u_0 \\ v_d &= f_y \mathrm{y}_d + v_0\,. \end{aligned} \tag{3.7}$$

If the frame camera is assumed to be intrinsically calibrated and the distortion is removed from the images, the vector for the calibrated intrinsic parameters is $\boldsymbol{\vartheta}_{\mathcal{F}_K} \in \mathbb{R}^4$. For the extrinsic parameters, $\mathbf{R}$ can firstly be parameterised using Euler angles $[r_z, r_y, r_x]$ and the vector for the unknown extrinsic parameters is $\boldsymbol{\vartheta}_{\mathcal{F}_E} \in \mathbb{R}^6$. Both $\boldsymbol{\vartheta}_{\mathcal{F}_K}$ and $\boldsymbol{\vartheta}_{\mathcal{F}_E}$ are defined as follows:

$$\boldsymbol{\vartheta}_{\mathcal{F}_K} = [u_0, v_0, f_x, f_y] \qquad \boldsymbol{\vartheta}_{\mathcal{F}_E} = [t_1, t_2, t_3, r_z, r_y, r_x] \tag{3.8}$$

3.2.2 Line-scan Camera

Now we consider the line-scan hyperspectral camera $\mathcal{H}$. In this work we focus on the spatial element of the camera and ignore the spectral capabilities, which means the following camera model would be applicable to line-scan cameras that capture other modalities. The line-scan camera can be regarded as a special form of a frame camera where all pixels are located along a single dimension or line on the image plane. This means we have a choice as to whether the line is aligned horizontally at $v = 0$ or vertically at $u = 0$. Calibrations devised by others [33, 35, 36, 38] have chosen to work with a vertical line, so we will adhere to this decision. The PCM of the line-scan camera projects a 3D point $\boldsymbol{P}[X, Y, Z]$ in $\mathcal{W}$ to a 2D pixel $\boldsymbol{p}[0, v]$ on the image plane, through its projection matrix $\mathbf{M}_{\mathcal{H}} \in \mathbb{R}^{3\times4}$ as follows:

$$s \begin{bmatrix} 0 \\ v \\ 1 \end{bmatrix} = \mathbf{M}_{\mathcal{H}} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} . \tag{3.9}$$

Where $\boldsymbol{P}$ and p are defined in homogeneous coordinates with the appended dimension (which are introduced in Section 2.1.2, and s is an unknown arbitrary scale factor. The projection matrix is defined as follows:

$$\mathbf{M}_{\mathcal{H}} = \begin{bmatrix} \mathbf{K}_{\mathcal{H}} & \mathbf{0} \end{bmatrix} \mathbf{E}_{\mathcal{H}} . \tag{3.10}$$

The extrinsics are unchanged but the intrinsics need to be modified. By knowing that all pixels have a u component of 0 then u_0 must also be 0. There is some confusion over the parameter f_x as there are three choices for its value: unchanged, $f_x = 1$ and $f_x = f_y$. Firstly, leaving it unchanged as an unknown intrinsic parameter is not a viable option due to the fact that $u = 0$ for all pixels, the optimal value would then always be $f_x = 0$. The second option, where $f_x = 1$, is a commonly made assumption [33, 34, 38, 39, 43] where the calibration focuses on estimating the view-plane of the line-scan camera, which is the 3D plane where all possible $\boldsymbol{P}$ exist that can be measured by the line-scan camera.

There are two problems with this assumption: errors in the u component are weighted less than those in the v component [36], and it does not align with the definition of f_x. As shown in Section 2.1.3.1, $f_x = F/l_x$ where F is the focal length of the PCM and l_x is the side length of a single pixel in the x-direction, both of which are in world units, and $F \gg l_x$. By assuming $f_x = 1$, then $l_x = F$ which cannot be true. This leaves us with the third option $f_x = f_y$ where the corresponding assumption is that the pixel's geometry is a square rather than a rectangle. The matrix $\mathbf{K}_{\mathcal{H}} \in \mathbb{R}^{3\times3}$ is shown as follows:

$$\mathbf{K}_{\mathcal{H}} = \begin{bmatrix} f_y & 0 & 0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} . \tag{3.11}$$

The PCM for the line-scan camera, shown in Equation (3.9), can be written in non-homogeneous pixel coordinates as follows:

$$\begin{aligned} 0 &= f_y \frac{r_{11}X + r_{12}Y + r_{13}Z + t_1}{r_{31}X + r_{32}Y + r_{33}Z + t_3} \\ v &= v_0 + f_y \frac{r_{21}X + r_{22}Y + r_{23}Z + t_2}{r_{31}X + r_{32}Y + r_{33}Z + t_3} . \end{aligned} \tag{3.12}$$

Similar to Equation (3.4), $(r_{11}, r_{12}, r_{13}, ...)$ are elements of $\mathbf{R}$ and (t_1, t_2, t_3) are elements of $\boldsymbol{t}$. Notice that we do not simplify the first equation any further as the definition of the projective transformation would be lost[1]. Now we consider the distortion effects of the line-scan camera due to its lens. Firstly, the normalised image coordinates (0, y) are similar to Equation (3.5) as follows:

$$\begin{aligned} 0 &= \frac{0}{f_y} = \frac{r_{11}X + r_{12}Y + r_{13}Z + t_1}{r_{31}X + r_{32}Y + r_{33}Z + t_3} \\ \mathrm{y} &= \frac{v - v_0}{f_y} = \frac{r_{21}X + r_{22}Y + r_{23}Z + t_2}{r_{31}X + r_{32}Y + r_{33}Z + t_3} . \end{aligned} \tag{3.13}$$

---

[1] In the projective transformation, we divide X and Y by Z. Simplifying the equation would remove the division.

The distortion is assumed to only occur along the y-direction and is parameterised using two-order radial (K_1, K_2) and one-order tangential T_1 [48]. The distorted normalised image coordinate $(0, \mathrm{y}_d)$ is then found as follows:

$$\mathrm{y}_d = \mathrm{y}(1 + K_1\mathrm{y}^2 + K_2\mathrm{y}^4) + 3T_1\mathrm{y}^2\,. \tag{3.14}$$

Only the y-component is shown as there is no distortion in the x-component. The distorted pixel $\boldsymbol{p}_d[0, v_d]$ on the image plane can be found by converting $(0, \mathrm{y}_d)$ to pixels as follows:

$$\begin{gathered} 0 = f_y\frac{r_{11}X + r_{12}Y + r_{13}Z + t_1}{r_{31}X + r_{32}Y + r_{33}Z + t_3} \\ v_d = f_y\mathrm{y}_d + v_0\,. \end{gathered} \tag{3.15}$$

The unknown intrinsic and extrinsic parameters of the line-scan camera can be combined into a vector $\boldsymbol{\vartheta}_{\mathcal{H}} \in \mathbb{R}^{11}$ which is defined as:

$$\boldsymbol{\vartheta}_{\mathcal{H}} = [v_0, f_y, K_1, K_2, T_1, t_1, t_2, t_3, r_z, r_y, r_x]\,. \tag{3.16}$$

It is important to consider the uncertainty in pixel measurements and how it will affect the calibration parameters for both cameras. The next section we discuss the probabilistic nature of the cameras.

3.2.3 Probabilistic Behaviour

Vision sensors exhibit uncertain behaviour which is seen as noise in the pixel locations on the image plane. Let us assume there is no motion of the cameras or the objects they are capturing. Therefore, the noise is only due to the internal hardware and electronics of the sensing circuity. For a pixel $\boldsymbol{p}[u, v]$, the uncertainty of its pixel location is assumed to be constant and independent for both components of $\boldsymbol{p}$, and thus modelled as a zero-mean Gaussian with a variance of σ_u^2. For I images, where each image contains N corresponding features, the u and v components of each feature are stacked into a single measurement

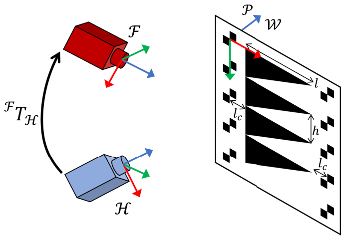


Figure 3.2: The coordinate frames of the cameras, world and calibration pattern. The pattern is annotated with the required dimensions. **T** represents the unknown rigid transformation between the cameras.

vector $\boldsymbol{u} \in \mathbb{R}^{2NI}$, where the noise is modelled as a zero-mean Gaussian with a diagonal covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}} \in \mathbb{R}^{2NI \times 2NI}$ as follows:

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{u}}, \boldsymbol{\Sigma}_{\boldsymbol{u}}) \,, \tag{3.17}$$

where $\boldsymbol{\mu}_{\boldsymbol{u}}$ is the mean of $\boldsymbol{u}$. These noisy measurements will induce uncertainties in all calibration parameters, where the noise in the parameters can also be modelled as zero-mean Gaussian variables. For the calibration parameters of the frame camera which are $\boldsymbol{\vartheta}_{\mathcal{F}_K}$ and $\boldsymbol{\vartheta}_{\mathcal{F}_E}$, their covariance matrices are defined as $\boldsymbol{\Sigma}_{\mathcal{F}_K} \in \mathbb{R}^{4\times 4}$ and $\boldsymbol{\Sigma}_{\mathcal{F}_E} \in \mathbb{R}^{6\times 6}$ respectively. Similarly, for the calibration parameter of $\mathcal{H}$ which is $\boldsymbol{\vartheta}_{\mathcal{H}}$, its covariance matrix is defined as $\boldsymbol{\Sigma}_{\mathcal{H}} \in \mathbb{R}^{11\times 11}$.

We have now described the appropriate models for camera calibration of both cameras, which are based on the PCM, and have shown their respective calibration parameters. There has also been a discussion of how the pixel noise will affect the calibration parameters. Now we introduce the procedure that is followed to calibrate the camera system.

3.3 Calibration Procedure

The cameras are rigidity combined into a single camera system. The steps for calibrating the camera system follow the works of [33] and [38]. The frame camera is assumed to be intrinsically calibrated with Zhang's method [37] using a checkerboard pattern. The calibration of the camera system involves capturing N known 3D Cartesian feature points using the line-scan camera in various different poses I. The optimised calibration parameters for the line-scan camera are $\widehat{\boldsymbol{\vartheta}}_{\mathcal{H}}$ and are found through the maximum likelihood estimation (MLE) by minimising the reprojection error:

$$\widehat{\boldsymbol{\vartheta}}_{\mathcal{H}} = \underset{\boldsymbol{\vartheta}_{\mathcal{H}}}{\operatorname{argmin}} ||\boldsymbol{u} - f_{\mathcal{H}}(\boldsymbol{\vartheta}_{\mathcal{H}}, \boldsymbol{x})||^2 \,, \tag{3.18}$$

where $\boldsymbol{u}$ is the pixel measurement vector as shown in Equation (3.17) and $\boldsymbol{x} \in \mathbb{R}^{3NI}$ is corresponding measurement vector of 3D Cartesian points with the X, Y and Z components stacked together. The function $f_{\mathcal{H}}$ represents the non-linear camera projection model for the line-scan camera as detailed in Equation (3.15). The problem can be solved by a non-linear least squares algorithm such as Levenberg-Marquardt or trust-region method.

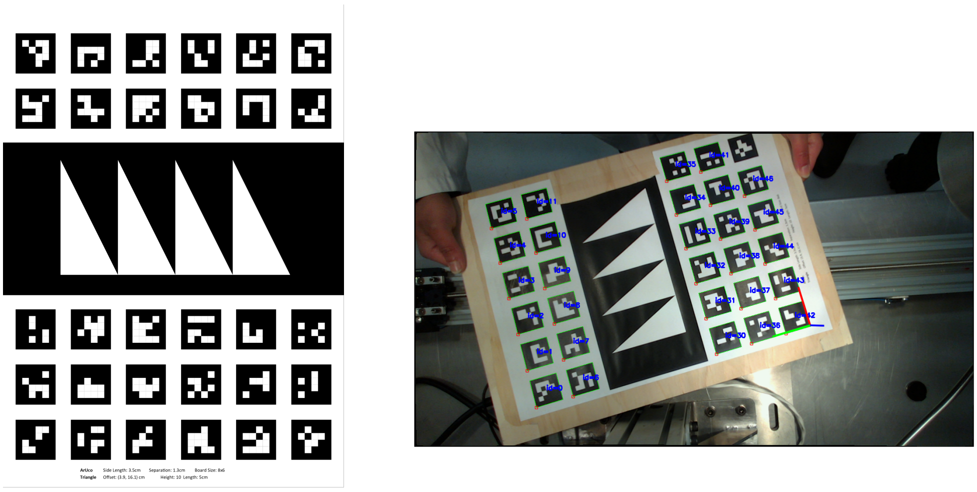
3.3.1 Pattern and Coordinate Frames

The cameras and the pattern are shown Figure 3.2 with their respective coordinate frames and the required dimensions of the pattern. All poses are relative to $\mathcal{W}$ and is attached to the pattern coordinate system $\mathcal{P}$. The cameras are assumed to be rigidly mounted together via an unknown transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{H}}$.

The first pattern is adopted from [38]. It is comprised of right-angled triangles which provide feature points for the line-scan camera and cross-corner markers for the frame camera to determine the pattern's pose. All points lie on the XY plane of the pattern. To avoid the calibration being under-determined, at least two different poses of the pattern must be captured, where the pattern contains at least three triangles and at least four cross-corner-markers. A new pattern was later devised where the cross-corner markers were replaced by ArUco tags [49] in order to automate and robustify the pose estimation of the pattern as shown in Figure 3.3.

3.3.2 Cross-ratio Point Correspondence

For the line-scan camera, its measurements can be seen as a line on the pattern which is known as the view-line. Given an image from the line-scan camera of the pattern, the



(a) (b)

Figure 3.3: In (a), the updated line-scan calibration board comprised of an ArUco [49] board and the triangle patterns. The ArUco board allows for automatic pose estimation from the frame camera images as shown in (b).

location of the view-line is difficult to determine and thus considered unknown. Therefore, we would not have the measurement vector $\boldsymbol{x}$ to solve the optimisation in Equation (3.18). In order to determine the view-line the cross-ratio η has been employed [33][34][36][38]. It is a scale invariant relationship which links a set of four collinear points, with another set, given that they are related by a projective transformation as follows:

$$\eta = \frac{(v_3 - v_1)(v_4 - v_2)}{(v_3 - v_2)(v_4 - v_1)} = \frac{(Y_3 - Y_1)(Y_4 - Y_2)}{(Y_3 - Y_2)(Y_4 - Y_1)} \,. \tag{3.19}$$

The values (v_1, v_2, v_3, v_4) are the v component of four collinear pixels and (Y_1, Y_2, Y_3, Y_4) are the Y component of the corresponding pattern coordinates in world units. If we examine Figure 3.4, $(p_1 ... p_8)$ are the feature points in pixels and $(P_1 ... P_8)$ are their corresponding pattern coordinates which we need to recover. Firstly, from Figure 3.4, the equations of the pattern lines $(L_1, L_2, ... L_N)$ with respect to the pattern coordinate system can be defined as:

$$\begin{aligned} Y &= h(i-1)/2 && i = (1,3,5,...N{-}1) \\ Y &= -Xh/l + h(2l_c + il)/2l && i = (2,4,6,...N) \end{aligned} \tag{3.20}$$

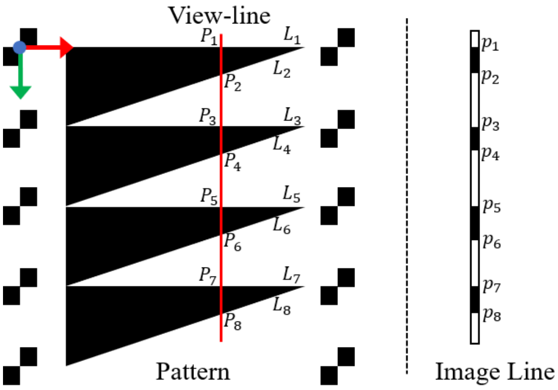


Figure 3.4: Red view-line of the line-scan camera which intersects the pattern lines ($L_1 ... L_8$) resulting in pattern feature points ($P_1 ... P_8$). The corresponding image line with the pixel feature points ($p_1 ... p_8$).

The following steps are taken to determine both X and Y components for all ($P_1 ... P_8$):

1. Y component of $(P_1, P_3, P_5, ... P_{N-1})$ are known.
2. Cross-ratio is used to determine Y component of $(P_2, P_4, P_6, ... P_N)$.
3. X component of $(P_2, P_4, P_6, ... P_N)$ is found using Equation (3.20).
4. Least squares linear fit of $(P_2, P_4, P_6, ... P_N)$.
5. X component of $(P_1, P_3, P_5, ... P_{N-1})$ is found by substituting into the linear fit from the previous step.

3.3.3 Common Reference Frame

The extrinsic transformation of pose I of the pattern with respect to the frame camera is $\mathbf{E}_{\mathcal{F}_I}$. This is estimated through the Perspective-n-Point (PnP) algorithm using the known cross-corner markers or ArUco tags. $\mathbf{E}_{\mathcal{F}_I}$ transforms feature points from $\mathcal{W}$ to the common reference frame of the frame camera as follows:

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{\mathcal{F}} = \mathbf{E}_{\mathcal{F}_I} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \tag{3.21}$$

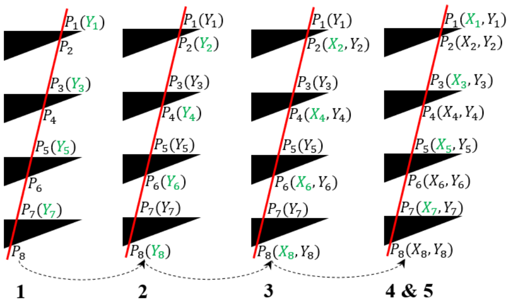


Figure 3.5: Steps in determining the feature points seen by the line-scan camera.

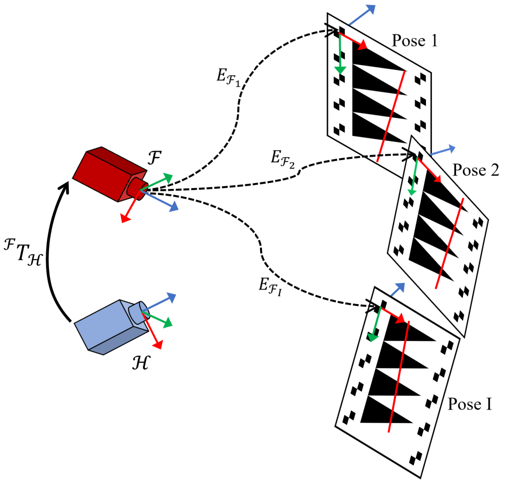


Figure 3.6: Demonstrating the extrinsics calculated from multiple views.

3.3.4 Closed Form Solution

A closed form solution for the calibration can be determined using direct linear transformation (DLT) [33][38][39]. This solution does not include the distortion parameters, thus it is used as the initial guess for the optimisation in Equation (3.18). DLT of Equation (3.12) results in two over-determined homogeneous equations with matrices $\mathbf{A}_1 \in \mathbb{R}^{NI\times 4}$ and

$\mathbf{A}_{23} \in \mathbb{R}^{NI \times 8}$ which are defined as follows:

$$\boldsymbol{A}_1 = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ X_{NI} & Y_{NI} & Z_{NI} & 1 \end{bmatrix}$$
$$\boldsymbol{A}_{23} = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{NI} & Y_{NI} & Z_{NI} & 1 & -v_{NI}X_{NI} & -v_{NI}Y_{NI} & -v_{NI}Z_{NI} & -v_{NI} \end{bmatrix}, \tag{3.22}$$

where (X_1, Y_1, Z_1) to (X_{NI}, Y_{NI}, Z_{NI}) are all 3D feature points with respect to the frame camera whose corresponding pixel coordinates in the line-scan camera are $(0, v_1)$ to $(0, v_{NI})$. The solutions to $\boldsymbol{A}_1$ and $\boldsymbol{A}_{23}$ are $\boldsymbol{m}_1 \in \mathbb{R}^4$ and $\boldsymbol{m}_{23} \in \mathbb{R}^8$ respectively. The closed form solution for the calibration parameters of the line-scan camera are obtained as such:

$$\boldsymbol{n}_1 = \begin{bmatrix} \boldsymbol{m}_1(1) & \boldsymbol{m}_1(2) & \boldsymbol{m}_1(3) \end{bmatrix}^\top$$
$$\boldsymbol{n}_2 = \begin{bmatrix} \boldsymbol{m}_{23}(1) & \boldsymbol{m}_{23}(2) & \boldsymbol{m}_{23}(3) \end{bmatrix}^\top$$
$$\boldsymbol{n}_3 = \begin{bmatrix} \boldsymbol{m}_{23}(5) & \boldsymbol{m}_{23}(6) & \boldsymbol{m}_{23}(7) \end{bmatrix}^\top$$
$$\gamma_1 = \frac{\pm 1}{||\boldsymbol{n}_1||}$$
$$\gamma_3 = \frac{\pm 1}{||\boldsymbol{n}_3||}$$

Intrinsic parameters:

$$v_0 = \gamma_3^2(\boldsymbol{n}_2 \cdot \boldsymbol{n}_3)$$
$$f_y = ||\gamma_3^2(\boldsymbol{n}_2 \times \boldsymbol{n}_3)||$$

Extrinsic parameters:

$$\boldsymbol{r}_1 = \gamma_1\boldsymbol{n}_1, \quad \boldsymbol{r}_2 = -(\boldsymbol{r}_1 \times \boldsymbol{r}_3), \quad \boldsymbol{r}_3 = \gamma_3\boldsymbol{n}_3$$
$$t_1 = \gamma_1\boldsymbol{m}_1(4), \quad t_2 = \frac{\gamma_3\boldsymbol{m}_{23}(4) - v_0t_3}{f_y}, \quad t_3 = \gamma_3\boldsymbol{m}_{23}(8)\,.$$

The vectors $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ are the rows of the rotation matrix $\mathbf{R}$. Note that γ_1 and γ_3 each have two solutions which account for the mirror pose of ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{H}}$, therefore their correct values need to be chosen such that they replicate the camera setup.

This section has outlined the full calibration procedure of the camera system which started with the MLE of the reprojection, the use of the cross-ratio to determine 3D feature points on the pattern, transforming feature points to the common reference frame of the frame camera , and the closed form solution. We have not yet considered any form of noise from the cameras, but have used pixel measurements to determine 3D feature points on the pattern and estimate the pose of the pattern with respect to the frame camera . In the next section we consider the uncertainty in all estimated parameters introduced by pixel noise.

3.4 Uncertainty Propagation

The uncertainty in the calibration arises from the pixel noise from both the frame camera and the line-scan camera. Therefore, we now look at error propagation through the calibration process in order to estimate the uncertainty in the optimised parameters $\widehat{\boldsymbol{\vartheta}}_{\mathcal{H}}$. This closely follows the work of Ozog and Eustice [32] by using first-order covariance propagation as follows:

$$\begin{aligned} \boldsymbol{b} &= g(\boldsymbol{a}) \\ \boldsymbol{\Sigma}_{\boldsymbol{b}} &\approx \mathbf{J}_g \boldsymbol{\Sigma}_{\boldsymbol{a}} \mathbf{J}_g^\top , \end{aligned} \tag{3.23}$$

where $\boldsymbol{a}$ is a vector corrupted by zero-mean noise (generally Gaussian) with a covariance $\boldsymbol{\Sigma}_{\boldsymbol{a}}$, passed into a non-linear function g which returns an output vector $\boldsymbol{b}$. The output is corrupted by noise due to $\boldsymbol{a}$ with a covariance of $\boldsymbol{\Sigma}_{\boldsymbol{b}}$, whose value can be approximated using $\mathbf{J}_g$ which is the jacobian matrix of g with respect to $\boldsymbol{a}$.

3.4.1 Unscented Transform

First order covariance propagation of non-linear functions rely on the ability to calculate the jacobian. These could be easily determined for implicit and explicit functions which

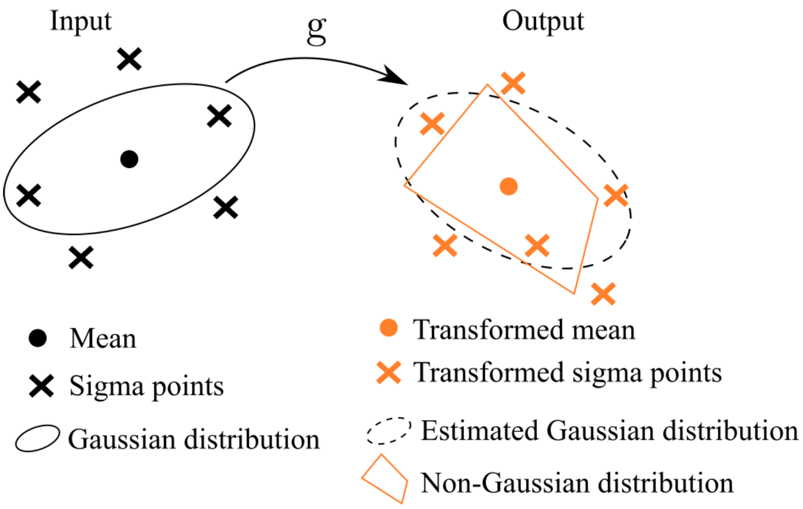


Figure 3.7: Visualisation of the unscented transformation used to estimate the distribution of an output noisy variable when a non-linear function g is applied to an input Gaussian variable. Even if the output variable is no longer Gaussian, the unscented transformation process will determine the best approximate Gaussian distribution.

contain a single closed-form expression. Some functions are highly complex with no closed-form expression due to multi-step processes, numerical or iterative methods, and matrix operations such as the inverse[2]. When carrying out uncertainty propagation with such functions the unscented transform (UT) can be used to approximate the covariance of the output variables. It assumes that the probability distribution of the noise in both the input and output variables is the same. For example, in Equation (3.23) if $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_x})$ then the UT would assume that the output vector $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu_y}, \boldsymbol{\Sigma_y})$. It is likely that the non-linear function g would alter the distribution to make it non-Gaussian, but UT would fit the best Gaussian distribution to the data as shown in Section 3.4.1.

The UT estimates the mean and covariance of the output variable through a weighted sum of a few calculated samples called sigma points $\mathcal{X}$. If $\boldsymbol{x} \in \mathbb{R}^N$, then $2N+1$ are sufficient to calculate the mean and covariance of $\boldsymbol{y}$ [50]. The sigma points for $\boldsymbol{x}$ are chosen as follows:

---

[2] In programs that can work with symbolic expressions, such as MATLAB, certain operations can lead to the phenomena of *expression swell*, where total size of the symbols rapidly increases factorially beyond the storage capabilities of the computer. The inverse of a non-diagonal symbolic matrix is one such example, where beyond a matrix of size 8 × 8 is practically impossible.

$$\begin{aligned}
\mathcal{X}_0 &= \boldsymbol{\mu}_{\boldsymbol{x}} \\
\mathcal{X}_i &= \boldsymbol{\mu}_{\boldsymbol{x}} + \left(\sqrt{(N+\lambda_{ut})\boldsymbol{\Sigma}_{\boldsymbol{x}}}\right)_i \quad i = 1, \ldots, N \\
\mathcal{X}_i &= \boldsymbol{\mu}_{\boldsymbol{x}} - \left(\sqrt{(N+\lambda_{ut})\boldsymbol{\Sigma}_{\boldsymbol{x}}}\right)_{i-N} \quad i = N+1, \ldots, 2N \\
W_0^{(m)} &= \lambda_{ut}/(N+\lambda_{ut}) \\
W_0^{(c)} &= \lambda_{ut}/(N+\lambda_{ut}) + \left(1-\alpha^2+\beta\right) \\
W_i^{(m)} &= W_i^{(c)} = 1/\{2(N+\lambda_{ut})\} \quad i = 1, \ldots, 2N \, ,
\end{aligned} \tag{3.24}$$

where α is a parameter that represents the spread of $\mathcal{X}$, κ is a parameter that represents how far $\mathcal{X}$ is from the $\boldsymbol{\mu}_{\boldsymbol{x}}$, β incorporates prior knowledge of the probability distribution ($\beta = 2$ for Gaussian distributions [50]), and $\lambda_{ut} = \alpha^2(N+\kappa) - N$. The weights for the mean and covariance are $W^{(m)}$ and $W^{(c)}$ respectively. Notice that the weights for $\mathcal{X}$ at the mean are larger than all other points. If $\mathcal{Y}$ is the sampled output of $\mathcal{X}$, then $\boldsymbol{\mu}_{\boldsymbol{y}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ can be approximated as follows:

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{y}} &\approx \sum_{i=0}^{2N} W_i^{(m)} \mathcal{Y}_i \\
\boldsymbol{\Sigma}_{\boldsymbol{y}} &\approx \sum_{i=0}^{2N} W_i^{(c)} \left(\mathcal{Y}_i - \boldsymbol{\mu}_{\boldsymbol{y}}\right) \left(\mathcal{Y}_i - \boldsymbol{\mu}_{\boldsymbol{y}}\right)^\top \, .
\end{aligned} \tag{3.25}$$

The UT is used to propagate noise for both cameras where a closed form solution of a function does not exist. The next section defines the covariance propagation for the extrinsics of the frame camera , where UT propagates noise due to the intrinsic parameters of the frame camera . The following section defines the covariance propagation of the calibration parameters of the line-scan camera, where UT propagates noise due to uncertain 3D feature points with respect to the frame camera .

3.4.2 Frame Camera

For the intrinsically calibrated frame camera, $\boldsymbol{\Sigma}_{\mathcal{F}_E}$ is determined by propagating covariances through the PnP algorithm. Ozog and Eustice [32] show that $\boldsymbol{\Sigma}_{\mathcal{F}_E}$ is a combination of the uncertainty in the feature pixels $\boldsymbol{\Sigma}_{\boldsymbol{u}}$, and the uncertainty in the intrinsic parameters

$\boldsymbol{\Sigma}_{\mathcal{F}_K}$ which can be approximated by a scaled UT as $\boldsymbol{\Sigma}_{UT-\mathcal{F}}$. Therefore, $\boldsymbol{\Sigma}_{\mathcal{F}_E}$ is defined as such:

$$\boldsymbol{\Sigma}_{\mathcal{F}_E} = (\mathbf{J}_f^\top \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} \mathbf{J}_f)^{-1} + \boldsymbol{\Sigma}_{UT-\mathcal{F}}\,, \tag{3.26}$$

where $\mathbf{J}_f \in \mathbb{R}^{6\times 4}$ is the Jacobian of the function for the PnP algorithm with respect to $\boldsymbol{\vartheta}_{\mathcal{F}_K}$. This function was assumed to be Equation (3.4) as the closed form for the PnP algorithm is difficult to express . $\boldsymbol{\Sigma}_{UT-\mathcal{F}}$ requires nine sigma points with the corresponding weights that are determined using $\boldsymbol{\vartheta}_{\mathcal{F}_K}$, $\boldsymbol{\Sigma}_{\mathcal{F}_K}$ and UT scaling parameters (α, β, κ) as shown in Equation (3.24).

3.4.3 Line-scan Camera

The uncertainty in the optimised line-scan camera parameters $\boldsymbol{\Sigma}_{\mathcal{H}}$ takes a similar form to Equation (3.26) and is referred to in the text as Ozog's method:

$$\boldsymbol{\Sigma}_{\mathcal{H}} = (\mathbf{J}_{\mathcal{H}}^\top \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} \mathbf{J}_{\mathcal{H}})^{-1} + \boldsymbol{\Sigma}_{UT-\mathcal{H}}\,, \tag{3.27}$$

where $\mathbf{J}_{\mathcal{H}} \in \mathbb{R}^{11\times 2NI}$ is the Jacobian of function f in Equation (3.18) with respect to $\boldsymbol{\vartheta}_{\mathcal{H}}$. $\boldsymbol{\Sigma}_{UT-\mathcal{H}}$ now arises from uncertainties in the world coordinates $\boldsymbol{x}$ given by the covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}} \in \mathbb{R}^{3NI\times 3NI}$. This covariance is required to calculate $\boldsymbol{\Sigma}_{UT-\mathcal{H}}$, and is determined by propagating the covariance $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ through the multi-step process of obtaining $\boldsymbol{x}$.

Firstly, the feature points on the pattern were determined using the corresponding pixel locations on the image line. The steps for estimating the feature points required the cross-ratio Equation (3.19), pattern line equations Equation (3.20), and a linear least squares fit. This process takes in uncertain pixel coordinates $\boldsymbol{u}$ and outputs their uncertain world coordinates $[X, Y, 0]_{\mathcal{W}}^\top$. Therefore, $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ gets propagated to these world coordinates. Then, the points are transformed to the frame camera via $\mathbf{E}_I$ per Equation (3.21) to give $\mathbf{x}$. $\mathbf{E}_I$

is equivalent to $\boldsymbol{\vartheta}_{\mathcal{F}_E}$ through a conversion from rotation matrix to Euler angles. For pose I, the uncertainty in $\boldsymbol{\vartheta}_{\mathcal{F}_E}$ is $\boldsymbol{\Sigma}_{\mathcal{F}_E}$ which is calculated in Equation (3.26). Propagating both uncertainties in $[X, Y, 0]_{\mathcal{W}}^{\top}$ and $\boldsymbol{\Sigma}_{\mathcal{F}_E}$ gives $\boldsymbol{\Sigma}_x$. Then $\boldsymbol{\Sigma}_{UT-\mathcal{H}}$ requires $2NI+1$ sigma points with the corresponding weights that are determined using $\mathbf{x}$, $\boldsymbol{\Sigma}_x$ and UT scaling parameters (α, β, κ) as shown in Equation (3.24).

Algorithm 1: Active calibration of the camera system

```
Input: $\boldsymbol{u}$, $\mathbf{x}$, $\boldsymbol{\Sigma}_{\boldsymbol{u}}$, $\boldsymbol{\Sigma}_x$
Output: $\widehat{\boldsymbol{\vartheta}}_{\mathcal{H}}$, $\widehat{\boldsymbol{\Sigma}_{\mathcal{H}}}$
function Active Calibration
    $Current\_Images \leftarrow [1]$
    for $i = 2$ to $I$ do
        $Current\_Images \leftarrow$ Append($i$)
        $\boldsymbol{\vartheta}_{\mathcal{H}} \leftarrow$ NonLinearCalibration($\boldsymbol{u} - f(\mathbf{x}, \boldsymbol{\vartheta}_{\mathcal{H}})$)
        $\mathbf{J}_{\mathcal{H}} \leftarrow \frac{\partial f}{\partial \boldsymbol{\vartheta}_{\mathcal{H}}}$
        $\boldsymbol{\Sigma}_{\mathcal{H}} \leftarrow (\mathbf{J}_{\mathcal{H}}^{\top} \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} \mathbf{J}_{\mathcal{H}})^{-1} + UT(\mathbf{x}, \boldsymbol{\Sigma}_x)$
        $\mathbf{J}_x \leftarrow \frac{\partial f}{\partial \mathbf{x}}$
        $\boldsymbol{\Sigma}_{x_u} \leftarrow \mathbf{J}_x^{\top} \boldsymbol{\Sigma}_x \mathbf{J}_x$
        $\boldsymbol{\Sigma}_T \leftarrow \boldsymbol{\Sigma}_{x_u} + \boldsymbol{\Sigma}_{\boldsymbol{u}}$
        $\mathbf{J} \leftarrow \mathbf{J}_{\mathcal{H}}^{\top} \boldsymbol{\Sigma}_T^{-1} \mathbf{J}_{\mathcal{H}}$
        $(\lambda_1 ... \lambda_{11}) \leftarrow$ EigenValues($\mathbf{J}$)
        if SumRelativeChange($\lambda_1 ... \lambda_{11}$) $> 0$ then
            $\widehat{\boldsymbol{\theta}}_{\mathcal{H}} \leftarrow \boldsymbol{\vartheta}_{\mathcal{H}}$
            $\widehat{\boldsymbol{\Sigma}_{\mathcal{H}}} \leftarrow \boldsymbol{\Sigma}_{\mathcal{H}}$
            if SteadyStateReached() then
                StopCalibration()
            end
        else
            $Current\_Images \leftarrow$ Remove($i$)
        end
    end
end
```

3.5 Active Calibration

Calibration should be performed with a variation of measurements that have sufficient parallax. We propose a technique to determine if new measurements will improve the estimation of the optimised parameters by looking at the observability. Firstly, the observation model for the line-scan camera can be defined as such:

$$\boldsymbol{u} = f(\boldsymbol{\vartheta}_{\mathcal{H}}, \mathbf{x}) + \nu_u + \nu_x \,. \tag{3.28}$$

The measurement noise of the observations ν_u is modelled as a zero-mean Gaussian with covariance $\boldsymbol{\Sigma}_{\boldsymbol{u}}$. The feature points $\mathbf{x}$ are prior parameters that introduce noise ν_x, which is also modelled as a zero-mean Gaussian with a covariance $\boldsymbol{\Sigma}_{x_u} \in \mathbb{R}^{2NI \times 2NI}$. This covariance is propagated from $\boldsymbol{\Sigma}_x$ by calculating the Jacobian $\mathbf{J}_x$ of the function f in Equation (3.18) with respect to $\mathbf{x}$. The observability analysis for an optimisation formulated as a non-linear least squares, can be performed by constructing the Fisher information matrix (FIM) [51][52]:

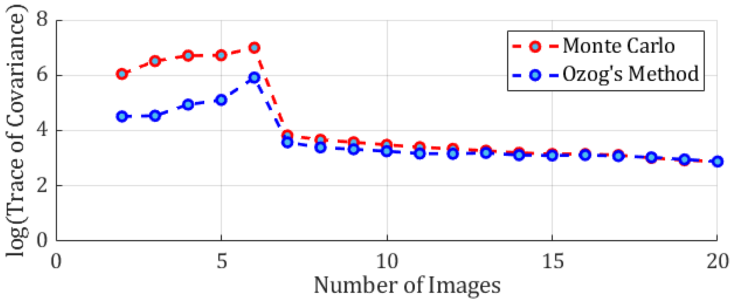
$$\mathbf{J}_{FIM} = \mathbf{J}_{\mathcal{H}}^{\top} \boldsymbol{\Sigma}_T^{-1} \mathbf{J}_{\mathcal{H}} \,. \tag{3.29}$$

$\mathbf{J}_{FIM} \in \mathbb{R}^{11 \times 11}$ and $\boldsymbol{\Sigma}_T \in \mathbb{R}^{2NI \times 2NI}$ is defined as $\boldsymbol{\Sigma}_T = \boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{\Sigma}_{x_u}$. The calibration is observable if $\mathbf{J}_{FIM}$ is full rank which ensures the Cramér-Rao lower bound exists [53][54]. The eigenvalues $(\lambda_1 ... \lambda_{11})$ of $\mathbf{J}_{FIM}$ generally relate to the observability of each parameter in $\boldsymbol{\theta}_l$. Therefore, by ensuring that $(\lambda_1 ... \lambda_{11})$ are large, the lower bound of the uncertainty for each parameter will be small. This concept is used to evaluate new measurements by calculating a metric of the sum of relative change between the current and previous $(\lambda_1 ... \lambda_{11})$. By only keeping measurements where the relative sum is positive, we ensure improved individual parameter observability. This metric also provides the capability of terminating the calibration once a steady-state is achieved. The proposed algorithm is described in Algorithm 1.

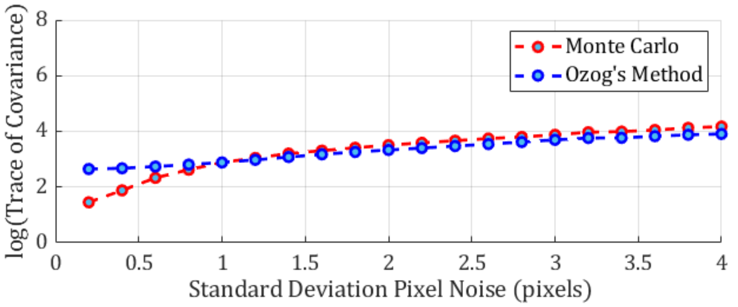
3.6 Experiments

3.6.1 Validation of Uncertainty Propagation

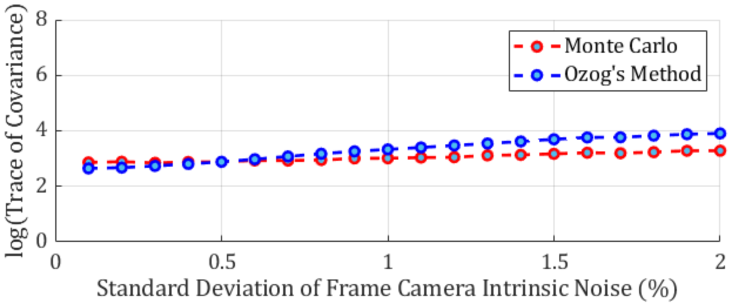
Experiments in simulation were carried out to confirm if the uncertainty $\boldsymbol{\Sigma}_{\mathcal{H}}$ was correctly propagated from $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ and $\boldsymbol{\Sigma}_{\mathcal{F}_K}$ using Ozog's method. A Monte Carlo simulation was performed with added pixel noise of STD σ_u and added noise in $\boldsymbol{\vartheta}_{\mathcal{F}_K}$. The trace of $\boldsymbol{\Sigma}_{\mathcal{H}}$



(a) Changing number of images



(b) Changing pixel image noise



(c) Changing frame camera intrinsic noise

Figure 3.8: $\mathbf{\Sigma}_{\mathcal{H}}$ estimated using Ozog's method was compared to a Monte Carlo simulation of 1000 runs. In (a), the trace is compared to a changing number of images, with fixed noises $\mathbf{\Sigma}_{\boldsymbol{u}} = 1$ and the standard deviation (STD) of the noise in $\boldsymbol{\vartheta}_{\mathcal{F}_K}$ is 0.3% of its mean value. In (b), $\mathbf{\Sigma}_{\boldsymbol{u}}$ was varied while the STD of the noise in $\boldsymbol{\vartheta}_{\mathcal{F}_K}$ is fixed. In (c), the STD of the noise in $\boldsymbol{\vartheta}_{\mathcal{F}_K}$ is varied with fixed pixel noise.

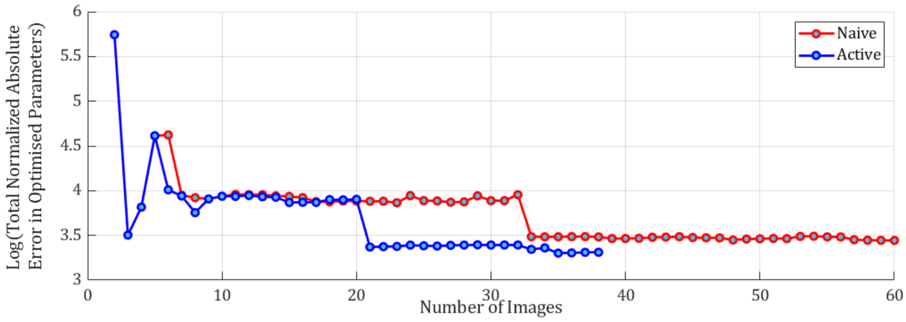
shows the results for changing number of images, changing $\mathbf{\Sigma}_{\boldsymbol{u}}$, and changing added noise in $\boldsymbol{\vartheta}_{\mathcal{F}_K}$, which can be seen in Figure 3.8a, Figure 3.8b and Figure 3.8c respectively. The estimated uncertainty by Ozog's method is valid for both changing noises and number of images. Large differences between the plots are a collective result of ignoring higher-order-terms in Equation (3.23) during the error propagation steps. Figure 3.8a also shows that the uncertainty is high in the beginning likely due to insufficient measurements for calibration.

3.6.2 Active Calibration in Simulation

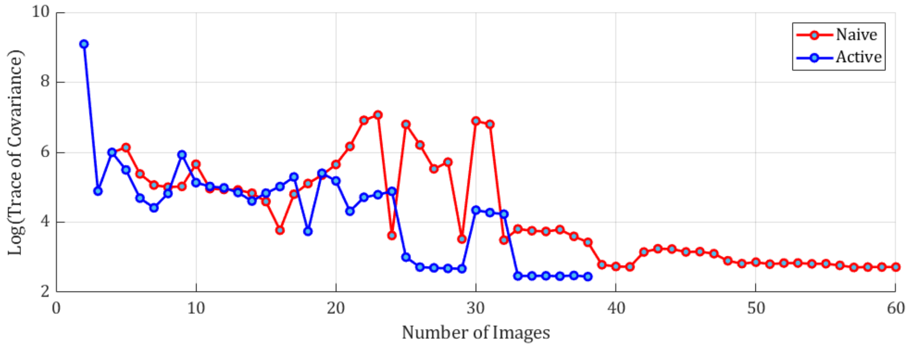
Experiments in simulation were performed using 60 poses of the pattern where ground truth was available. The algorithm was compared to a naive approach which calibrated using all images. The results for the total normalised absolute errors are shown in Figure 3.9a, the trace of $\mathbf{\Sigma}_{\mathcal{H}}$ is shown in Figure 3.9b, the metric to determine if new measurements from an image improve observability of the calibration as per Algorithm 1 is shown in Figure 3.9c, and absolute errors in the optimized translation, rotation, and intrinsic parameters are shown in Figure 3.10a, Figure 3.10b, and Figure 3.10c respectively. The algorithm chose the best 38 images out of the 60, where the last image was the 54^{th} at which the metric reached steady state and terminated the calibration as shown in Figure 3.9c. When comparing the results for the final optimized parameters, the active run showed a 26.4% drop in the total normalized error and a 46.8% drop in the trace of $\mathbf{\Sigma}_{\mathcal{H}}$. The time to complete the calibration was 76s for active and 19s for naive on an Intel Core i7-8650U four-core processor in MATLAB 2020B.

3.6.3 Active Calibration of Real Camera System

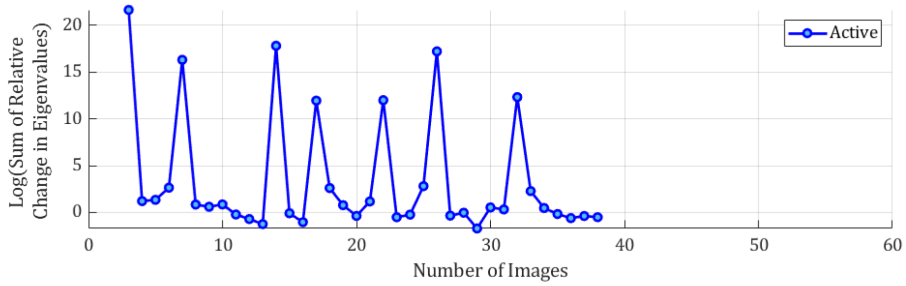
Active calibration of the camera system shown in Figure 3.1 was performed. Naive and active runs were compared using 40 poses of the pattern. The Figure 3.11a shows the trace of $\mathbf{\Sigma}_{\mathcal{H}}$ for both runs. The algorithm used the best 22 images from the 40 available. It did not reach a steady state so all 40 images were evaluated. When comparing the results for the final optimised parameters, the active run showed a 35.4% drop in the trace of $\mathbf{\Sigma}_{\mathcal{H}}$. The active calibration parameters were then used to transform and project a hyperspectral view-line onto an image of the pattern as shown in Figure 3.11b. The uncertainty in camera parameters is used to calculate the 1-σ error boundary of the pixel coordinates. The time to complete the calibration was 52s for active and 5s for naive on an Intel Core i7-8650U four-core processor in MATLAB 2020B.



(a) Total normalised absolute error in simulation

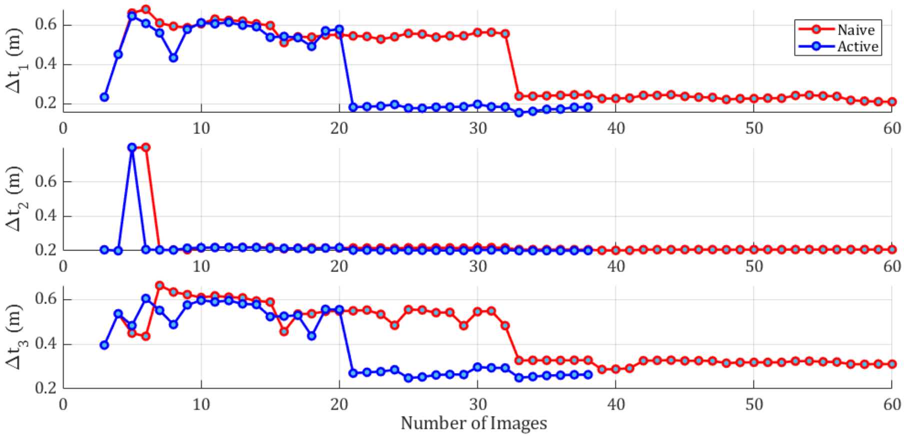


(b) Trace of covariance in simulation

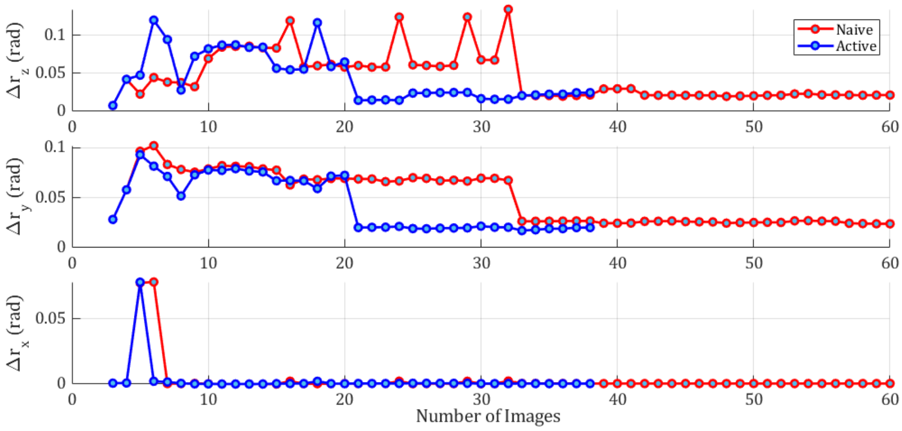


(c) Sum of relative change in eigenvalues in simulation

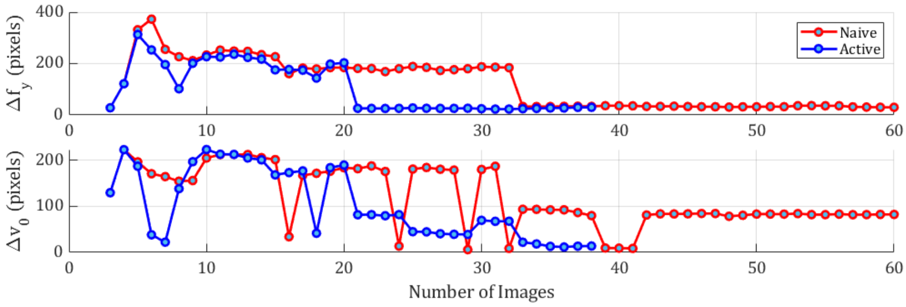
Figure 3.9: Active and naive runs were compared in simulation. The total normalised absolute errors in $\boldsymbol{\theta}_l$ is shown in (a), the trace of $\boldsymbol{\Sigma}_{\mathcal{H}}$ is shown in (b), and the sum of relative chance in eigenvalues, which is a metric used in the active calibration algorithm, is shown in (c). For both runs, $\sigma_u = 1$ and the STD of the noise in $\boldsymbol{\vartheta}_{\mathcal{F}_K}$ is 0.3% of its mean value. Note that the first calibration result has been omitted due to significantly large error.



(a) Absolute errors in translation parameters in simulation

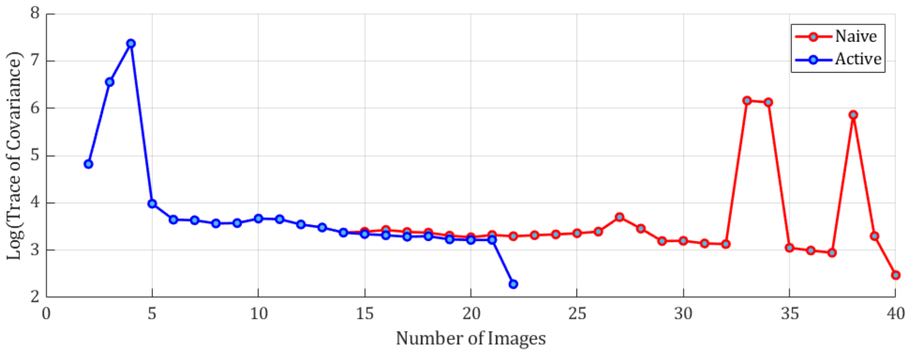


(b) Absolute errors in rotation parameters in simulation

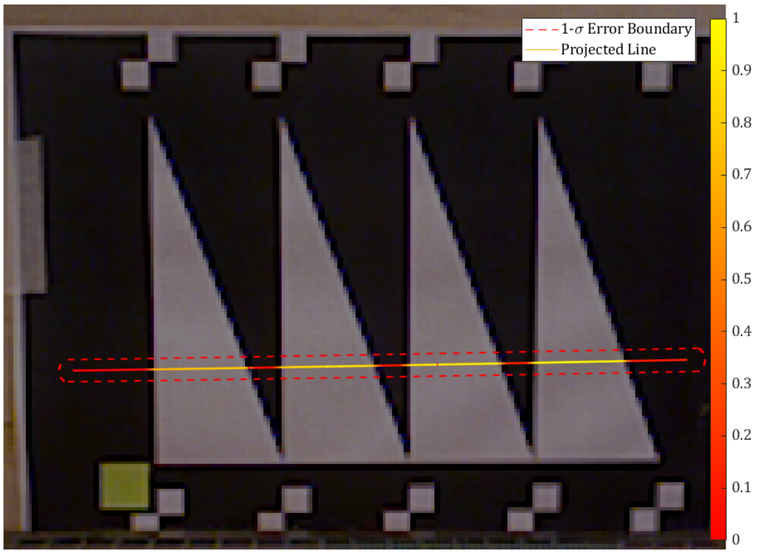


(c) Absolute errors in intrinsic parameters in simulation

Figure 3.10: Active and naive runs were compared in simulation. The absolute errors in t_1, t_2, and t_3 are shown in (a), absolute errors in r_z, r_y, and r_x are shown in (b), and absolute errors in f_y and v_0 are shown in (c). For both runs, $\sigma_u = 1$ and the STD of the noise in $\boldsymbol{\vartheta}_{\mathcal{F}_K}$ is 0.3% of its mean value. Note that the first calibration result has been omitted due to significantly large error.



(a) Trace of covariance in hardware experiments



(b) Projected spectral view-line on pattern

Figure 3.11: Active and naive runs were performed on a real camera system. The trace of $\mathbf{\Sigma}_{\mathcal{H}}$ is shown in (a). For both runs, $\sigma_u = 1$ and the STD of the noise in $\boldsymbol{\vartheta}_{\mathcal{F}_K}$ was retrieved from MATLAB's Camera Calibration toolbox. In (b), a hyperspectral view-line was projected to a frame camera image with its 1-σ error boundary.

3.6.4 Discussion of Calibration Experiments

The results of the experiments show an advantage to using the active calibration algorithm, even though it is comparatively slower due to its iterative nature. The poor results of the naive runs can be attributed to calibrating using measurements with insufficient parallax. The error plots in Figure 3.10a, Figure 3.10b and Figure 3.10c are roughly reflected by the total error evaluation metric in Figure 3.9c. This evaluation metric gave equal weighting to all parameters, therefore, large variations in some parameters, such as the distortion coefficients, greatly influenced the plot.

The metric used in the active calibration algorithm only considers the sign of the relative change in eigenvalue sum. If the metric is positive, the latest measurement has likely increased parameter observability, but the magnitude of the metric is disregarded. In terms of early termination, there may be a risk of ignoring later observations which may improve parameter estimation. The algorithm focuses on keeping measurements that increase the metric until the eigenvalues stabilise, at this point it assumes that no further measurements will improve the results. The algorithm may also terminate too early, using fewer than 10 images, which generally resulted in higher uncertainties as shown in Figure 3.8a, Figure 3.9b, and Figure 3.11a. Therefore, a minimum number of images should be set.

3.7 Conclusions

A probabilistic calibration for a line-scan and frame camera system is proposed, which propagated noise in the pixel measurements to the optimised parameters. This method of uncertainty propagation provided the groundwork for an observability analysis of the optimisation-based calibration, which was then exploited to evaluate new measurements based on their contribution towards improving parameter observability. The algorithm of active calibration demonstrated that not all new measurements were beneficial. Ignoring such measurements resulted in calibration parameters closer to ground truth with lower uncertainties. The observability metric was also used to terminate the calibration once reaching a steady state value.

In order to capture hyperspectral images of an object, an appropriate light source must be used that provides the necessary radiant energy in the spectral bands of the line-scan camera. When the light source is close to the object, the intensity of the incident light and its directions will vary over the object's surface. The variation in the incident light will affect the captured hyperspectral image. The next chapter will look to spatially model the radiant energy distributed from a real light source with the calibrated camera system, particularly its inherent asymmetric distribution. The light source model could then be used to infer the incident radiant energy and its direction, with respect to the source, at any point in space.

Chapter 4

Probabilistic Light Source Modelling

This chapter focuses on spatially modelling a light source in a near-field environment that is used with the hyperspectral camera, using a realistic model that is data-driven to capture asymmetric details in the distributed radiant flux. This spatial light source model allows for the estimate of the irradiance over the surface of an object being observed by the hyperspectral camera. It captures both the variation in irradiance and the direction of light rays over the surface, which can aid in improving the estimation in the intrinsic material properties of the surface, such as diffuse reflectance, that do not depend on the light source.

Image formation relies on reflected light from objects which originate from external light sources. If we consider hyperspectral cameras, they require light sources that can illuminate an object in all the measurable spectral bands of the camera. The variation of the incident light on a surface depends on the relative distance between the light source and surface. When this distance is small, the incident radiant flux and the light ray directions vary considerably. This is caused by the geometry of the light source and its asymmetric distribution of radiant flux. Existing work with hyperspectral cameras generally assume no variation of incident light, and existing light source modelling enforce strict properties on the distribution, such as symmetry.

Understanding the behaviour of light sources using models aids in understanding the interaction of their emitted light with the environment. For example, in computer graphics the use of light source models in rendering applications give more realistic appearances to objects in a scene [55][56][57][58]. If the light source is considered to be very far away (i.e. at infinity), regardless of its geometry, the incident light rays on an object can be considered to be parallel. If the geometry of the source is not important, than the incident radiant flux at any point on an object will only vary across the spectral bands. In this work, hypercubes are captured where the light source is close to the object (referred to as near-field), which means that the geometry of the source is important as both the incident ray directions and radiant flux will vary spatially.

A light source model should achieve the following: have a specified location, capture the geometry of the source, define the direction of the emitted rays and their radiant intensity, which is the radiant flux in a given solid angle direction (further details in Section 2.4.4). The simplest geometry for a source is a point. With this geometry, light source models assume a parametric distribution profile of radiant intensity about the source, which generally enforce properties of symmetry or isotropy in the radiant intensity profile to simplify the model. By capturing measurements of known targets using cameras, the parametric models can be fit to the measurements. In reality, the radiant intensity profile of real light sources do not have properties of symmetry or isotropy due to their internal components (i.e. filament), and inherent design favouring certain directions (i.e. down-lights). The parametric models alone cannot be altered to predict asymmetric features. Given that we capture discrete measurements of the radiant intensity profile in space, using images from the camera, the measurements should be used to define the inherited asymmetries. It is possible to augment the parametric light source model with a data-driven model, that takes in the discrete measurements from the hyperspectral camera, and enforces the continuous nature of the radiant intensity profile. A Gaussian process (GP) model is appropriate for such purposes and has a probabilistic component, that could capture the uncertainty in the light source model especially in areas with fewer measurements.

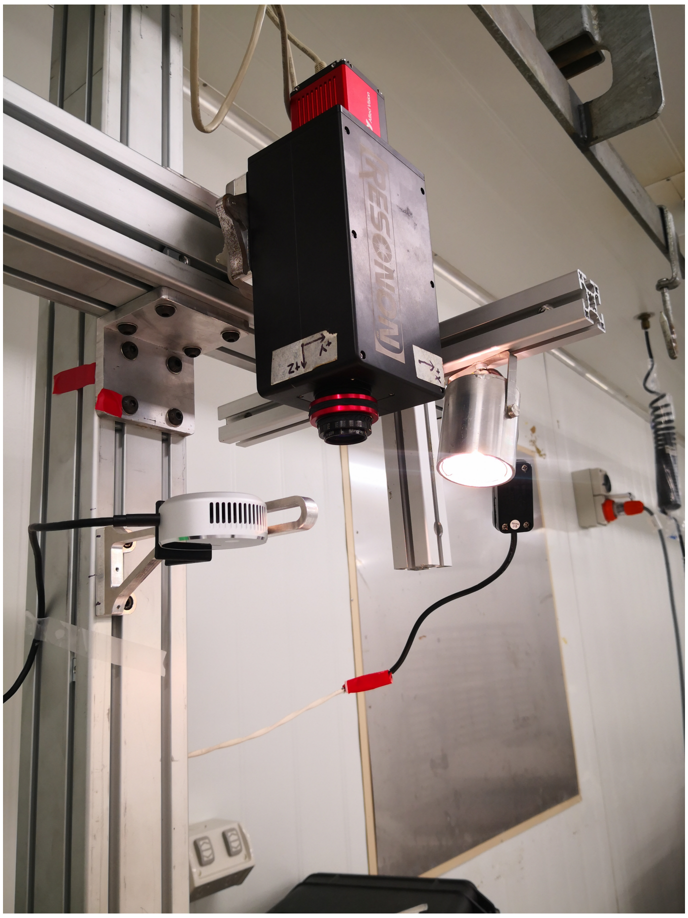


Figure 4.1: Rigidly mounted camera system comprised of a Resonon Pika NIR-320 hyperspectral line-scan camera, an Intel Realsense L515 RGB-D camera, and a single halogen light source.

4.1 Literature Review

To build such a model, the behaviour of light must first be established to describe its properties and interactions with matter. There are various light models that range from ray optics, which describes light as rays in the direction of energy flow, and Quantum optics, where the interactions occur at the quantum scale with the elementary particle of the photon, which exhibits both wave and particle like behaviour. Further details of these light models are discussed in Section 2.3.2. As the light models increase in complexity, more properties of light can be accurately described. For computer vision and computer graphics, especially when working with cameras, ray optics is sufficient to capture the distribution of radiant flux, model the interactions with objects and their materials at both a macro and micro scale, and describe the process of image formation (discussed further in Section 2.1).

The light source model establishes how the radiant flux is distributed spatially in particular spectral dimensions. There are three major components to modelling a light source: the geometry of the source, the radiant properties of the geometry, and the manner in which radiant power is distributed [59]. The complexity of the light source model depends on the application. The geometry of the source describes the physical size, shape, and location. The source can be treated as rays emanating at an infinite distance, a point, or a 3D model of the physical components from which the rays emanate inside the bulb. Most applications do not require the level of detail present in the last option, as they are more interested in how the emitted light interacts with the environment, not in the source itself. One study that used a detailed 3D light source model, compared the radiant intensity produced when the filament inside the bulb was modelled as a helix versus a cylinder. Through Monte Carlo simulations, it was shown that the less complex geometry of the cylinder had an almost identical profile as the helix [59].

When assuming a point source model, how the light is distributed from the source is defined by its radiant intensity distribution (RID). When a light source emits equally in all directions, the source is referred to as being isotropic. The simpler model would assume the light source be isotropic, but real light sources are highly directional, so inaccuracies in lighting and shadows can occur [60][61]. If the light source is considered a distant source at infinite, then all rays are parallel and are assumed to arrive with equal intensity; the most common example of such is the sun. These types of lighting models are generally used for improving the rendered appearance of a 3D modelled environment. This is typically carried out using direct path-tracing techniques, where the path of light rays that originate from an unknown distant light source are triangulated [56][55]. Starting from the camera, many light rays emanate from each pixel to the scene, where they interact by bouncing or reflecting off surfaces. These interactions are recorded, and eventually, most rays should coalesce into the general direction of the light source. Furthermore, through an optimisation framework, material properties such as roughness can be obtained, which can help to further improve the appearance of rendered scenes [55].

The light source modelling procedure can be treated as a calibration using either spheres [62][63][64] or planar [61][65][66][60] targets. Spheres are simple to parameterise, provide a smooth positively curved surface without discontinuities from which surface normals can be easily computed. This makes them ideal for triangulating the location of a source with respect to a camera. However, the changing surface normals will affect the radiant flux that is received at different areas of the sphere, which makes the RID estimation difficult. Using a planar target which has constant surface normals, the RID can be estimated by knowing the reflectance properties of the target material[60][61].

4.1.1 Sphere Based Calibration

If a light source is treated as a nearby point and is captured using a camera, then its locations can be defined with respect to the camera's coordinate system. A technique to estimate this location is through the use of calibration spheres with frame cameras. One of the earliest works captured images of three reflective spheres, with known relative positions and radii, and used the detected highlights on the surface of the spheres to triangulate the location of a single light source [62]. This method is known as forward calibration and is the most straightforward approach, as it starts from the camera, projects rays into 3D space and determines the nearest intersection of the rays. In comparison, backward calibration starts from the unknown 3D location of the source, follows a ray that reflects off the sphere that is captured by the camera, which projects it into the image space. The unknown location of the source is determined through a non-linear least squares of the reprojection error. The backward calibration was shown to be more accurate than the forward calibration when estimating the location of two light sources using seven spheres [63].

There are several advantages to using spherical geometries for estimating the location of the source. For the size of the sphere only one parameter of its radius is required, while the location is constrained through a single central 3D point. The sphere's surface is essential as it has smooth positive curvature without any discontinuities, and the surface normal can be easily determined at any point. Forward calibration of a light source was estimated using a hollow sphere, where the reflections on both the outer and inner surfaces were detected and used to triangulate the location of the source [67]. A more realistic light source estimation problem was proposed, which contained point, directional and ambient sources [64]. In this approach, a single image of two known Lambertian spheres was captured. A pixel difference between the spheres in the image removed lighting from the ambient and directional sources, from which the point sources could be estimated through highlights. Then removing the effects of the points source, the ambient and directional sources could be estimated. This method relied on the assumption that the spheres had a Lambertian surface, as the camera's viewing direction did not affect the lighting that was seen on the surface of the spheres.

4.1.2 Planar Based Calibration

Calibrations with spherical targets generally only consider isotropic light sources, as the varying curvature of the sphere makes it difficult to capture the variation of radiant intensity from the source. The non-isotropic behaviour of real light sources is defined through a RID, which describes the directional variation of the radiant intensity about the light source using a profile. The RID effectively adds parameters to encode the shape of the distribution profile [61][65][66][60]. Furthermore, the RID curves can have a multitude of shapes that can be modelled as polynomials, where the coefficients of the polynomial terms would need to be estimated when modelling the light source[60]. Finally, it should be noted that an isotropic source would have a circular-shaped RID.

Planar based calibrations can be used to capture non-isotropic RID using images of a Lambertian target, which is assumed to only reflect light in a diffuse manner. The identification of both highlights, and symmetrical properties between the light source, camera, and target, can be used to formulate an optimisation [60]. The target can be made from a non-Lambertian material, where the amount of reflected light changes depending on the viewing angle, by incorporating a bidirectional reflectance distribution function (BRDF) into the optimisation, which describes the amount of reflected light intensity on the surface, given directions of both the incident light from the source and outgoing light towards the camera [60]. Further details of a BRDF can be found in Section 2.5.1. This calibration has been shown to improve 3D construction results due to accurately capturing the lighting conditions [65]. Others have approached the calibration through a backwards method, where they start from the light source, that illuminates a Lambertian plane and capture an image of the plane using a camera. The unknown parameters introduced by the plane and the camera are shown to be multiplicative, and are combined into a single unknown parameter which is estimated alongside the light source parameters [61].

4.1.3 Gaussian Processes of Fields

Both spherical and planar based calibrations model the light source using parametric models which assume a fixed profile and are not able to capture asymmetries that exist in the distribution. A probabilistic non-parametric method of building regression models is through a GP. A GP is defined by its mean and covariance functions, which can be thought of as an infinite collection of random variables that are all jointly Gaussian distributed [68]. Therefore, GP models show continuous behaviour.

The radiant intensity of a light source can be thought of as an intensity field across space. Rather than train a parametric model that estimates the behaviour of a field, a GP, which already behaves like a field, could be used to model the radiant intensity. Similar work has been shown with magnetic fields, where the squared exponential (SE) covariance function is reshaped to fit the properties of a magnetic field having zero divergence and curl [69]. Modelling was done by capturing the magnetic field strength using a three-axis magnetometer in discrete locations. An alternative application used GPs to model the laminar flow of ocean currents that were planar, time-invariant and incompressible [70]. The property of incompressibility is realised by ensuring the divergence of the flow field is zero. Again, the covariance function is reshaped to fit this property, and when compared with the standard unaltered version, it showed higher accuracy in modelling the flow field.

4.2 Cameras and Light Source

We now formally introduce the light source modelling problem with the calibrated camera system. Let us consider a single non-isotropic near-field point light source $\mathcal{L}$ illuminating

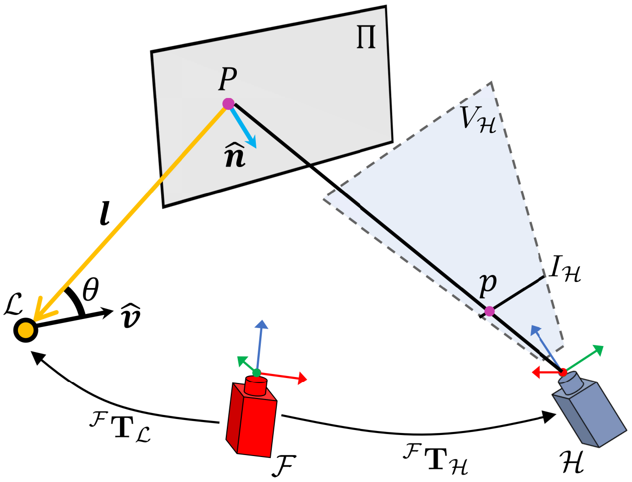


Figure 4.2: The non-isotropic point light source $\mathcal{L}$ with the principal direction vector $\hat{\boldsymbol{v}}$ that has maximum radiant intensity. The calibrated camera system comprised of the red, green, and blue colour channels (RGB) frame camera $\mathcal{F}$ and line-scan hyperspectral camera $\mathcal{H}$, that are rigidly mounted by the homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{H}}$. $\mathcal{L}$ is rigidly mounted to the hyperspectral camera by the known homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{L}}$. The light vector $\boldsymbol{l}$ carries radiant flux to P, which is then reflected off the Lambertian surface Π and measured by the hyperspectral camera at p on the image $I_{\mathcal{H}}$.

a flat homogeneous Lambertian surface Π, which is captured by a multi-modal camera system consisting of a line-scan hyperspectral camera $\mathcal{H}$ and a 2D RGB frame camera $\mathcal{F}$ as shown in Figure 4.2. $I_{\mathcal{H}}$ is the line image of the hyperspectral camera which is captured on its view-plane $V_{\mathcal{H}}$. The camera system is fully calibrated as described in Chapter 3. The fixed homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{H}} \in \mathbb{R}^{4\times 4}$ is of the hyperspectral camera with respect to the frame camera which comes from the calibrated extrinsic parameters. The light source is rigidly attached to the frame camera by the homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{L}} \in \mathbb{R}^{4\times 4}$ which is assumed to be known. The direction vector $\hat{\boldsymbol{v}} \in \mathbb{R}^3$ represents the principal direction of the light source which is also inline with its Z-axis. The light vector $\boldsymbol{l} \in \mathbb{R}^3$ represents the ray that is emitted by the light source and strikes the 3D point P on Π with the surface normal $\hat{\boldsymbol{n}} \in \mathbb{R}^3$. $\boldsymbol{l}$ is defined as originating from P in the direction towards the light source to ensure the dot product between $\boldsymbol{l}$ and $\hat{\boldsymbol{n}}$ is positive [61]. The radiance reflected from P in the direction of the hyperspectral camera is measured by the pixel p on $I_{\mathcal{H}}$.

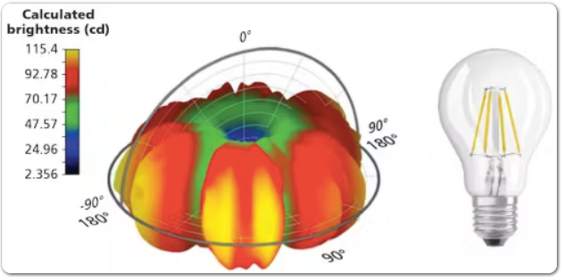


Figure 4.3: An example of a 3D radiant intensity distribution (RID) for a real bulb [74]. The measured brightness is noted in candela (cd) which is a unit of radiant intensity specifically for visible light.

4.3 Light Source Modelling

The modelling process starts from the light source and its distributed radiant flux (with units $[W]$) about its geometry, which is described by a parametric model that introduces unknown parameters. The light arrives at a known planar surface, and is reflect in the direction of the hyperspectral camera, where the radiant energy is converted to a pixel value. Between the light source and the surface, radiometric conversions occur which account for radiant flux that is directional, surface area dependent, or both. Further details of the radiometric units can be found in Section 2.4. This section ends with the parametric solution to the light source model.

4.3.1 Radiant Intensity Distribution

For real light sources, the RID is a 3D distribution that depends on many factors such as the shape of the illuminating components (typically metallic filaments), and their placement and direction in the bulb [59][71]. A point ${}^{\mathcal{L}}\boldsymbol{P}$, which is a 3D point with respect to the light source, can be defined using spherical coordinates (r, θ, φ), where r is the Euclidean distance, θ is the zenith angle, and φ is the azimuth angle (further details of this coordinate system can be found in Section 2.4.1). To acquire an accurate 3D RID as shown in Figure 4.3, measurements of the the light source would need to be captured using a goniophotometer [72][73] over both spherical angles. To our knowledge, there are no such publicly available datasets. However, measurements for a single plane were available [72] and are shown in Figure 4.5.

Although, RID of a real source is asymmetric, as highlighted in Figure 4.5, it is commonly modelled as a parametric radially symmetric function [60][61][66], because it is impossible to predict and encode the asymmetrical directions prior to fitting the parametric model. Therefore, we will also assume that the RID of the light source is radially symmetric, where the radiant intensity of the light source will only vary across the θ angle. One can imagine that for a fixed θ, there exists a ring of equal radiant intensity as you vary the φ. Therefore, the coordinate system at the light source can be simplified to only two components (r, θ). The conversion of ${}^{\mathcal{L}}\boldsymbol{P}[X, Y, Z]$ from Cartesian coordinates to this new 2D coordinate system is defined as:

$$\begin{aligned} r &= \|\boldsymbol{l}\| \\ \theta &= arctan2\left(Y, \sqrt{X^2 + Z^2}\right) . \end{aligned} \tag{4.1}$$

A cosine relationship can be used to express a symmetric RID as follows:

$$\begin{gathered} \Phi(\theta) = \Phi_0 \cos^{\mu} \theta \, . \\ \textbf{Unknown Parameters: } \Phi_0, \mu \quad \textbf{Units: } Wsr^{-1} \end{gathered} \tag{4.2}$$

where Φ is the radiant intensity $[Wsr^{-1}]$, Φ_0 is a maximum radiant intensity parameter found at the source in the direction of $\hat{\boldsymbol{v}}$, and μ describes the directional "laser-like" behaviour of the light source. Both Φ_0 and μ must be greater than 0. The effect of the parameters on the cosine model can be seen in Figure 4.4.

4.3.2 Attenuation

Point sources experience a fall-off in radiant flux with increasing distance which is known as attenuation. Assuming that the light rays are travelling in a vacuum, where there is no interacting medium, the attenuation is only due to the geometry of the source as the rays are emitted in various directions[1]. If we consider a fixed solid angle about the source, the area subtended by the angle will increase proportionally to the distance squared. Therefore, the radiant flux carried by the rays will be distributed over a larger area which

---

[1] There would be no attenuation of radiant intensity if the rays are emitted parallel to each other. This could be achieved by a planar light source or a source located at infinity.

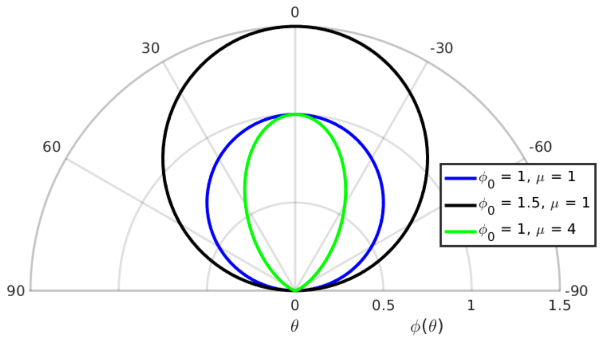


Figure 4.4: A polar plot of the cosine model used to parameterise a symmetrical radiant intensity distribution (RID) is shown. The parameters Φ_0 and μ are the maximum radiant intensity and directional "laser-like" behaviour of the light source. Plots of three different combinations of the parameters are shown. If the blue is the default, then black line is the model for a larger Φ_0, and green line is the model for a larger μ.

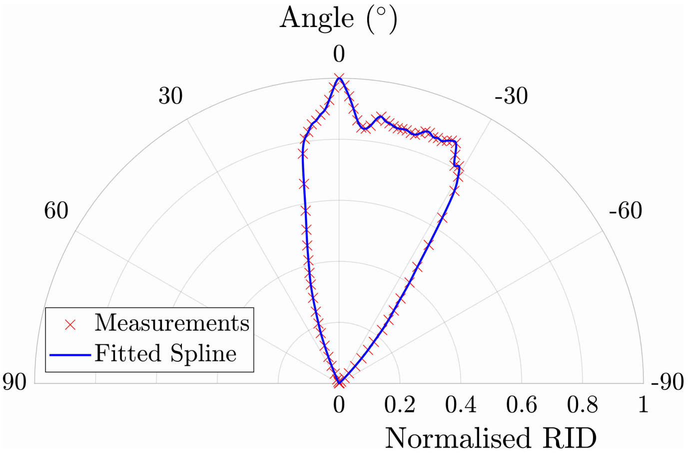


Figure 4.5: Measured RID on the inclination plane of an iGlobe 5W GU10 down-light bulb provided by [72]. A cubic spline has been fit to the measurements and is used in our simulator.

effectively reduces radiant intensity. Typically, this can be modelled by the inverse square law, but it has drawbacks of a singularity as the distance approaches 0, the radiant intensity decaying too rapidly and the shape of the decaying function can not be altered [75]. An alternative form is used in computer graphics which resolves these issues by changing the geometry of the source to a disk with a radius r_d [76][77]. The attenuation can then be modelled as:

$$f_{att}(r) = \frac{1}{\left(\frac{r}{r_d}+1\right)^2}\,. \tag{4.3}$$

$$\textbf{Unknown Parameters: } r_d \quad \textbf{Units: } 1$$

Note that the attenuation is only between the source and the point on the illuminated surface. The attenuation between the surface and the viewer does not need to be considered because the apparent area of the surface, as seen by the viewer, decreases by the same factor as the radiant intensity. Therefore, no attenuation occurs.

4.3.3 Radiant Intensity Model

Finally, to model the radiant intensity field of the light source we combine Equation (4.2), Equation (4.3), and $\hat{\boldsymbol{l}}$ which is the direction vector of $\boldsymbol{l}$ as follows:

$$\boldsymbol{s}(r,\theta) = -\Phi(\theta)\cdot f_{att}(r)\cdot \hat{\boldsymbol{l}}\,. \tag{4.4}$$

$$\textbf{Unknown Parameters: } \Phi_0, \mu, r_d \quad \textbf{Units: } Wsr^{-1}$$

The radiant intensity $\boldsymbol{s}$, with units $[Wsr^{-1}]$, is returned as a vector along the direction $-\hat{\boldsymbol{l}}$ and will arrive at the surface point P from a single direction. The irradiance E, with units $[Wm^{-2}]$, at P is the sum of the radiant flux that arrives at the surface from all directions per unit surface area (see Section 2.4.3 for further details). By only considering the light from the light source, the incident radiant flux at P is only due to $\boldsymbol{s}$. Given that a surface is defined by $\hat{\boldsymbol{n}}$, if $\hat{\boldsymbol{n}}$ is inline with $\hat{\boldsymbol{l}}$, which can be formally described as $\hat{\boldsymbol{n}} \bullet \hat{\boldsymbol{l}} = 0$, then the norm of $\boldsymbol{s}$ will return the E. A planar slice of the irradiance for a simulated light source is shown in Figure 4.6.

4.3.4 Measuring Radiance with Hyperspectral Cameras

We are only interested in the radiant flux that has been reflected off P in the direction of the hyperspectral camera. This quantity is known as the radiance L with units $[Wm^{-2}sr^{-1}]$, which in this case is the radiant flux that is reflected towards the hyperspectral camera per unit foreshortened area of the surface, per unit solid angle direction (further details

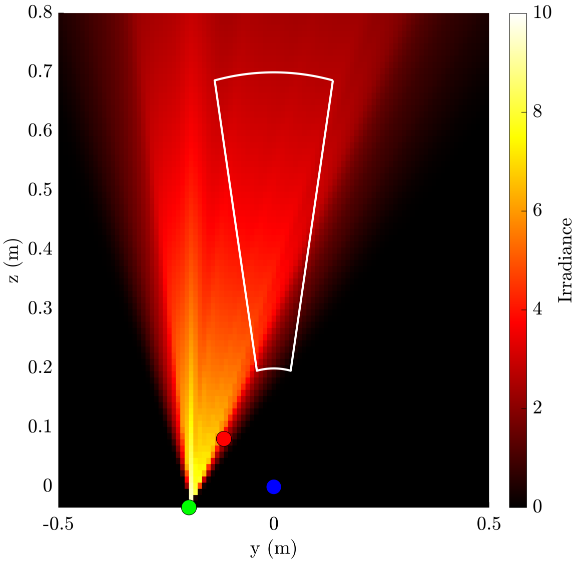


Figure 4.6: A planar slice of the irradiance as seen on the view-plane of the hyperspectral camera. The white outline shows the 2D viewing frustrum. The green, red and blue markers represent the location of the light source, frame camera, and hyperspectral camera respectively.

can be found in Section 2.4.5). The reflected radiance depends on the incident irradiance and the material properties at P. The amount of radiance that is reflected at P is defined by the BRDF, which is a function of the reflected radiance to the irradiance for a single combination of a incident and a reflected directions (further details can be found in Section 2.5.1). Given that P is on Π, which is of a Lambertian material, the reflected radiance is equal in all solid angle directions. Therefore, the reflected radiance headed towards the hyperspectral camera can be determined using the light transport equation (as defined in Section 2.5.2) of the Lambertian BRDF (as defined in Section 2.5.5) for all incident directions[2] as follows:

$$L(r,\theta) = -\frac{\rho}{\pi}\left(\boldsymbol{s}(r,\theta) \bullet \hat{\boldsymbol{n}}\right) \, , \qquad (4.5)$$

$$\textbf{Unknown Parameters: } \Phi_0, \mu, r_d \quad \textbf{Units: } Wm^{-2}sr^{-1}$$

where ρ is the reflectance at P. The radiance is measured by the hyperspectral camera through pixel p on $I_{\mathcal{H}}$. This measurement process involves a complex arrangement of lenses which guide light rays towards the sensor to accumulate and measure the radiant

---

[2] The general form of the outgoing radiance considers all incoming directions of light regardless of the light sources. In this work we only consider the light from a single source. Therefore, there is only a single incident direction.

energy. The sensor at p is exposed to the radiance for a fixed duration of time Δt resulting in:

$$H(p) = L(p)\Delta t \,, \tag{4.6}$$

where H is the exposure, with units $[J]$, which is the radiant energy that is received by the sensor at p. The exposure is then mapped to a pixel intensity reading through the camera response function (CRF) f_{crf}. The measurement model of the hyperspectral camera is then given by:

$$I(p) = f_{crf}(H(p)) + I_d(p) + \nu_p \,, \tag{4.7}$$

where I is the resulting pixel intensity reading. I_d is the dark current caused by thermal effects in the camera's sensor, which is measured by capturing an image with the lens covered. The measurement is assumed to be corrupted by zero-mean Gaussian noise ν_p with a variance of σ_p^2. The CRF is always increasing and monotonic, and is generally non-linear per frequency band, but recovering it is an ill-posed problem [78]. In this work, we have assumed that f_{crf} can be modelled as a linear function, which we empirically validated by capturing images of a diffuse surface using a Resonon Pika NIR-320 hyperspectral line-scan camera with a single light source. For the same setup, the normalised pixel intensity, with $I_d(p)$ subtracted, is plotted against changing exposure time of four different spectral bands as shown in Figure 4.7. The close-to-linear response for each of the spectral bands shows that the linear approximation of f_{crf} is valid as long as the pixel has not been saturated for the spectral band.

To build the light source model, measurements of the radiance are required. Given that the actual measurements are I and that f_{crf} is an unknown linear function, an approximation is made where we instead work with $\tilde{L}$ which encompasses L, f_{crf}, and Δt. The measurement model Equation (4.7) can then be simplified as follows:

$$I(p) \approx \tilde{L}(p) + I_d(p) + \nu_p \,. \tag{4.8}$$

This approximation only holds as long as Δt is fixed and that the f_{crf} of the hyperspectral camera is both linear and fixed.

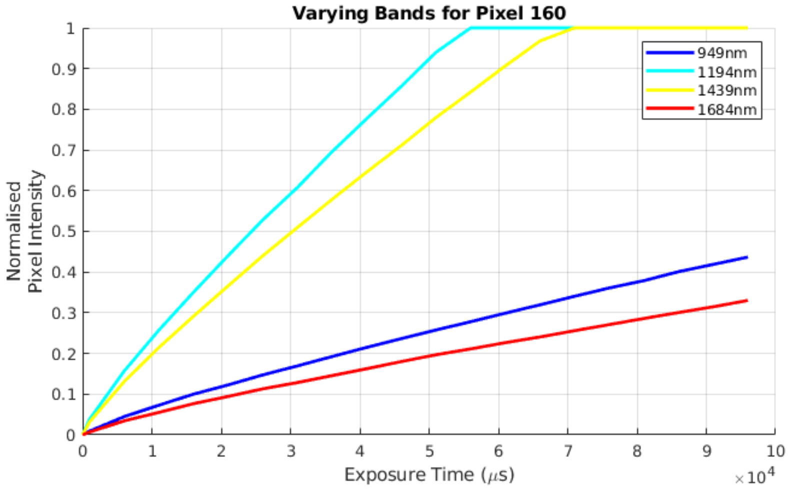


Figure 4.7: The normalised pixel intensity (with dark current subtracted), captured by the Resonon Pika NIR-320 hyperspectral line-scan camera of a diffuse surface illuminated by a single light source, against changing exposure times of four different spectral bands. The plots highlight the close-to-linear response of the camera response function (CRF) for each spectral band. The spectral bands 1194nm and 1439nm saturate at approximately $5.5\mu s$ and $7.0\mu s$ respectively in the experiment.

4.3.5 Least Squares Solution

The radiant intensity model for a single band is shown in Equation (4.4) that has the unknown parameters $\boldsymbol{\psi} = [\Phi_0, \mu, r_d]$. Recovering these parameters for the light source can be formulated as a non-linear least squares of Equation (4.5) as follows:

$$\widehat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}}{\operatorname{argmin}} \Big\| \sum_{r} \sum_{\theta} \Big(\tilde{L}(r,\theta) - f_L(\boldsymbol{\psi}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{l}}, \boldsymbol{\rho}, \hat{\boldsymbol{n}}, r, \theta) \Big) \Big\|^2, \tag{4.9}$$

where $\widehat{\boldsymbol{\psi}}$ is the optimised parameters and f_L is the nonlinear function of Equation (4.5). This problem can be solved by well-known algorithms such as Levenberg-Marquardt or trust-region methods. The solution is formulated for a single spectral band. To model multiple spectral bands, the optimisation would be repeated for the radiance measurements of each spectral band, and would result in a separate set of optimised parameters.

The parametric model for a non-isotropic disk source has been defined for a single spectral band, which includes the RID and attenuation functions that both introduce unknown parameters. By capturing a Lambertian surface with the hyperspectral camera, the radiance received by the camera is approximated using the pixel intensity, which is then used to fit parameter model as shown above. Now we consider the proposed data-driven GP light source model.

4.4 Gaussian Process Irradiance

Although we know that the RID is an asymmetric function, in Equation (4.2), it is still being modelled as a symmetric one. Correctly parameterising the asymmetry for a given source, however, is extremely challenging. We believe that the data should drive the asymmetry in the model, and due to the irradiance being continuous in nature, a GP is proposed to enhance the light source model presented above.

A GP can be seen as a distribution over functions and is used to model the irradiance of the light source as follows:

$$E(\boldsymbol{x}) \sim GP(m(\boldsymbol{x}), K(\boldsymbol{x}, \boldsymbol{x}')) \,. \tag{4.10}$$

The measurement vector $\boldsymbol{x} = [r, \theta]$ stacks the components of the points into a single vector. The points are defined in the coordinate system described in Equation (4.1). $K(\boldsymbol{x}, \boldsymbol{x}')$ is the covariance function which is chosen to be the standard SE as the space of $E(\boldsymbol{x})$ is smooth. $m(\boldsymbol{x})$ is the mean function and generally is assumed to be $\mathbf{0}$, where the model is fully captured by $K(\boldsymbol{x}, \boldsymbol{x}')$. In the literature, non-zero mean functions have been used to enforce known models in the otherwise purely data-driven approach, e.g. known shapes have been imposed through the mean function when fitting GPs on point-cloud data [79]. In a similar fashion, we propose a non-zero mean function to enforce the general profile of $E(\boldsymbol{x})$ by combining both Equation (4.2) and Equation (4.3) as such:

$$m(\boldsymbol{x}) = \Phi(\theta) \cdot f_{att}(\mathbf{r}) \,, \tag{4.11}$$

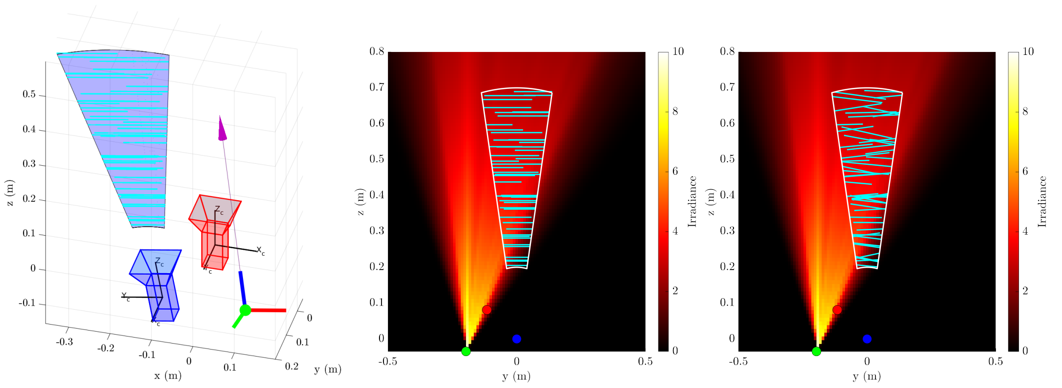
where this mean function is referred to as the “light-source-mean” and the unknown parameters $[\Phi_0, \mu, r_d]$ are now its hyper-parameters. The training measurements for $E(\boldsymbol{x})$ are found by first acquiring $\tilde{L}(\boldsymbol{x})$ from Equation (4.8), and then rearranging Equation (4.5) as follows:

$$E(\boldsymbol{x}) = \frac{\pi}{\rho\left(\hat{\boldsymbol{l}} \bullet \hat{\boldsymbol{n}}\right)} \tilde{L}(\boldsymbol{x})\,. \tag{4.12}$$

4.5 Experiments

4.5.1 Light Source Modelling in Simulation

A light simulator was created in MATLAB to test the GP light-source-mean model for a single spectral band as shown in Figure 4.8a. The simulator had a known frame and line-scan camera system, and an asymmetric RID disk source which sampled the measured RID in Figure 4.5. A flat planar Lambertian target of known reflectance value with a width of 0.2m, was randomly placed in the view-plane of the line-scan camera. A single experiment involved capturing 50 samples of the target where only the portion of the target inside the view-plane was measured and used for training. The measured training data was used to fit light source models for the methods of least squares regression and GPs with zero-mean, constant-mean and the proposed light-source-mean. In the first set of results, 50 experiments were carried out and averaged where the surface normal of the target is fixed and perpendicular to the XY plane of the line-scan camera as shown in Figure 4.8b. The predicted irradiance on the view-plane for each method and its absolute error to ground truth can be seen in Figure 4.9, and the distribution of absolute errors inside and outside of the view-plane were visualised using a violin plot as shown in Figure 4.11. In the second set of results, 50 experiments were carried out and averaged again, where the surface normal of the target is randomly offset to the hyperspectral camera's XY plane by an angle between -10° and 10° as shown in Figure 4.8c. The second experiment was used to test for the effects of Lambert's cosine law, where the effective irradiance arriving at the target varies due to the cosine of the angle between the surface normal and the light direction. Angle offsets of such nature can occur when collecting real data using a physical reflectance target. Once again, the predicted irradiance on the view-plane for each method and its absolute error to ground truth can be seen in Figure 4.10, and the distribution of absolute errors inside and outside of the view-plane were visualised using a violin plot as shown in Figure 4.12. The results across all experiments demonstrate that



(a) **(b)** **(c)**

Figure 4.8: The MATLAB non-isotropic asymmetric light simulator used for testing the light-source-mean model for a single spectral band. In (a), a view of the simulator is shown where the red coloured camera object is the frame camera, blue coloured camera object is the line-scan hyperspectral camera, and the green sphere is the light source. The blue area is the 2D viewing frustum of the line-scan hyperspectral camera, where the turquoise coloured lines represent separate measurements of the irradiance in space using a known flat Lambertian reflectance target with a width of 0.2m. In (b) and (c), the light source, frame and line-scan cameras are denoted in green, red, and blue dots, respectively. The 2D frustum of the hyperspectral camera is displayed in white, where again the turquoise coloured lines represent measurements of the irradiance in space using the known reflectance target. In (b), the reflectance target is placed at random locations within the view-plane where the target's surface normal is perpendicular to the hyperspectral camera's XY plane (i.e. the target is parallel to the view-line of the camera), while in (c) the reflectance target surface normal is also randomly offset to the hyperspectral camera's XY plane by an angle between -10° and 10°.

the proposed GP with light-source-mean model had the least error both inside and outside the view-plane when compared to the other models. The proposed model can be seen to combine the inferenced irradiance from the parametric least squares model outside of the view-plane, and data-driven GP with zero-mean model within the view-plane. The results between the two different experiments, where the angle of the target is fixed and varied, demonstrate that modelling of the irradiance is invariant to angle offsets in the target and that the proposed model still outperformed the others.

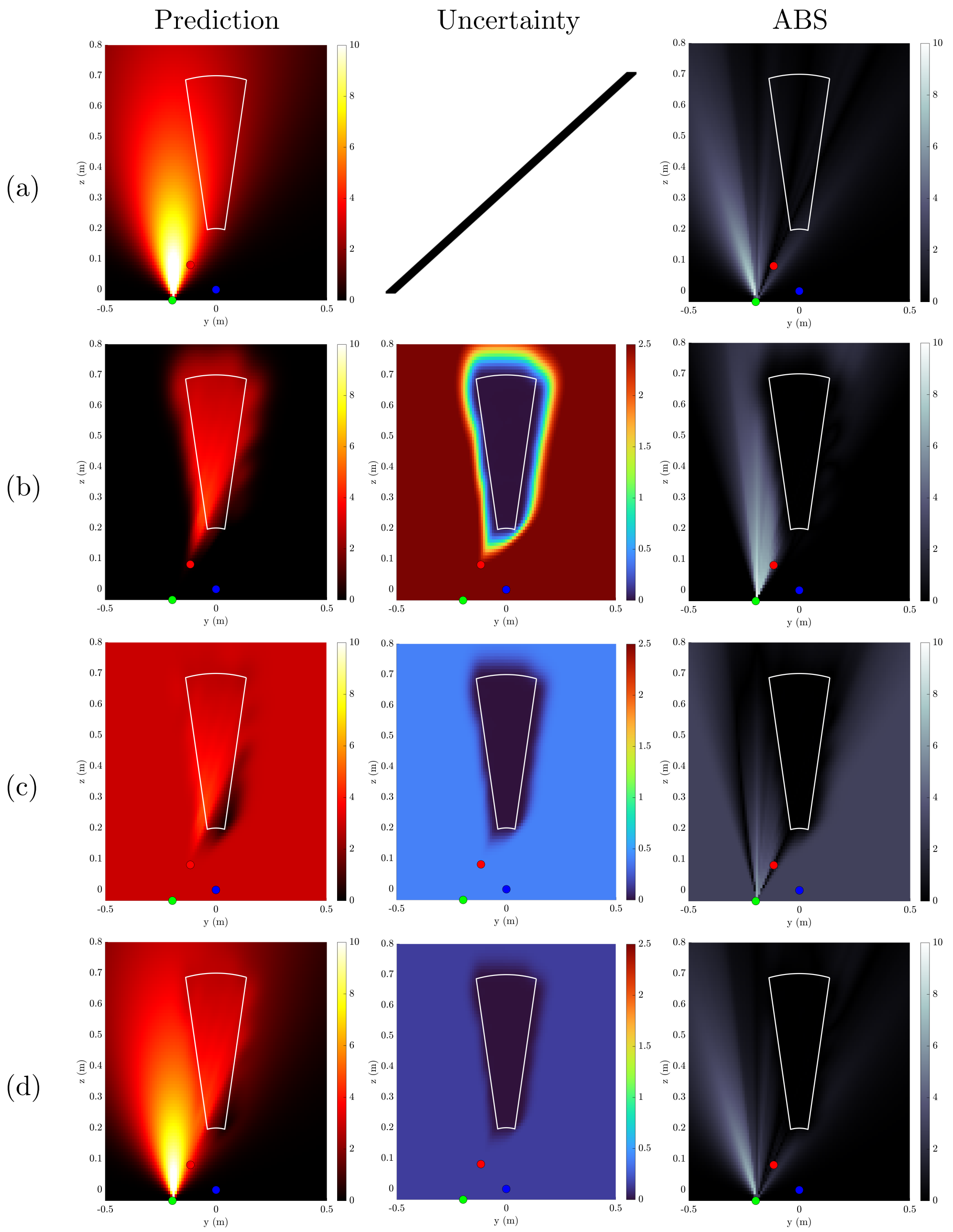


Figure 4.9: Comparison of inference methods for irradiance on the view-plane of the hyperspectral camera in simulation, where the reflectance target is placed at random locations within the view-plane, and the target's surface normal is fixed and perpendicular to the hyperspectral camera's XY plane. In the columns; (left) irradiance, (middle) uncertainty, and (right) absolute error. In the rows are the methods (a) least squares (b) GP with zero-mean (c) GP with constant-mean and (d) GP with light-source-mean. The light source, frame and line-scan cameras are denoted in green, red, and blue dots, respectively. The 2D frustum of the hyperspectral camera is displayed in white.

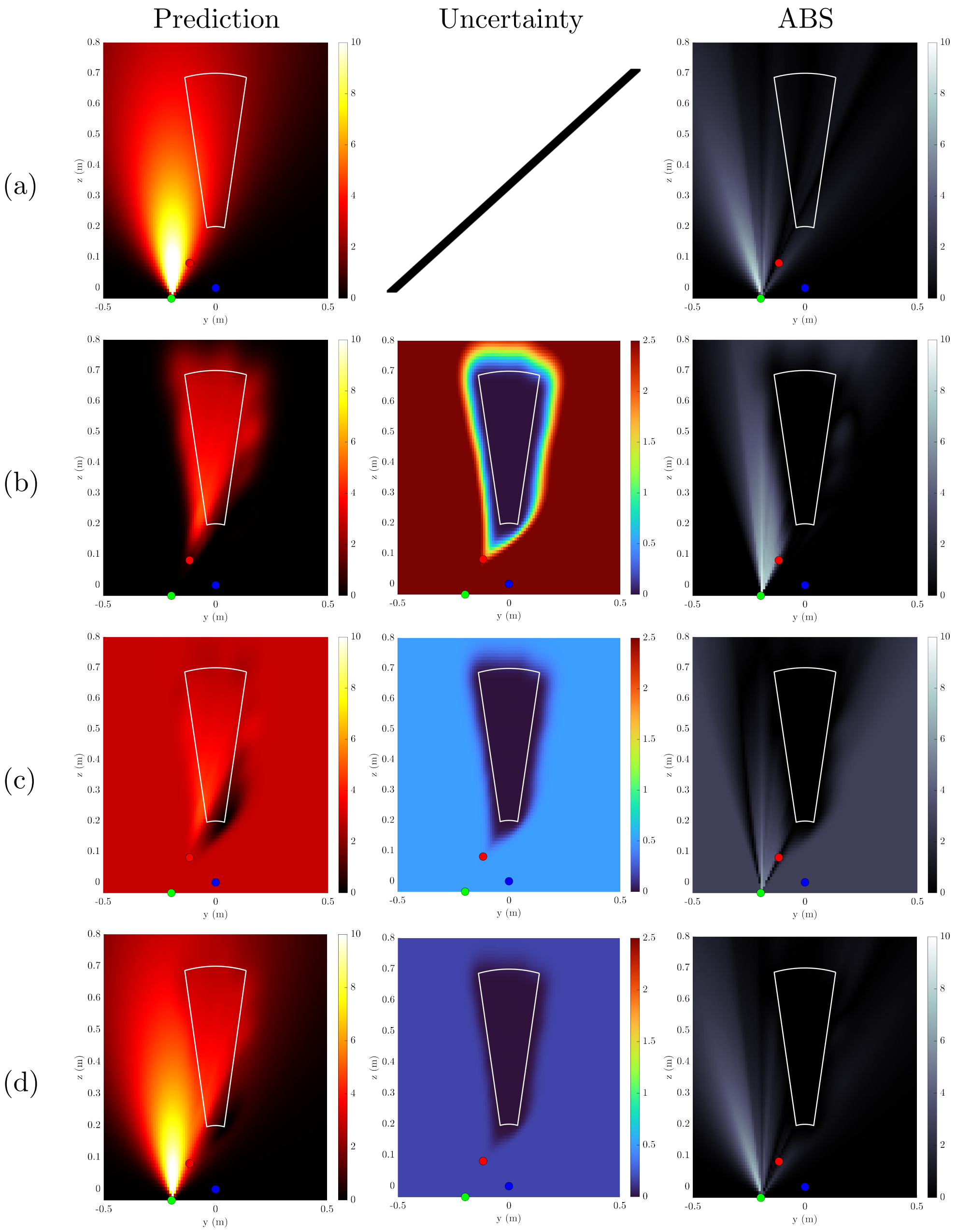


Figure 4.10: Comparison of inference methods for irradiance on the view-plane of the hyperspectral camera in simulation, where the reflectance target is placed at random locations within the view-plane, and the target's surface normal is randomly offset to the hyperspectral camera's XY plane by an angle between -10° and 10°. In the columns; (left) irradiance, (middle) uncertainty, and (right) absolute error. In the rows are the methods (a) least squares (b) GP with zero-mean (c) GP with constant-mean and (d) GP with light-source-mean. The light source, frame and line-scan cameras are denoted in green, red, and blue dots, respectively. The 2D frustum of the hyperspectral camera is displayed in white.

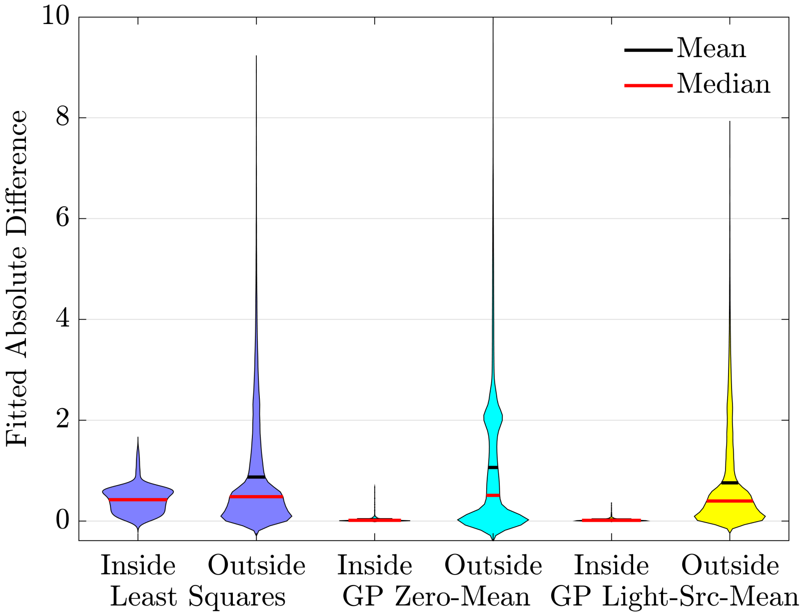


Figure 4.11: A violin plot showing the distribution of absolute errors in simulation of the irradiance for the least squares, GP with zero-mean and GP with light-source-mean both for points inside and outside the view-plane. In these results, the reflectance target is placed at random locations within the view-plane, and the target's surface normal is fixed and perpendicular to the hyperspectral camera's XY plane.

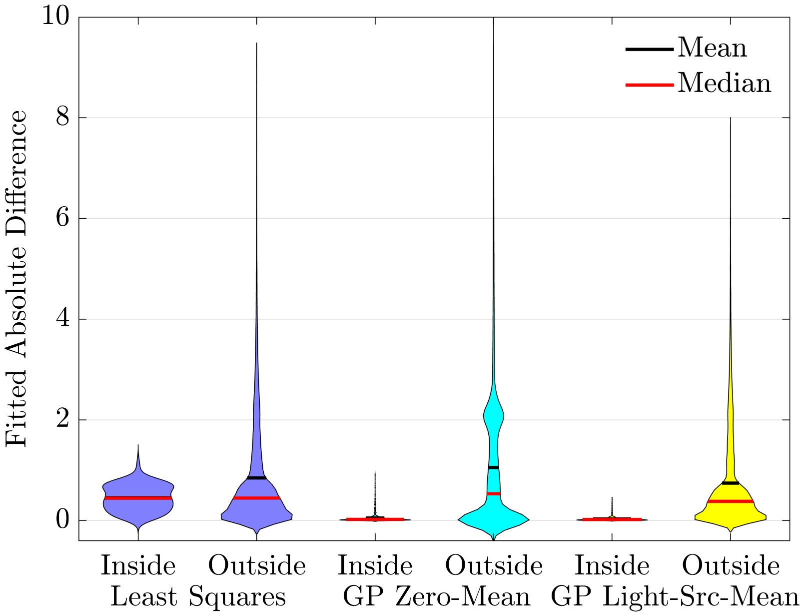


Figure 4.12: A violin plot showing the distribution of absolute errors in simulation of the irradiance for the least squares, GP with zero-mean and GP with light-source-mean both for points inside and outside the view-plane. In these results, the reflectance target is placed at random locations within the view-plane, and the target's surface normal is randomly offset to the hyperspectral camera's XY plane by an angle between -10° and 10°.

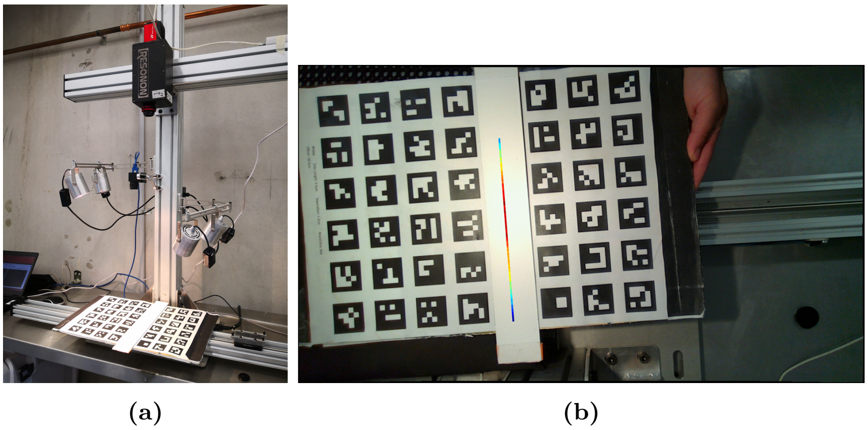


Figure 4.13: The hardware setup used to collect data for modelling a light source with a hyperspectral camera. In (a), the ArUco board with a diffuse white reflectance stripe used to capture data with the line-scan hyperspectral and frame camera system for a single light. In (b), is a frame camera image with a projected view-line image from the hyperspectral camera on the diffuse reflectance stripe.

4.5.2 Light Source Modelling with Hardware Setup

A light source modelling experiment was conducted with the hardware setup presented in Figure 4.1. The camera system consisted of a Resonon Pika NIR-320 hyperspectral line-scan camera and Intel Realsense L515 RGB-D frame camera. A single light source was rigidly mounted together with the camera system. The hyperspectral camera had 164 contiguous bands. We perform extrinsic calibration of the camera system following the method outlined in Chapter 3. A Lambertian target of known reflectance was attached to a board with ArUco [49] fiducial markers and the pose of the board was estimated using images from the frame camera as shown in Figure 4.13. The Lambertian target was comprised of a metal stripe with a white Avian-B reflectance coating [80], whose reflectance was calibrated using a Spectralon® diffuse reflectance standard [81]. To acquire the light source position and principal direction, we follow [64] using a reflective hemisphere attached to a ChArUco [49] board by detecting the bright spots on the hemisphere as shown in Figure 4.14. The irradiance was modelled separately for each band of the hyperspectral camera, and the results for one of these bands is shown in Figure 4.15. The measurements taken by the hyperspectral camera are seen to be non-continuous on its view-plane. When comparing the least squares and the GP with light-source-mean, it can be seen that the GP irradiance has captured some asymmetry and that there is a sharper fall-off with angle.

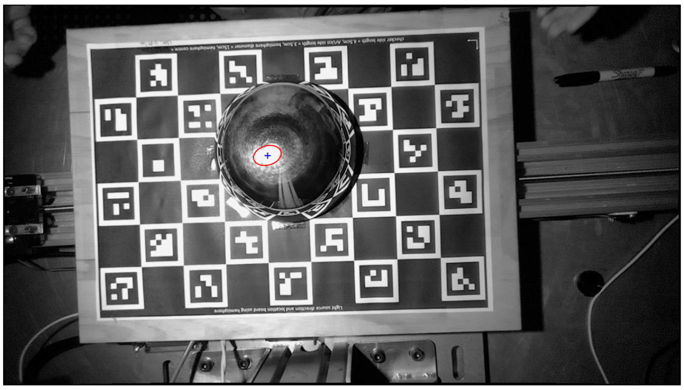
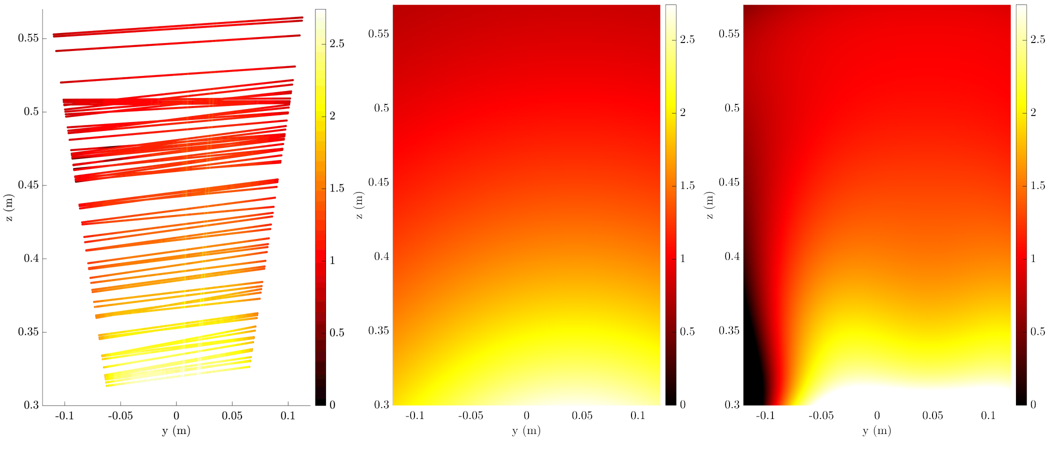


Figure 4.14: A metal reflective hemisphere attached to a ChArUco [49] board to triangulate the light source's position and principal direction w.r.t. the frame camera. An ellipse is fit to the brightest pixels on the hemisphere, shown in red, and its centre, shown in blue, is a single measurement. Multiple measurements of the board are recorded in varying poses.



(a) **(b)** **(c)**

Figure 4.15: The modelled irradiance on the view-plane of the hyperspectral camera for the setup shown in Figure 4.1. (a) measurements (b) fitted least squares, and (c) fitted GP with light-source-mean. These plots are for a single band.

4.6 Conclusions

The calibrated camera system from Chapter 3 was used to model the irradiance of a non-isotropic light source that is used to capture images with the hyperspectral camera. Initially, an appropriate parametric non-isotropic disk source model was introduced. It is shown that this model could be fit using least squares of hyperspectral measurements of a known Lambertian target. The parametric model assumes the RID of the light source is symmetric, but there exist inherent asymmetries which it cannot capture. A proposed data-driven GP model with a unique non-zero mean function, that is derived from the parametric model, imposes the knowledge of the general profile of irradiance while also capturing asymmetries that exist in the measurements. Experiments in simulation and hardware, show that the proposed GP model captures greater detail with less error when compared to the parametric model and other GP models. The proposed model successfully captures asymmetries in the irradiance for areas where measurements are possible.

The calibrated camera system now has a rigidly attached light source that has been fully modelled, where the irradiance and the direction of the light ray from the source can be determined for any point in 3D space for each of the spectral bands of the hyperspectral camera. This information defines the incident light on an object with unknown material properties. Given that the reflected light off a material, in the form of radiance, is measured by the hyperspectral camera, material properties that are independent to lighting and shape can be estimated. The next chapter takes advantage of all the modellings and calibrations that have been performed so far to estimate material properties.

Chapter 5

Reflectance Estimation with Light and Shape Information

Hyperspectral cameras measure the reflected light spectra from a surface with a given material on an object of interest. The reflected light is in the form called radiance which varies over the spectral bands and depends on the viewing direction, the amount of incident irradiance from the light source, the surface shape, and the properties of the material. These properties are unique and could be used to distinguish the material. One such material property is the diffuse reflectance. Estimating these properties from the radiance is a heavily underdetermined inverse problem as there are more unknown parameters than measurements. This chapter will focus on improving reflectance estimation, from radiance measurements captured by a hyperspectral camera, by incorporating known or modelled information. The information comes in the form of surface shape, which we can be acquired from depth images or 3D models, and light irradiance and direction which comes from the light source modelling discussed in Chapter 4.

The diffuse reflectance represents the proportion of incident light that is reflected by a material in a diffuse manner which varies over the spectral dimension. Determining this quantity is important for hyperspectral imaging to ensure consistency between measurements, especially when there are changes to lighting conditions and when capturing hypercubes with different camera sensors. For hyperspectral imaging, the calibrated reflectance is commonly determined by a simple normalisation between a dark reference, which is an image with the lens covered, and a white reference, which is an image of a highly reflective diffuse material. However, the calibrated reflectance is not robust for near-field experiments as it cannot capture the surface variation in lighting and shape. Material interaction with light can be described by physics-based reflectance models, where the diffuse reflectance is treated as an unknown parameter which is recovered through an optimisation. The reflectance here is then termed as estimated reflectance.

For a given material, the manner in which light interacts with its matter must be described by an appropriate physics-based reflectance model. For example, given a reflective material such as a mirror, almost all the incident radiant flux is reflected in a single direction. Therefore, the reflectance model for this example, which is known as the specular model, must capture this behaviour. In terms of reflectance models, specular represents one end of the extreme, where the other is a Lambertian (or diffuse) model. It assumes that the incident radiant flux is reflected equally in all directions. Generally, materials exhibit both specular and diffuse reflections, which is best captured using the dichromatic reflectance model (DRM). Further discussion of all reflectance models can be found in Section 2.5.

5.1 Literature Review

A property of the material that remains invariant to lighting, viewing direction and shape is the diffuse reflectance, which is referred to as reflectance in this chapter. This material property is the proportion of light reflected off the surface in a diffuse manner, which can be acquired from the measured hyperspectral radiance. Reflectance can be estimated either through a calibration procedure, which uses a target with known diffuse properties, or through an optimisation where the reflected light from an object is assumed to follow a physics-based model. Calibrated reflectance is generally not robust to lighting changes [26] and assumes the measured objects and the diffuse target to be of similar shape. As a result, calibrated reflectance has been shown to be less accurate for classification and regression applications [82]. Recovering the reflectance via optimisation is an ill-posed inverse problem, assumptions on the lighting, shape and material can help to constrain the problem [83][84].

A common reflectance model is the dichromatic reflectance model (DRM) which was first proposed by Shafer [85]. It assumes the radiance spectra measured by a camera can be fit to a 2D dichromatic hyperplane whose axes separate the specular and diffuse components [83]. The dichromatic reflectance model (DRM) is relatively simple in how it separates the two forms of reflection, but solving its inverse is ill-posed without any constraints. The diffuse component of the DRM has a term known as the shading factor, which describes the interaction of the incident light vector with the surface normal. The incident illumination is always smooth, and by assuming that most shapes are smooth, a common assumption is that shading factor is also smooth [83][86][87][88][84]. This constraint has been used as a regularisation term in a non-linear objective function, that also aims at reducing the palette of reflectance values by assuming that small patches of neighbouring pixels are of the same material [83][88][82]. Smoothness constraints have been applied to shape, illumination, reflectance and shading in the form of priors to recover reflectance using only a single image [84].

Recent work has used spectral similarity measures which compare the spectra of two pixels to determine their similarity with each other [86][87]. For a given pixel and its neighbourhood, cost functions to independently solve for diffuse and specular components are constructed, where the similarity measure is use to weight the cost due to a neighbouring pixel. These cost functions were found to be quadratic in both specular and diffuse components. The reflectance could then be calculated by rearranging the DRM. This style of approach has shown greater accuracy when compared to previous methods which had strict smoothness assumptions. We build upon this approach by reformulating the cost function to include known irradiance and light ray directions from the light source modelling obtained in Chapter 4, and known surface shape. The next section will now formally introduce the problem which includes the surface, camera system and light source.

5.2 Problem Statement

We now formally introduce the reflectance estimation problem with the calibrated camera system and modelled light source. An irregular surface, with a material that behaves according to the DRM, is illuminated by a near-field non-isotropic disk light source $\mathcal{L}$ and captured by a multi-modal camera system consisting of a line-scan hyperspectral camera $\mathcal{H}$ and a 2D RGB frame camera $\mathcal{F}$ as shown in Figure 5.1. The camera system is fully calibrated as described in Chapter 3. The fixed homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{H}} \in \mathbb{R}^{4\times4}$ is of the hyperspectral camera with respect to the frame camera which comes from the calibrated extrinsic parameters. $\mathcal{L}$ has a principal direction vector $\hat{\boldsymbol{v}}$ with maximum radiant intensity, a radius of r_d, and is rigidly attached to the frame camera by the known homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{L}}$.

The irradiance of the light source is spatially modelled using the Gaussian process (GP) with light-src-mean function as described in Chapter 4. The light vector $\boldsymbol{l} \in \mathbb{R}^3$ represents the ray that is emitted by the light source and arrives at the 3D surface point P with the surface normal $\hat{\boldsymbol{n}} \in \mathbb{R}^3$. $\boldsymbol{l}$ is defined as originating from P in the direction towards the light source to ensure the dot product between $\boldsymbol{l}$ and $\hat{\boldsymbol{n}}$ is positive [61]. At P, the radiance is reflected in both a diffuse and specular manner which in Figure 5.1 is represented as red and blue arrows respectively. The diffuse and specular radiance reflected in the direction of the hyperspectral camera is measured by the pixel p on $I_{\mathcal{H}}$.

The DRM is a crucial element of this work and must be formally introduced. In the next section we derive its equation from the known bidirectional reflectance distribution function (BRDF).

5.3 Dichromatic Reflectance Model

The amount of radiance L that is reflected at P is defined by the BRDF, which is a function of the reflected L to the irradiance E for a single combination of an incident and a reflected directions. Further details can be found in Section 2.5.1. Given that P is on a surface with a material that behaves according to the DRM, L is a linear combination of

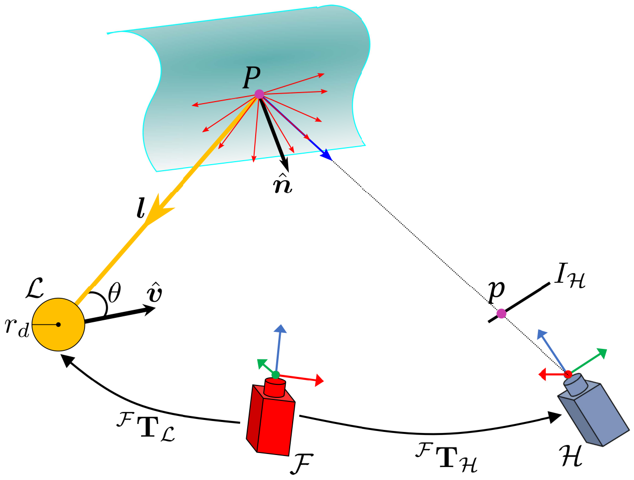


Figure 5.1: A calibrated camera system captures a surface with a material that is modelled using the dichromatic reflectance model (DRM). The calibrated camera system is comprised of a red, green, and blue colour channels (RGB) frame camera $\mathcal{F}$ and a line-scan hyperspectral camera $\mathcal{H}$, that are rigidly mounted by the homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{H}}$. A modelled non-isotropic disk light source $\mathcal{L}$ of radius r_d, with the principal direction vector $\hat{\boldsymbol{v}}$, is rigidly attached to $\mathcal{F}$ by the homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{L}}$. The light vector $\boldsymbol{l}$, which is a zenith angle θ from $\hat{\boldsymbol{v}}$, carries radiant flux to the surface point P with surface normal $\hat{\boldsymbol{n}}$, but $\boldsymbol{l}$ is defined as originating from P. The radiant flux is reflected off P as radiance in both a diffuse manner (red arrows) and a specular manner (blue arrow). The diffuse and specular radiance in the direction of the hyperspectral camera is measured by p on the image $I_{\mathcal{H}}$.

light that is reflected in a diffuse manner and specular manner. Therefore, the L headed towards the hyperspectral camera can be determined using the light transport equation (as defined in Section 2.5.2) of the DRM's BRDF (as defined in Section 2.5.6) for all incident directions[1] as follows:

$$L(p, \lambda) = k(p) \cdot E(p, \lambda) + g(p) \cdot \rho(p, \lambda) \cdot E(p, \lambda)\,, \tag{5.1}$$

where p is the pixel of P in the hyperspectral camera that measures L for the spectral band λ, due to the incident irradiance E which varies over λ. The specular coefficient k is the proportion of E that is reflected in a specular manner at P and depends on the viewing direction. The diffuse component at P is made up of ρ which is the material property of reflectance that varies with λ, and the shading factor g which is the proportion of diffuse

---

[1] The general form of the outgoing radiance considers all incoming directions of light regardless of the light sources. In this work we only consider the light from a single source. Therefore, there is only a single incident direction.

light due to the interaction between the incident light ray and surface shape [83]. It is defined as follows:

$$g(p) = h\frac{\hat{\boldsymbol{l}} \bullet \hat{\boldsymbol{n}}}{\pi}\,, \tag{5.2}$$

where $\hat{\boldsymbol{l}}$ is the direction vector of $\boldsymbol{l}$ and h is an unknown constant based on the finite solid angle at p due to the light source.

5.3.1 Estimated Irradiance

The irradiance E in Equation (5.1) is estimated by sampling the built GP light source model of $\mathcal{L}$ for which requires P for all pixels of the hyperspectral camera . These 3D points can be determined by ray casting p to a 3D model of the object, which is only possible due to the calibration between the hyperspectral camera and the frame camera . Once recovering all P, the 3D points are transformed to the light source using ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{L}}$, then converted to 2D light source spherical coordinates as described in Equation (4.1), which are then used to sample the trained GP light source model in Equation (4.10) for each λ. Further details of the light source modelling can be found in Chapter 4.

5.3.2 Discretised DRM

Given that E is recovered for all p and λ, Equation (5.1) can be rearranged and discretised using the measured p and λ of the hyperspectral camera as follows:

$$\boldsymbol{R}_{p,\lambda} = \frac{\boldsymbol{L}_{p,\lambda}}{\boldsymbol{E}_{p,\lambda}} = \boldsymbol{k}_p + \boldsymbol{g}_p\boldsymbol{\rho}_{p,\lambda}\,, \tag{5.3}$$

where $\boldsymbol{R}$ is the ratio of reflected radiance to incident irradiance which will be used in all later formulations. The variables $\boldsymbol{R}, \boldsymbol{L}, \boldsymbol{\rho}, \boldsymbol{E} \in \mathbb{R}^{u \times v \times \lambda}$ are all 3D hypercubes and $\boldsymbol{k}, \boldsymbol{g} \in \mathbb{R}^{u \times v}$ are both 2D images.

In this section the DRM has been formally introduced, which is assumed to be the reflectance model that describes the interaction of all materials in this work. The irradiance arriving on a surface from the light source is estimated using the established light source models from Chapter 4, which is then used to calculate a ratio of measured radiance to irradiance. A discretised form of the DRM, which works with the measured pixels and spectral bands of the hyperspectral camera , converts continuous quantities to matrices and hypercubes. This next section will formulate the reflectance estimation.

5.4 Reflectance Estimation

Now we propose a reflectance estimation method that incorporates light source directions from the light source modelling, and surface normals which we assume are given. This method builds upon existing work that formulates $\boldsymbol{g}$ and $\boldsymbol{k}$ from Equation (5.3) as quadratic programming problems and solves them independently [86][87]. The solutions to the optimisations will allow us to calculate reflectance as:

$$\boldsymbol{\rho}_{p,\lambda} = \frac{\boldsymbol{R}_{p,\lambda} - \boldsymbol{k}_p}{\boldsymbol{g}_p} \,. \tag{5.4}$$

5.4.1 Quadratic Programming Solution

Both g and k are recovered for all relevant pixels through independent unconstrained quadratic programming formulations where the general solution for a vector $\boldsymbol{z} \in \mathbb{R}^{u \cdot v}$ is defined as such:

$$\begin{aligned} \boldsymbol{z}^* = \quad & \arg\min_{\boldsymbol{z}} \quad && \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{Q}\boldsymbol{z} + \boldsymbol{q}^\top \boldsymbol{z} \\ & \text{s.t.} && \boldsymbol{b}_l \leq \boldsymbol{z} \leq \boldsymbol{b}_u \,, \end{aligned} \tag{5.5}$$

where $\boldsymbol{z}^* \in \mathbb{R}^{uv}$ is the vectorised solution of all pixels which are restricted within the upper and lower bounds of $\boldsymbol{b}_u$ and $\boldsymbol{b}_l \in \mathbb{R}^{uv}$ respectively. The matrix $\mathbf{Q} \in \mathbb{R}^{uv \times uv}$ is a sparse square symmetric matrix that quantifies the similarity between pixels and their neighbourhoods, and the vector $\boldsymbol{q}$ enforces some cost on individual pixels. There are no equality and inequality constraints for this problem. The optimisation can be solved with a convex interior-point algorithm. The objective functions that populate $\boldsymbol{Q}$ and $\boldsymbol{q}$, for both g and k, are defined in the following sections for *one pixel*, whereas the optimisation solves **all pixels** at once.

Although the irradiance has been removed in Equation (5.3), the number of equations given by $\boldsymbol{R}$ is $(uv\lambda)$, and the number of unknowns given by $(\boldsymbol{\rho}, \boldsymbol{k}, \boldsymbol{g})$ is $(uv\lambda + 2uv)$. Therefore, recovering the reflectance is still an underdetermined problem and further constraints are required. The additional constraints are derived from smoothness in the shading factor over a pixel and its neighbourhood, which are weighted based on pixel similarity. The next section will discuss one such similarity metric that is used to weight the cost introduced by individual pixels.

5.4.2 Spectral Angle Mapper

It is important to consider whether a pixel and its neighbourhood contain the same material as this will constrain the behaviour of $\boldsymbol{g}$ and $\boldsymbol{k}$. The similarity between two pixels can be measured using the scale-invariant spectral angle mapper (SAM) metric [89] which is defined as follows:

$$\mathrm{SAM}\left(\boldsymbol{r}_p, \boldsymbol{r}_q\right) = \arccos\left(\frac{\boldsymbol{r}_p \bullet \boldsymbol{r}_q}{||\boldsymbol{r}_p|| \cdot ||\boldsymbol{r}_q||}\right) , \tag{5.6}$$

where p and q are pixels with spectra vectors $\boldsymbol{r}_p$ and $\boldsymbol{r}_q \in \mathbb{R}^{\lambda}$ respectively, which both come from the hypercube $\boldsymbol{R}_{p,\lambda}$. A closer inspection of SAM reveals that it calculates the 2D angle between any two N-dimensional vectors. Therefore, its output range is between 0 and $\frac{\pi}{2}$ radians, which is not useful as weighting terms. An alternative metric is constructed using SAM with an exponential function as follows:

$$\zeta_{p,q} = \exp\left(\frac{-\,\mathrm{SAM}\left(\boldsymbol{r}_p, \boldsymbol{r}_q\right)^2}{\gamma}\right) . \tag{5.7}$$

The value $\zeta_{p,q}$ returns the similarity between 0 and 1, where 1 represents the pixels being comprised of the same material. The parameter γ is used to alter the strictness in the similarity, where a larger value gives a higher similarity measure for the same pixels. The following notation for ζ will appear in this work:

$$\zeta_p = \zeta_{p,\mathbf{1}} \tag{5.8}$$

$$\bar{\zeta} = 1 - \zeta\,, \tag{5.9}$$

where Equation (5.8) compares $\boldsymbol{r}_p$ with a vector of ones, of equal size, to check if the pixel is grey[2], and Equation (5.9) calculates the difference between two pixels or between a pixel and vector of ones. This similarity measure will be used to weight the quadratic cost functions that recovery shading factor and specular coefficient.

5.4.3 Recovering Shading Factor

In order to recover the shading factor a cost function must be constructed which we will derive from its smoothness property. Firstly, the shading factor from Equation (5.2) can be written in the following form:

$$\begin{gathered} g(p) = hd\,, \\ \text{where} \\ d = \frac{\hat{\boldsymbol{l}} \bullet \hat{\boldsymbol{n}}}{\pi} \end{gathered} \tag{5.10}$$

where $\hat{\boldsymbol{l}}$ comes from the light source model, $\hat{\boldsymbol{n}}$ is given, and the constant h is unknown. Therefore, the optimisation will involve estimating h with the known d, which is then used to recover g.

Firstly, an intermediate variable is defined as $a = \ln(h)$ that assumes $h > 0$, which must hold as $g > 0$ is always true. The cost function $f_g(a)$ is used for the pixel-wise recovery of a, that involves three similarity cases between a pixel p and a pixel q from its surrounding neighbourhood $\mathcal{N}(p)$. The cost function is shown as follows:

$$f_g(a) = f_1(a) + f_2(a) + f_3(a)\,. \tag{5.11}$$

---

[2] A grey pixel has equal intensity across all its spectral bands and has no specularity

Case 1: Pixels p and q are of the same material but neither are grey.

If p and q are of the same material, then their unknown reflectance spectra $\boldsymbol{\rho}_p$ and $\boldsymbol{\rho}_q \in \mathbb{R}^\lambda$ respectively would be equal for all λ. We start by defining $\boldsymbol{r}_p$ using the discretised DRM in Equation (5.3) as such:

$$\boldsymbol{r}_p = k_p + g_p \cdot \boldsymbol{\rho}_p \,, \tag{5.12}$$

where (k_p, g_p) are both scalar parameters that do not vary with λ. We can determine the variation in $\boldsymbol{r}_p$ in Equation (5.12) over λ by calculating the standard deviation (STD) σ_p as follows:

$$\sigma_p = g_p \cdot \mathrm{std}(\boldsymbol{\rho}_p, \lambda) \,, \tag{5.13}$$

where $\mathrm{std}(\boldsymbol{\rho}_p, \lambda)$ is the STD of the reflectance over its spectral bands. Given that p and q are both of the same material then $\boldsymbol{\rho}_p = \boldsymbol{\rho}_q = \boldsymbol{\rho}$, the ratio of STDs in $\boldsymbol{r}_p$ and $\boldsymbol{r}_q$ can be shown to be equal to their ratio of shading factors, which can then be redefined using Equation (5.2) as follows:

$$\frac{\sigma_p}{\sigma_q} = \frac{g_p \cdot \mathrm{std}(\boldsymbol{\rho}, \lambda)}{g_q \cdot \mathrm{std}(\boldsymbol{\rho}, \lambda)} = \frac{g_p}{g_q} = \frac{h_p \cdot d_p}{h_q \cdot d_q} \,. \tag{5.14}$$

By applying the natural log to Equation (5.14) the ratios are transformed to differences as follows:

$$a_p - a_q + \ln(\sigma_q) - \ln(\sigma_p) + \ln(d_p) - \ln(d_q) = 0 \,, \tag{5.15}$$

then $f_1(a)$ can be written as the sum of costs for all p as such:

$$\begin{gathered} f_1(a) = \sum_{p} \sum_{q \in \mathcal{N}(p)} \zeta_{p,q} \bar{\zeta}_p \bar{\zeta}_q \left(\breve{a} + \breve{\sigma} + \breve{d} \right)^2 \\ \text{where} \\ \breve{a} = a_p - a_q \\ \breve{\sigma} = \ln(\sigma_q) - \ln(\sigma_p) \\ \breve{d} = \ln(d_p) - \ln(d_q) \ , \end{gathered} \tag{5.16}$$

where the weights promote pixels that are similar but not grey.

Case 2: Pixels p and q are of different material but neither are grey.

If p and q are not of the same material, then their unknown reflectance spectra $\boldsymbol{\rho}_p$ and $\boldsymbol{\rho}_q$ respectively would not be equal for any λ. In this case we assume the shading factor remains smooth which can be defined as the shading factors being approximately equal. This can be redefined using Equation (5.2) as follows:

$$\begin{gathered} g_p \approx g_q \\ (h)_p \cdot d_p \approx (h)_q \cdot d_q \, . \end{gathered} \tag{5.17}$$

By applying the natural log and rearranging the following form is given:

$$a_p - a_q + \ln(d_p) - \ln(d_q) \approx 0 \, , \tag{5.18}$$

then $f_2(a)$ can be written as the sum of costs for all p as such:

$$\begin{gathered} f_2(a) = \sum_{p} \sum_{q \in \mathcal{N}(p)} \left(1 - \zeta_{p,q} \zeta_p \zeta_q - \zeta_{p,q} \bar{\zeta}_p \bar{\zeta}_q \right) \left(\breve{a} + \breve{d} \right)^2 \\ \text{where} \\ \breve{a} = a_p - a_q \\ \breve{d} = \ln(d_p) - \ln(d_q) \ , \end{gathered} \tag{5.19}$$

where the weights promote pixels that are different but not grey.

Case 3: Pixels p and q are both equally grey and have no specularity.

If p and q are both equally grey $\boldsymbol{\rho}_p$ and $\boldsymbol{\rho}_q$ would be equal for all λ, and it would not be possible to separate the diffuse and specular components as the STDs of $\boldsymbol{r}_p$ and $\boldsymbol{r}_q$ are both 0. Therefore, grey pixels are assumed to have no specular component which means that $k = 0$. Rather than calculating the STD we consider the average of $\boldsymbol{r}_p$ from Equation (5.12) over the λ by calculating the mean μ_p as follows:

$$\mu_p = g_p \cdot \text{mean}(\boldsymbol{\rho}_p, \lambda)\,, \tag{5.20}$$

where mean$(\boldsymbol{\rho}_p, \lambda)$ is the mean of the reflectance over its spectral bands. Given that p and q are both equally grey then $\boldsymbol{\rho}_p = \boldsymbol{\rho}_q = \boldsymbol{\rho}$, and the ratio in means of p and q can be shown to be equal to their ratio of shading factors which can then be redefined using Equation (5.2) as follows:

$$\frac{\mu_p}{\mu_q} = \frac{g_p \cdot \text{mean}(\boldsymbol{\rho}, \lambda)}{g_q \cdot \text{mean}(\boldsymbol{\rho}, \lambda)} = \frac{g_p}{g_q} = \frac{h_p \cdot d_p}{h_q \cdot d_q}\,. \tag{5.21}$$

By applying the natural log to Equation (5.21) the ratios are transformed to differences as follows:

$$a_p - a_q + \ln(\mu_q) - \ln(\mu_p) + \ln(d_p) - \ln(d_q) = 0\,, \tag{5.22}$$

then $f_3(a)$ can be written as the sum of costs for all p as such:

$$\begin{gathered} f_3(a) = \sum_p \sum_{q \in \mathcal{N}(p)} \zeta_{p,q} \zeta_p \zeta_q \left(\breve{a} + \breve{\mu} + \breve{d}\right)^2 \\ \text{where} \\ \breve{a} = a_p - a_q \\ \breve{\mu} = \ln(\mu_q) - \ln(\mu_p) \\ \breve{d} = \ln(d_p) - \ln(d_q)\ , \end{gathered} \tag{5.23}$$

where the weights promote pixels that are grey.

To summarise the three cases, for a pixel p and its neighbourhood $\mathcal{N}(p)$: $f_1(a)$ enforces smoothness in g between pixels of the same material that are not grey, $f_2(a)$ enforces smoothness in g between pixels of different materials that are not grey, and $f_3(a)$ enforces smoothness in g between grey pixels that have no specularity. The solution to Equation (5.11) via quadratic optimisation, as shown in Equation (5.5), with no bounds is a vector, which is then converted to a 2D image of the same dimensions as $\boldsymbol{g}$, which gives $\boldsymbol{a}^* \in \mathbb{R}^{u \times v}$. We can then calculate $\boldsymbol{g}$ as follows:

$$\boldsymbol{g} = \exp(\boldsymbol{a}^*)\boldsymbol{d}\,, \tag{5.24}$$

where $\boldsymbol{d} \in \mathbb{R}^{u \times v}$ is the 2D image of d for all pixels, and all operations are element-wise.

5.4.4 Recovering Specular Coefficient

We will not formulate the cost function to recover the specular coefficient, which is a similar process to the shading factor. The specular coefficient can be thought of as explaining the remaining portion of reflected light that is not diffuse. It is recovered in a similar manner to the shading factor by first defining an intermediate variable:

$$b = \ln\left(h \cdot \mathrm{mean}(\boldsymbol{\rho}, \lambda)\right)\,, \tag{5.25}$$

where h is the unknown constant in the shading factor as shown in Equation (5.10) and the mean is of the unknown ρ for a pixel over its λ. Note that although the value of h for all pixels can be acquired through the solution $\boldsymbol{a}^*$ as described in Section 5.4.3, for the formulation of k, we assume it to be unknown. The significance of b is that it is comprised of all unknown variables that relate to the diffuse portion of reflected light. This results in the two optimisations being independent of each other and thus could be run in parallel.

The cost function $f_k(b)$ is used for the pixel-wise recovery of b, that involves two similarity cases between a pixel p and a pixel q from its surrounding neighbourhood $\mathcal{N}(p)$. The cost function is shown as follows:

$$f_k(b) = f_4(b) + f_5(b)\,. \tag{5.26}$$

Case 4: Pixels p and q are of the same material but neither are grey.

Similar to case 1, if p and q are of the same material, then their unknown reflectance spectra $\boldsymbol{\rho}_p$ and $\boldsymbol{\rho}_q \in \mathbb{R}^\lambda$ respectively would be equal for all λ. Therefore, the following expression is true mean$(\boldsymbol{\rho}_p, \lambda) =$ mean$(\boldsymbol{\rho}_q, \lambda)$. In Equation (5.14), it was shown that the ratio of STDs in $\boldsymbol{r}_p$ and $\boldsymbol{r}_q$ is equal to their ratio of shading factors. We can then modify Equation (5.14) to include the means of $\boldsymbol{\rho}$ for both pixels as follows:

$$\frac{\sigma_p}{\sigma_q} = \frac{g_p \cdot \mathrm{mean}(\boldsymbol{\rho}_q)}{g_q \cdot \mathrm{mean}(\boldsymbol{\rho}_q)} = \frac{d_p}{d_q} \cdot \frac{h_p \cdot \mathrm{mean}(\boldsymbol{\rho}_q)}{h_q \cdot \mathrm{mean}(\boldsymbol{\rho}_q)}\,. \tag{5.27}$$

By applying the natural log to Equation (5.27) the ratios are transformed to differences as follows:

$$b_p - b_q + \ln(\sigma_q) - \ln(\sigma_p) + \ln(d_p) - \ln(d_q) = 0\,, \tag{5.28}$$

then $f_4(b)$ can be written as the sum of costs for all p as such:

$$\begin{gathered} f_4(b) = \sum_p \sum_{q \in \mathcal{N}(p)} \zeta_{p,q} \bar{\zeta}_p \bar{\zeta}_q \left(\breve{b} + \breve{\sigma} + \breve{d} \right)^2 \\ \text{where} \\ \breve{b} = b_p - b_q \\ \breve{\sigma} = \ln(\sigma_q) - \ln(\sigma_p) \\ \breve{d} = \ln(d_p) - \ln(d_q)\ , \end{gathered} \tag{5.29}$$

where the weights promote pixels that are similar but not grey.

Case 5: Pixels p and q of different material or grey generally have no specularity.

If p and q are not of the same material, then their unknown reflectance spectra $\boldsymbol{\rho}_p$ and $\boldsymbol{\rho}_q$ respectively would not be equal for any λ. In this case, we assume the specular component is generally negligible when compared to the diffuse component, which is already assumed for grey pixels. This means that $k \approx 0$ as shown in Equation (5.20). Similar to Equation (5.21), the ratio of means in $\boldsymbol{r}_p$ and $\boldsymbol{r}_q$ can be written as such:

$$\frac{\mu_p}{\mu_q} = \frac{g_p \cdot \text{mean}(\boldsymbol{\rho}, \lambda)}{g_q \cdot \text{mean}(\boldsymbol{\rho}, \lambda)} = \frac{d_p}{d_q} \cdot \frac{h_p \cdot \text{mean}(\boldsymbol{\rho}, \lambda)}{h_q \cdot \text{mean}(\boldsymbol{\rho}, \lambda)} \,. \tag{5.30}$$

By applying the natural log the ratios are transformed to differences as follows:

$$b_p - b_q + \ln(\mu_q) - \ln(\mu_p) + \ln(d_p) - \ln(d_q) = 0 \,, \tag{5.31}$$

then $f_5(b)$ can be written as the sum of costs for all p as such:

$$\begin{gathered} f_5(b) = \sum_p \sum_{q \in \mathcal{N}(p)} \left(1 - \zeta_{p,q} \bar{\zeta}_p \bar{\zeta}_q\right) \left(\breve{b} + \breve{\mu} + \breve{d}\right)^2 \\ \text{where} \\ \breve{b} = b_p - b_q \\ \breve{\mu} = \ln(\mu_q) - \ln(\mu_p) \\ \breve{d} = \ln(d_p) - \ln(d_q) \ , \end{gathered} \tag{5.32}$$

where the weights promote pixels that are different or grey.

To summarise the two cases, for a pixel p and its neighbourhood $\mathcal{N}(p)$: $f_4(b)$ enforces smoothness in g between pixels of the same material that are not grey, and $f_5(b)$ assumes that k is negligible for pixels that are either of different material or grey. The quadratic optimisation, as shown in Equation (5.5), requires the following bounds which are described for a single pixel p as follows:

$$\ln(\mu_p - \min(\boldsymbol{r}_p, \lambda)) - \ln(d_p) \leq b_p \leq \ln(\mu_p) - \ln(d_p) \,, \tag{5.33}$$

where $\min(\boldsymbol{r}_p, \lambda)$ is the **non-zero** minimum in $\boldsymbol{r}_p$. These bounds will ensure that the recovered k is always positive and upper-bounded. The solution to Equation (5.26) via quadratic optimisation with the bounds above is a vector, which is then converted to a 2D image of the same dimensions as $\boldsymbol{k}$, which gives $\boldsymbol{b}^* \in \mathbb{R}^{u \times v}$. We can then calculate $\boldsymbol{k}$ as follows:

$$\boldsymbol{k} = \boldsymbol{\mu} - \exp(\boldsymbol{b}^*)\boldsymbol{d}\,, \tag{5.34}$$

where $\boldsymbol{d} \in \mathbb{R}^{u \times v}$ is the 2D image of d for all pixels, $\boldsymbol{\mu} \in \mathbb{R}^{u \times v}$ is the 2D image of μ_p for all pixels, and all operations are element-wise.

5.5 Experiments

5.5.1 Reflectance Estimation in Simulation

The proposed reflectance estimation method is validated using the unbiased physically-based renderer Mitsuba [90] to generate hyperspectral-like scenes with ground truth reflectance. The renderer works in the visible spectrum and is capable of binning into contiguous channels between wavelengths of 360nm to 830nm. The scenes are rendered with a frame camera sensor and a single known spot-light source using 100 channels, where the width of each channel bin is 4.75nm. The renderer outputs hypercubes of radiance, reflectance, and irradiance, alongside with a normal map and a 3D position map relative to the frame camera. The Stanford bunny is rendered with a rough plastic (RP) material at surface roughness values of root mean square (RMS) of 0.3, 0.5 and 1.0. Mixture materials with the RP are rendered with smooth dielectric and conductor materials at 50/50 and 80/20 linear proportions. Figure 5.2 shows the maps of surface normals, surface light source direction vectors, and the resulting dot product. The reflectance is estimated using the methods of Robles[83], Krebs[87], and our proposed method. Figure 5.3 shows the output shading factor, specular coefficient and reflectance of the methods for one material. When comparing the shading factors, it can be seen that the proposed method produces an output that closely resembles the dot product between the light source directions and

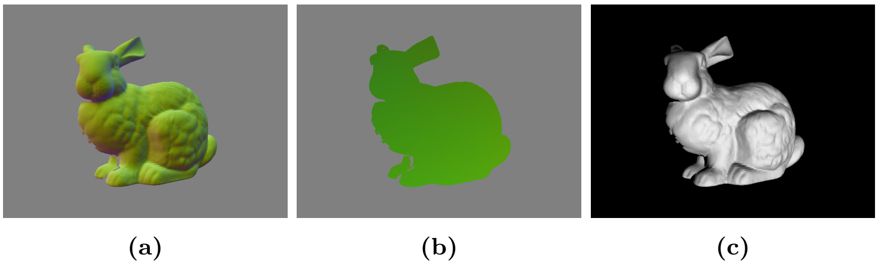

(a) (b) (c)

Figure 5.2: The variation in shape and lighting for the Stanford bunny. (a) surface normals, (b) light source direction map, and (c) resulting dot product of (a) and (b).

surface normals in Figure 5.2. The estimated reflectance is also more homogeneous. The outputs were compared to ground truth using the SAM metric and sum of squared error (SSE) of normalised reflectance per band. These results are shown in Table 5.1. Our proposed method has the highest accuracy for all materials, achieving average reductions of 82.1% in SAM and 96.8% in SSE when compared to the other two methods.

5.5.2 Reflectance Estimation of Lamb Cut Sample

The calibrated hardware shown in Figure 5.4 was used to scan a lamb cut sample. When the cut is scanned, it is placed on a tray that moves using a stepper motor which simultaneously triggers the acquisition of camera images. The light source that was used to capture hyperspectral images had been modelled prior to the data collection using the GP model outlined in Chapter 4. Depth images were used to construct a 3D model, where the hyperspectral pixels are ray-traced onto the model to acquire their surface intersection points and surface normals. The irradiance that arrived on the surface was estimated by querying the built GP model using the surface intersection points. Figure 5.5 shows the map of the surface normals, the map of light source direction vectors on the surface, and the resulting dot product. The hyperspectral line-scan images were combined into a single hypercube and the reflectance was calculated using four methods of calibrated, Robles[83], Krebs[87], and our proposed method. The results for the shading factor, specular coefficient and reflectance are shown in Figure 5.6. The calibrated reflectance does not estimate the specular coefficient or shading factor. If we compare the shading factors, it can be seen that the proposed method produces an output that closely resembles the dot product between the light source directions and surface normals in Figure 5.5. The

Table 5.1: Quantitative results of the pixel averaged estimated reflectance of different materials simulated with Mitsuba [90]. A rough plastic (RP) material is rendered in a pure state and as a mixture where the roughness is quantified in Root mean square (RMS). Three different estimated reflectance methods are compared of Robles [83], Krebs [87], and the proposed method. The spectral similarity to ground truth is calculated the spectral angle mapper (SAM) measure. The error with ground truth is calculated using the sum of squared error (SSE).

Material	**Method**	**SAM (std) rad (rad)**	**SSE (no units)**
RP 0.3 RMS	Robles [83]	0.023 (0.038)	1.256E+00
	Krebs [87]	0.017 (0.033)	1.919E-04
	Ours	**0.004 (0.004)**	**1.286E-05**
RP 0.5 RMS	Robles [83]	0.031 (0.025)	2.445E+00
	Krebs [87]	0.016 (0.016)	6.793E-05
	Ours	**0.002 (0.001)**	**4.038E-06**
RP 1.0 RMS	Robles [83]	0.040 (0.016)	2.973E+00
	Krebs [87]	0.024 (0.008)	6.210E-05
	Ours	**0.005 (0.002)**	**4.101E-06**
80% RP 0.3 RMS 20% smooth dielectric	Robles [83]	0.023 (0.038)	1.225E+00
	Krebs [87]	0.017 (0.032)	1.918E-04
	Ours	**0.004 (0.004)**	**1.291E-05**
50% RP 0.3 RMS 50% smooth dielectric	Robles [83]	0.023 (0.038)	1.107E+00
	Krebs [87]	0.017 (0.032)	1.916E-04
	Ours	**0.004 (0.004)**	**1.212E-05**
80% RP 0.3 RMS 20% smooth conductor	Robles [83]	0.023 (0.038)	1.180E+00
	Krebs [87]	0.017 (0.032)	1.913E-04
	Ours	**0.004 (0.004)**	**1.289E-05**
50% RP 0.3 RMS 50% smooth conductor	Robles [83]	0.023 (0.038)	1.095E+00
	Krebs [87]	0.017 (0.033)	1.919E-04
	Ours	**0.004 (0.004)**	**1.225E-05**

estimated reflectance from the proposed method has the least texture and similar regions are seen to be more homogeneous.

5.6 Conclusions

A reflectance estimation method is established for near-field hyperspectral imagery that incorporates known surface shape and lighting information. The method takes advantage of a calibrated camera system, as discussed in Chapter 3, and the irradiance and directions from a modelled light source, as discussed in Chapter 4. Materials are assumed to behave

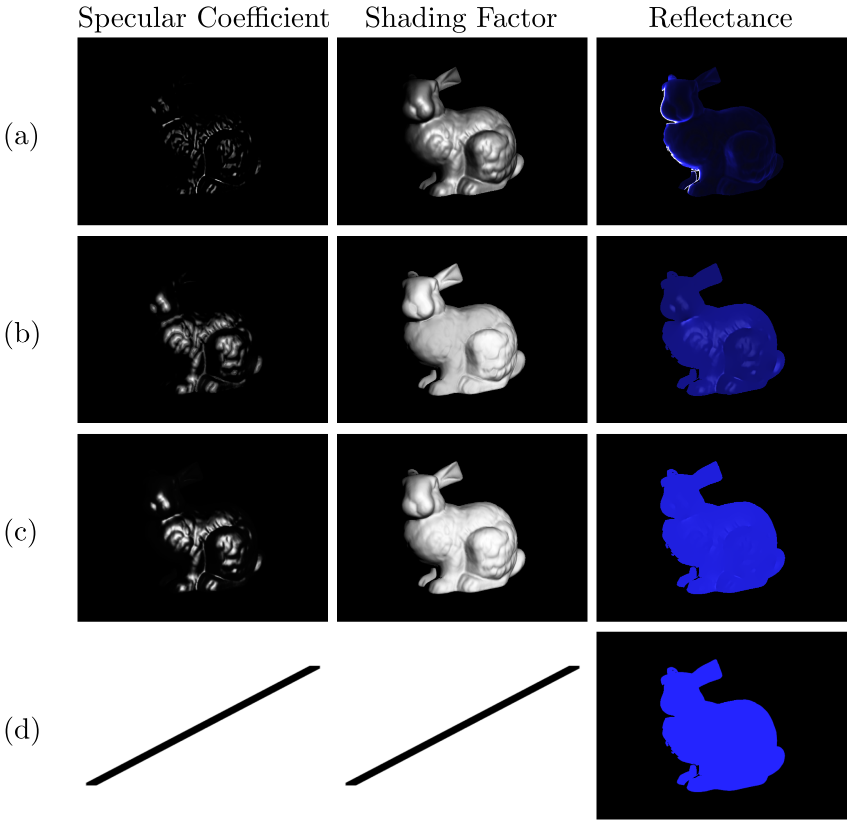


Figure 5.3: Outputs for the reflectance estimation of the rough plastic (RP) 0.3 root mean square (RMS) simulated material. In the columns: (left) specular coefficient, (middle) shading factor, and (right) reflectance. Rows contain results from: (a) Robles [83], (b) Krebs [87], (c) our method, and (d) ground truth. The reflectance is presented as a false colour image where channel bins with centre-wavelengths of 452nm, 547nm, and 737nm are mapped to RGB.

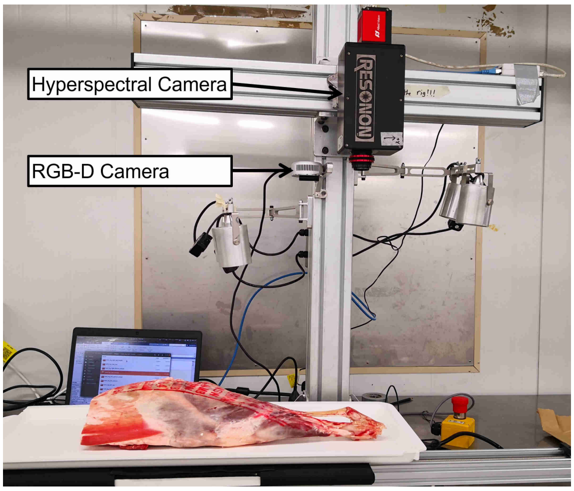


Figure 5.4: Rigidly mounted camera system comprised of a Resonon Pika NIR-320 hyperspectral line-scan camera and an Intel Realsense L515 RGB-D camera. While a number of light sources are present, we only use one. The lamb cut is moved via stepper motor which simultaneously triggers data collection.

(a) **(b)** **(c)**

Figure 5.5: The variation in shape and lighting for a lamb cut sample. (a) The surface normals, (b) the light source direction map, and (c) the resulting dot product of (a) and (b).

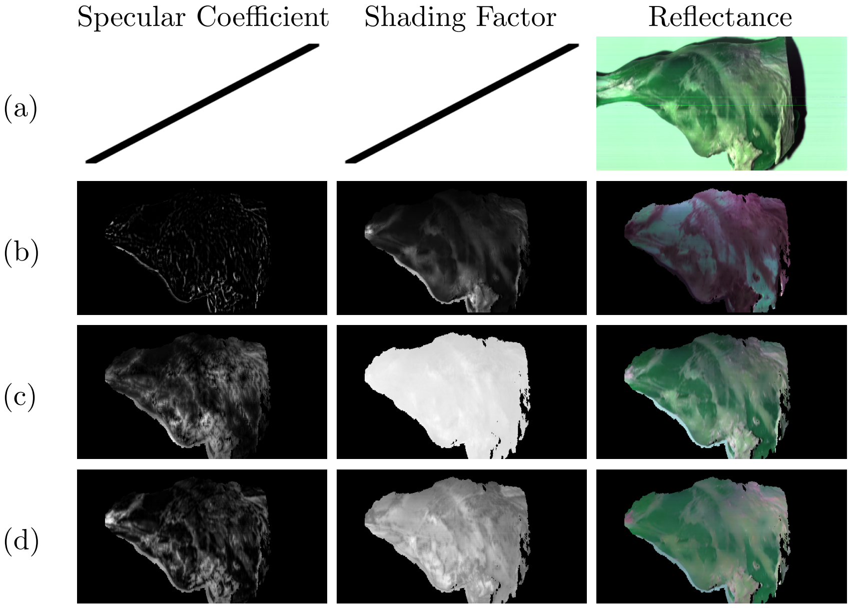


Figure 5.6: Outputs for the reflectance estimation of the lamb cut sample. In the columns: (left) specular coefficient, (middle) shading factor, and (right) reflectance. Rows contain results from different methods: (a) calibrated, (b) Robles [83], (c) Krebs [87], and (d) our method. The reflectance is presented as a false colour image where channel bins with centre-wavelengths of 1114nm, 1217nm, and 1339nm are mapped to RGB. The calibrated method only recovers reflectance.

according to the DRM, which describe the reflected light off a surface as a linear combination of both specular and diffuse components. The proposed method solves reflectance estimation per pixel as two independent quadratic programming problems for specular and diffuse parameters. This is possible by including general smoothness constraints on a property called the shading factor, which comes from the diffuse component and describes the interaction between the light source directions and surface normals. Strict smoothness is not enforced thanks to the use of the SAM similarity measure and terms which incorporate the known shape and lighting information. The proposed reflectance estimation method is compared with existing optimisation-based methods using an unbiased spectral renderer, and was shown to be superior across various materials. Hardware experiments were conducted by scanning a lamb cut using the calibrated camera system on a rig. The estimated reflectance using the proposed method was seen to be more homogeneous in areas with the same material.

The reflectance is a property that can be used to characterise a material. It refers to the diffuse component of light that is reflected at the surface, which arises from light rays penetrating into the material and scattering internally. The penetration depth is quite small, typically around 10mm, and depends on the material and the wavelength of light. If we consider objects that have different layered materials within this penetration depth, the reflected diffuse light would depend on the scattering of light rays with multiple layers. Therefore, in the final chapter a case study is examined to investigate if subcutaneous fat depth on a lamb cut can be identified through hyperspectral imaging using estimated reflectance. The fat depth is acquired through a unique process using computed tomography (CT) scans, and different regression models are compared.

Chapter 6

Case Study: Subcutaneous Fat Depth Estimation of Lamb Cuts

Hyperspectral imaging presents an opportunity to analyse consumable products for spatially varying attributes that are not typically visible. It has the advantage of capturing imaging in a non-destructive, remote, and non-invasive manner. One such attribute on lamb carcasses is the subcutaneous fat depth which can relate to overall fat quantity. Existing methods to determine overall fat rely on subjective measures from experts. Therefore, the full hyperspectral pipeline from the line-scan hyperspectral camera system calibration in Chapter 4, light source modelling in Chapter 5, and reflectance estimation using known shape and lighting in Chapter 6, are used to build regression models of subcutaneous fat depth with estimated reflectance. The ground truth fat depth is acquired from a novel process using computed tomography (CT) scans. A comparison between pixel-wise and hypercube-wise regression models is explored for this problem.

Consumable products must follow strict guidelines in regards to composition, nutritional value, quality and contamination. Some of these attributes relate to the chemical compositions of the product, and either cannot be seen in the visible spectrum of light or are internal features. Investigating such attributes of consumable products is paramount in the field of food science which generally rely on chemical analysis techniques. These techniques tend to be destructive to the samples, invasive, time consuming, and give overall averaged measurements [22, 91, 92]. Fat content analysis of animal products, such as beef or lamb, can be carried out using chemical techniques but only using small samples of the product and would render it unusable. X-ray imaging could be used for large cuts and whole animals, but is impractical, expensive, and there is a risk of radiation damaging the product due to overexposure. The current practise for whole carcasses grading is to use a single measurement of carcass fat at the 12th rib, known as GR8 fat depth measurement. This is a single point measure that correlates to overall fat deposition on the animal, carried out by an expert assessor through point fat depth measurements and visual inspection of shape.

Imaging techniques can be applied through illuminating materials and measuring the reflected light after it has interacted with surface and sub-surface matter. The visible portion of the electromagnetic spectrum typically limits the information to the surface level. Hyperspectral imaging capture hypercubes in the near-infrared (NIR) range of the electromagnetic spectrum which can provide subsurface features, that are typically not visible, which are encoded in its measured spectral bands. The measured hypercube depends on a number of factors such as illumination and surface shape, as discussed in Chapter 5, estimating the material property of reflectance provides a hypercube that is invariant to these factors.

While using the reflectance as a feature removes variations causes by external factors, the use of a large feature space (i.e. the number of spectral bands) impacts the training time and require a large model complexity. Features that show no correlation with the output fat depth may degrade the models. Therefore, dimensionality reduction of the hypercube is required prior to training regression models. Given that we are building a fat depth model of a real lamb cut, there will be imbalances in the dataset as the fat is not distributed equally. This will have to be considered when comparing the fat depth regression models.

6.1 Literature Review

In the literature, the problem considered in this study case was first investigated using lamb short loin cuts [82], which built a Gaussian process (GP) regression model for fat depth on the surface of the short loin using estimated reflectance. The reflectance here did not consider varying illumination and surface shape. Therefore, the fat depth models that were trained on one cut could not generalise to other samples. Prior to this work, others have captured the influence of subcutaneous fat depth on NIR bands through spectroscopy of tissue phantoms which replicate human skin [93] and Monte Carlo simulations of human tissue [94] up to fat depths of 12mm and 15mm respectively. Subcutaneous fat depth of humans has been estimated using point NIR light emitting diode (LED) sensor devices, where ground truth fat depth of participates can be acquired from CT scans [95] and skin-fold caliper measurements [96]. The fat depth models were accurate up to 20mm, but required close proximity of the sensor to the surface and multiple measurements at varying light source distances from the measured point. In most cases, point measurements are captured to represent the fat depth in local areas, the surfaces examined are assumed to be flat, sensor-to-surface and sensor-to-light distances are fixed and generally small, and the incoming light direction and intensity are not considered.

The case study can be considered a form of meat quality assessment. Therefore, existing methods for quality assessment must be explored to understand their disadvantages. These methods can be grouped into manual measurements carried out by experts, alternative imaging technologies such as X-ray, and existing spectral imaging technologies, which includes hyperspectral, that have been used for a diverse range of meat assessment applications.

For applications that use spectral imaging, most involve building supervised machine learning models of the desired attribute in the meat using reflectance data. Recent developments have focused on applying deep learning-based models to hyperspectral imaging problems, after their considerable success with red, green, and blue colour channels (RGB) vision problems.

6.1.1 Manual Measurements

Non-invasive techniques for measuring body fat of a carcass are necessary to ensure minimal disruption of specimens, which will reduce the chance of contamination from chemicals or bacteria and help to maintain the strict food standards for the meat [18]. The most common non-invasive technique used to assess fat of live animals are subjective measurements made by an expert assessor. One of these measurements is known as the body condition score (BCS), which is a meat quality estimator that observes the fat reserves present in live animals [97]. It is a highly subjective measure where the assessor uses a five-point scale system that is visually determined from the shape of the animal's back and pelvic regions [98]. Such methods are vastly inaccurate, time-consuming, and require expert individuals who could be subject to bias [98][99][97][14].

A fat measure specifically for lamb and sheep carcasses is the girth rib (GR) site [100], which is performed by measuring the depth of tissue at the 12th rib 110mm from the mid-line of the sheep carcass. Multiple studies that have been reviewed have shown some significance in predicting lean meat yield of lamb carcasses when using the single point GR site measurement as a predictor [101]. Recent experiments have shown that estimating overall carcass fat using both the GR site and hot carcass weight is highly variable [102]. The inclusion of other site measurements, with the GR site and hot carcass weight, have shown to reduce the variability [103]. Irrespective, these site measurements require direct contact with the carcass which introduces the potential for contamination of the meat, and GR site generally has low correlation to overall metrics for low fat animals [104, 105].

6.1.2 Alternative Imaging Technologies

Forms of precise measurement technology are necessary to give consistent objective measures not possible by humans. Existing imaging technologies such as dual-energy X-Ray absorptiometry (DEXA), CT and magnetic resonance imaging (MRI) are designed specifically for capturing internal compositions. They have had years of development and research prior to hyperspectral imaging, and produce reliable and accurate scans making them ideal for estimating subcutaneous fat depth [106][82]. There are major disadvantages that make them impractical for such purposes. The initial cost for the equipment is vastly expensive as these machines are complex and require additional safety measures when installed. Running the scans on specimens is also a time-consuming process [18][82]. DEXA and CT use X-Ray technology, where there is a risk of over-exposure to harmful X-Ray radiation to the carcass and surroundings [18]. These reasons make such technologies more suited to strict laboratory environments which adhere to health and safety protocols. Ultrasound scanners have also been developed for farm use as live animals can be safely scanned while they stand [106]. The performance of these scanners for determining animal composition has been investigated for many years, such as in determining the fat depth of lambs [107][108] and horses [99]. The issues with using ultrasound scanners is that they are not accurate compared to other imaging technologies, require direct contact, and there is difficulty in differentiating between different layers of tissue.

6.1.3 Spectral Imaging Technologies

Spectral imaging encompasses hyperspectral and other forms of imaging, such as multispectral and Raman spectroscopy, that capture reflected or scattered light from a material which is typically outside the visible range of the electromagnetic spectrum. There have been numerous works on meat analysis using spectral imaging due to the large number of spectral bands and the interaction of NIR light with common molecular bonds that are present in the chemical composition of the meat [22][109]. Classification problems using spectral imaging focus on detecting subtle differences across a single type of animal meat. This includes the identification of different types of muscle tissue for lamb [12], separating free-range and caged chicken meats [110], and determining the diet of ether grass or grain for cows, from their beef carcasses [111]. For these problems support vector machine (SVM) or partial least squares discriminant analysis (PLS-DA) models are trained and tested. Regression problems consider characteristics of meats such as pH [112], moisture [91], fat content [22], tenderness [112], protein content [91], and salt content [113][15]. The adulteration of beef [109] and lamb [20] have also been formulated as regression problems that are able to accurately determine the proportion of added meats. For these regression problems, partial least squares regression (PLS) models demonstrated the best fits and highest accuracy. All approaches have a common dimensionality reduction process typically using principal component analysis (PCA).

Recent years have seen a push towards learning approaches, based on deep neural networks, after their massive success on RGB computer vision problems such as segmentation [114] and object detection [115]. In these works a new class of learning models known as convolutional neural networks (CNNs) have been developed specifically to process image data using the convolution operation. As opposed to non-deep-learning models, CNNs are trained with either patches or the entire image, and are naturally designed to account for spatial correlations between pixels. Early work on hyperspectral classification using CNNs achieved comparable results to commonly used non-learning models [116] by training two separate CNNs which extract spectral and spatial features. A CNN that extracted both spectral and spatial features together was found to outperform existing methods [117][118][119]. More recently efficient CNNs, such as MiniNet-v2 [120], that have shown to work well with RGB autonomous driving datasets using fewer parameters, show promising results with hyperspectral imaging [121]. The main problem with using deep-learning based models for hyperspectral imaging is the limited labelled samples that are available for training [116, 118, 119]. Many look to data augmentation to avoid any over-fitting of their models.

6.2 Problem Statement

Given a calibrated camera system of a line-scan hyperspectral camera $\mathcal{H}$, a 2D frame colour depth channels (RGB-D) camera $\mathcal{F}$, and a modelled non-isotropic near-field disk light source $\mathcal{L}$ as shown in Figure 6.1. A lamb cut sample is captured with the camera system assuming that its surface reflectance properties are described by the dichromatic reflectance model (DRM), and CT scans of the cut are used to determined ground truth fat depth labels. The aim of this work is to estimate fat depth in lamb cuts using the measurements from $\mathcal{H}$. The fat depth estimation is formulated as a regression problem using estimated reflectance, which is acquired from the radiance as outlined in Chapter 5. The estimated reflectance is in the form of a 3D hypercube $\in \mathbb{R}^{u \times v \times \lambda}$ where u and v are its spatial dimensions in pixels, and λ is spectral bands. The predicted fat depth is a 2D matrix $\in \mathbb{R}^{u \times v}$ and the regression of the fat depth is defined as the estimation of a function $f : \mathbb{R}^3 \mapsto \mathbb{R}^2$.

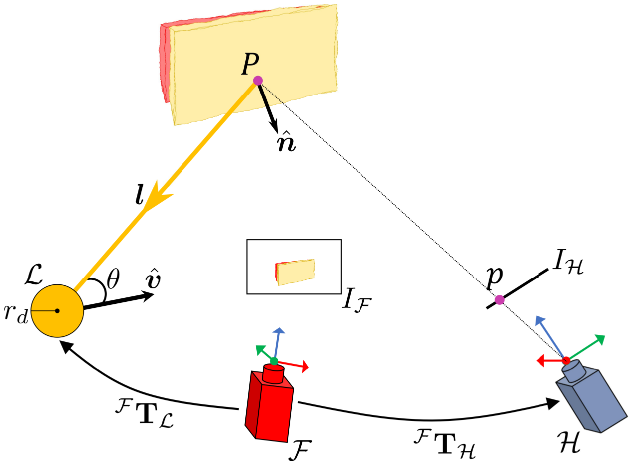


Figure 6.1: A calibrated camera system captures captures a lamb cut surface where its reflectance properties are modelled using the dichromatic reflectance model (DRM). The calibrated camera system is comprised of a RGB frame camera $\mathcal{F}$ and a line-scan hyperspectral camera $\mathcal{H}$, that are rigidly mounted by the homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{H}}$. A modelled non-isotropic disk light source $\mathcal{L}$ of radius r_d, with the principal direction vector $\hat{\boldsymbol{v}}$, is rigidly attached to $\mathcal{F}$ by the homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{L}}$. The light vector $\boldsymbol{l}$, which is a zenith angle θ from $\hat{\boldsymbol{v}}$, carries radiant flux to the surface point P with surface normal $\hat{\boldsymbol{n}}$, but $\boldsymbol{l}$ is defined as originating from P. The radiant flux is reflected off P as radiance in both a diffuse manner (red arrows) and a specular manner (blue arrow). The diffuse and specular radiance in the direction of $\mathcal{H}$ is measured by p on the image $I_{\mathcal{H}}$.

6.3 Hyperspectral Reflectance

The reflectance estimation follows the methods of Chapter 5. In Figure 6.1, a lamb cut sample whose reflections are described using the DRM, is captured by $I_{\mathcal{H}}$ which is the line image of the line-scan hyperspectral camera and $I_{\mathcal{F}}$ represents both the 2D colour and depth images of the frame camera, where the depth image is aligned to the colour image. The camera system is fully calibrated as described in Chapter 3. The fixed homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{H}} \in \mathbb{R}^{4\times4}$ is of the hyperspectral camera with respect to $\mathcal{F}$ which comes from the calibrated extrinsic parameters. The non-isotropic near-field disk light source $\mathcal{L}$ has a principal direction vector $\hat{\boldsymbol{v}}$ with maximum radiant intensity, a radius of r_d, is rigidly attached to $\mathcal{F}$ by the known homogeneous transformation ${}^{\mathcal{F}}\mathbf{T}_{\mathcal{L}}$, and its irradiance is spatially modelled using the GP with light-src-mean function as described in Chapter 4. The light vector $\boldsymbol{l} \in \mathbb{R}^3$ represents the ray that is emitted by the light source and arrives at the 3D surface point P with the surface normal $\hat{\boldsymbol{n}} \in \mathbb{R}^3$. $\boldsymbol{l}$ is defined as originating from P in the direction towards $\mathcal{L}$ to ensure the dot product between $\boldsymbol{l}$ and $\hat{\boldsymbol{n}}$ is positive [61].

The process to estimate the reflectance for lamb cuts is discussed in the following sections which comes from Chapter 5.

6.3.1 Hyperspectral Ray Casting

For the radiance measurement at p by hyperspectral camera, the 3D coordinate P must be determined in order to estimate surface normals and the irradiance from the light source. The point P can be determined by the intersection of the vector $\boldsymbol{P}$ with the surface of the lamb cut that is being viewed. Given that there is a depth image of the sample from $\mathcal{F}$, the corresponding point cloud can be easily acquired. A 3D occupancy grid map is then constructed using the point cloud, which can then be used for ray casting as shown in Figure 6.2. Once P is acquired, the camera calibration parameters of the hyperspectral camera can be used to re-project the hyperspectral measurements to I_F as shown in Figure 6.3.

6.3.2 Estimating the Irradiance

The irradiance E arriving on the surface of the sample is assumed to only come from light source $\mathcal{L}$ and is estimated for all measured spectral bands λ by inference of the modelled GP distribution of $\mathcal{L}$ from Chapter 4. Each spectral band has a separate GP model with its optimised hyperparameters and the irradiance is returned as a hypercube $\in \mathbb{R}^{u \times u \times \lambda}$.

A map of the normalised irradiance for a lamb cut sample is shown in Figure 6.4 for a single spectral band. Alongside the irradiance, the light source direction vectors $\hat{\boldsymbol{l}}$ can be estimated for all P using the known pose of the light source. By calculating the dot product between $\hat{\boldsymbol{l}}$ and $\hat{\boldsymbol{n}}$, a property that is proportional to the shading factor from the DRM can be acquired as described in Section 5.3. Maps of the $\hat{\boldsymbol{l}}$, $\hat{\boldsymbol{n}}$ and their respective dot product are shown in Figure 6.5.

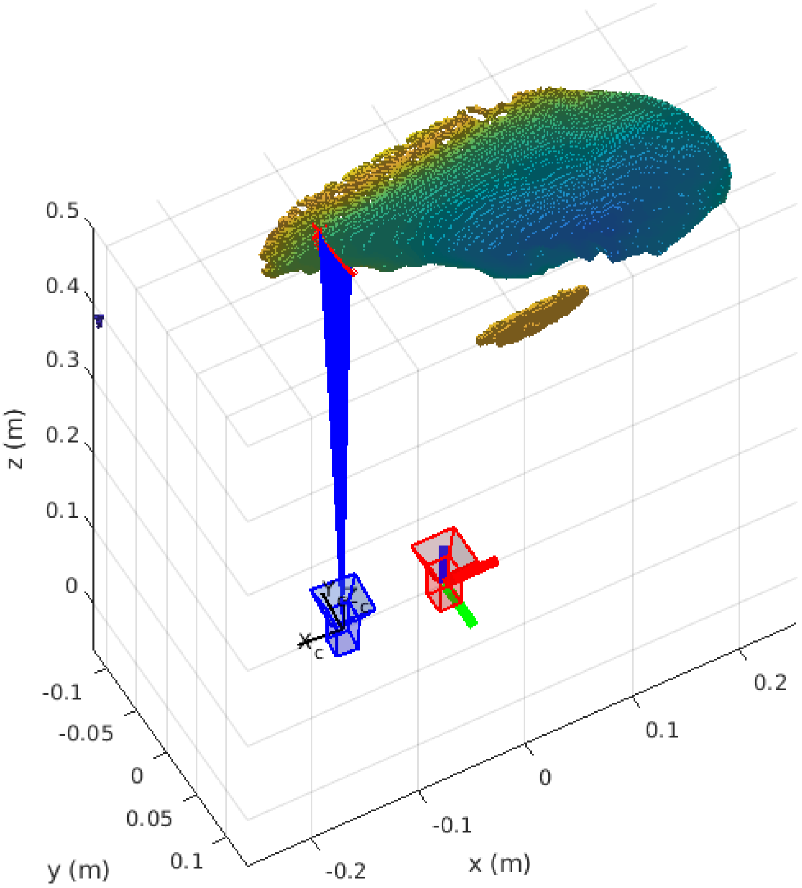


Figure 6.2: Ray casting 3D occupancy grid map of lamb cut. The camera in red is the frame camera $\mathcal{F}$ which is the world coordinate system. The blue camera is the line-scan hyperspectral camera $\mathcal{H}$. The rays originating from the the hyperspectral camera are shown in blue and the intersection points on the cut are shown as red crosses.

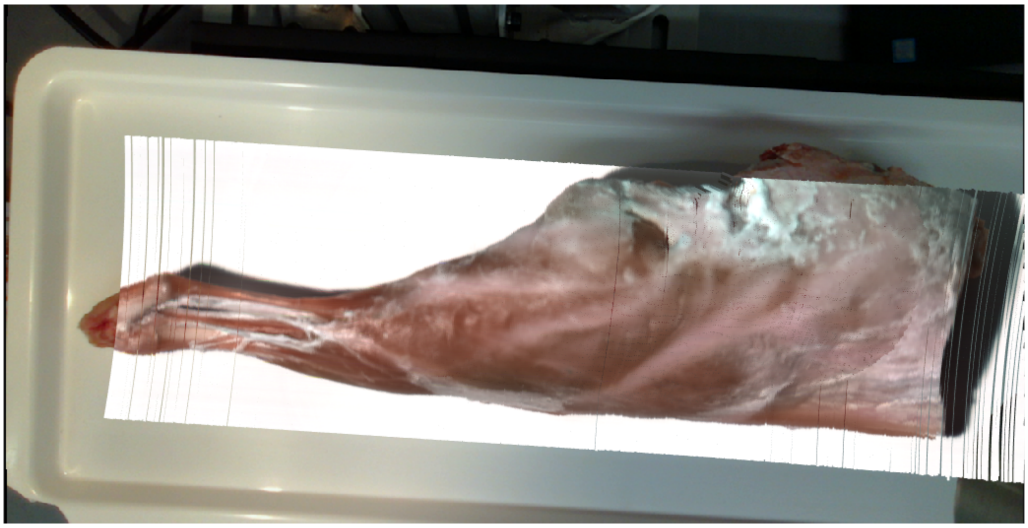


Figure 6.3: Re-projected hyperspectral radiance false colour image lines overlaid onto the RGB image of the left leg lamb cut. The false colour radiance comes from channel bins with centre-wavelengths of 1114 nm, 1217 nm, and 1339 nm being mapped to RGB.

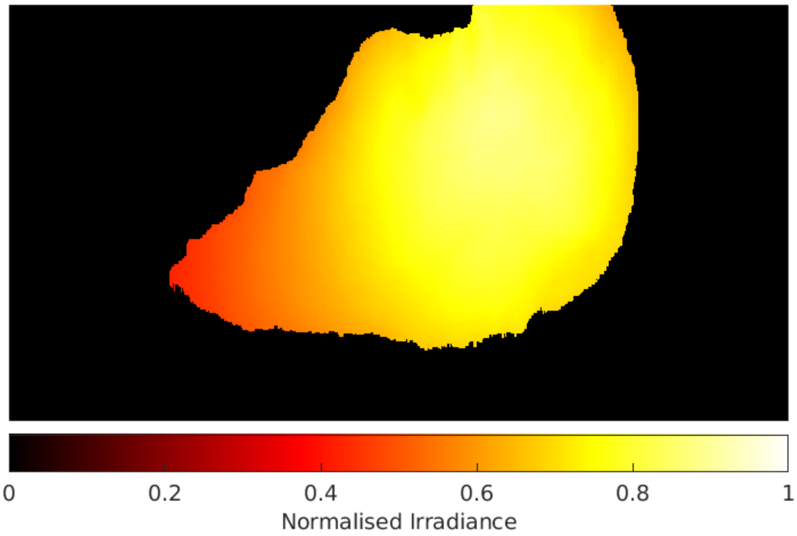


Figure 6.4: Normalised irradiance only from the light source $\mathcal{L}$ arriving at the surface of the left leg lamb cut sample, which is estimated from the modelled GP model for a single spectral band.

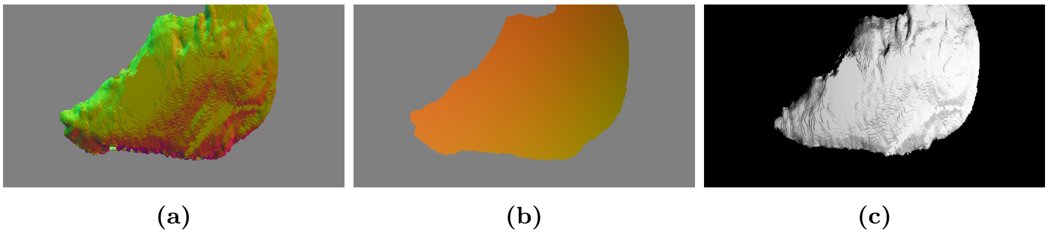


Figure 6.5: The variation in shape and lighting for the left leg lamb cut sample. (a) The surface normals, (b) the light source direction map, and (c) the resulting dot product of (a) and (b) which relates to the shading.

6.3.3 Estimating the Reflectance

The reflectance is estimated using three different methods: calibration, Krebs [87], and the proposed method from Chapter 5. The last two methods are formulated as an optimisation problem by assuming the reflections of the sample material are described by the DRM. Krebs only requires irradiance, while the method from Chapter 5 requires irradiance, surface normals, and light source direction vectors. The resulting reflectance is calculated for two different lamb cut samples using the three methods. The false colour images of the reflectances are shown in Figure 6.6. The reflectance estimated by the method from Chapter 5 is found to show the least texture, which is expected from the reflectance property.

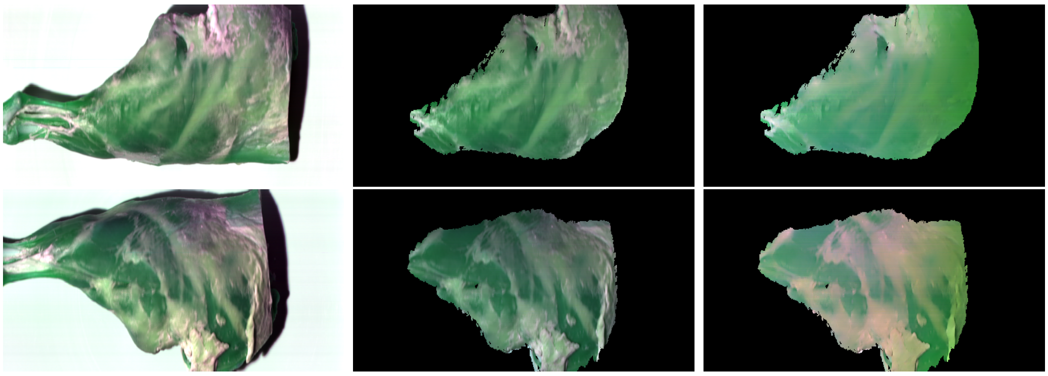

(a) (b) (c)

Figure 6.6: Outputs for the reflectance estimation of both lamb cut samples, top row is for the left cut and bottom row is for the right cut. Columns: (a) calibrated, (b) Krebs [87], and (c) method from Chapter 5. The reflectance is presented as a false colour image where channel bins with centre-wavelengths of 1114 nm, 1217 nm, and 1339 nm are mapped to RGB.

Now with estimated reflectance for the lamb cuts, the next section discusses how the corresponding ground truth fat depth is determined, using the CT scans, for every pixel in the reflectance hypercube.

6.4 Fat Depth Ground Truth

The ground truth fat depth of the lamb cut samples is obtained through CT scans. A CT measures the X-rays attenuation and is an ideal modality for distinguishing muscle from fat components [122, 123]. The CT measurements are in the form of a series of sequential 2D slices. They are stacked into a 3D tensor to build a 3D grid of the X-ray absorption. The voxels have a size of $0.56 \times 0.56 \times 0.7$ millimetres and are stored as DICOM images.

6.4.1 CT Reconstructions

To acquire the subcutaneous fat depth of the lamb cut samples and align CT scans to the data of the camera system shown earlier, 3D reconstructed meshes were created from the 2D CT scan slices. This is achieved using marching cubes algorithm with thresholds on the voxel absorption [124]. Three meshes are created which isolate the fat, muscle and outer contour of the lamb cut. The fat and muscle are shown together in Figure 6.7.

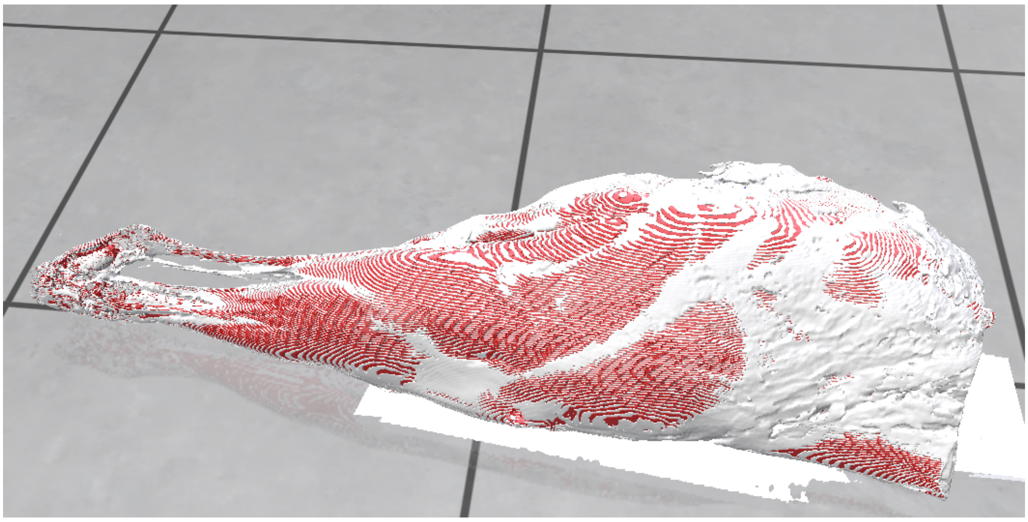


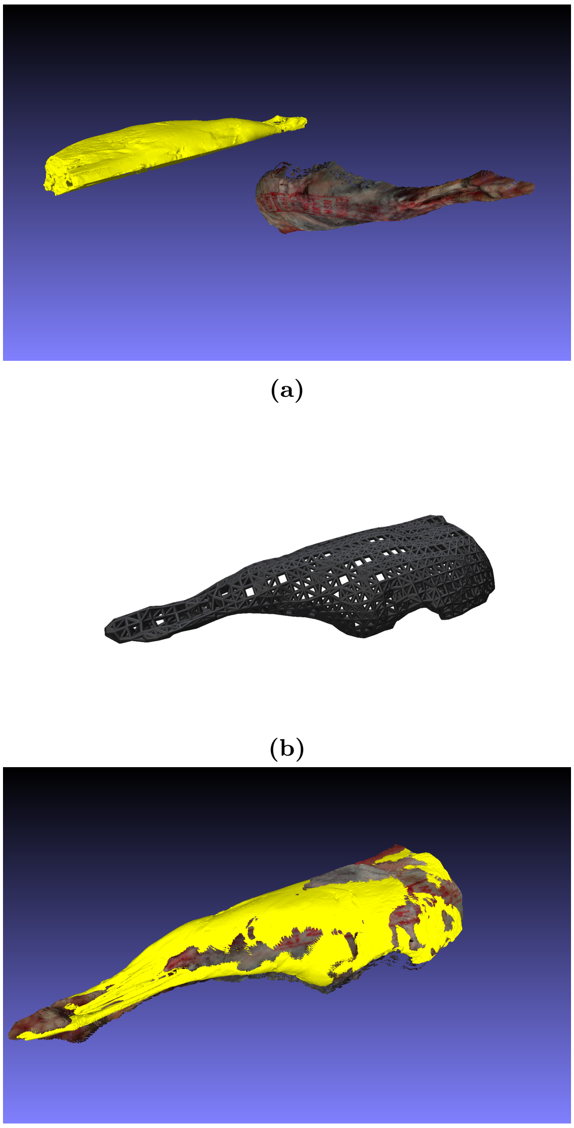
Figure 6.7: Surface reconstruction of the muscle (in red) and fat (in white) from the computed tomography (CT) scans by thresholding the CT data and applying marching cubes [124] visualised using the Polyscope viewer [125].

6.4.2 CT and Depth Alignment

The same sample has been processed separately using a CT scanner and the camera system. The CT scans are reconstructed into meshes and the RGB-D camera from the camera system allows us to create a 3D depth reconstruction. Ideally, these two separate 3D models of the sample should perfectly align in a rigid manner. However, this is not what happens in practice, as slight deformations occur when transporting the sample between the hyperspectral rig and the CT scanner. Further sagging is caused by changes in temperature and moisture. To solve this problem, the outer contour is aligned to the depth reconstruction in a non-rigid manner by using embedded deformation [126] with manually selected 3D correspondences. The deformation is then applied to the fat and muscle meshes to align them to the depth reconstruction. The alignment of the CT fat mesh to the depth reconstruction using the deformation is shown in Figure 6.8.

6.4.3 Fat Depth Ray Casting

The fat depth must be acquired for all radiance measurements at p. This is estimated by ray casting the ray ${}^{\mathcal{H}}\boldsymbol{P}$, which comes from the 3D point P defined in the coordinate system of the line-scan hyperspectral camera $\mathcal{H}$, with the two aligned meshes of fat and



(a)

(b)

(c)

Figure 6.8: Non-rigid alignment of computed tomography (CT) mesh to depth reconstruction using embedded deformation. (a) misaligned CT fat mesh shown in yellow and depth reconstruction, (b) deformation graph of CT fat mesh, and (c) aligned CT fat mesh to depth reconstruction.

Table 6.1: Different outcomes of ray casting onto the CT meshes.

P_f Found	P_m Found	$d_f < d_m$	Outcome
×	×	-	No intersection
×	✓	-	Only muscle
✓	×	-	Only fat
✓	✓	False	Fat under muscle
✓	✓	True	Fat depth

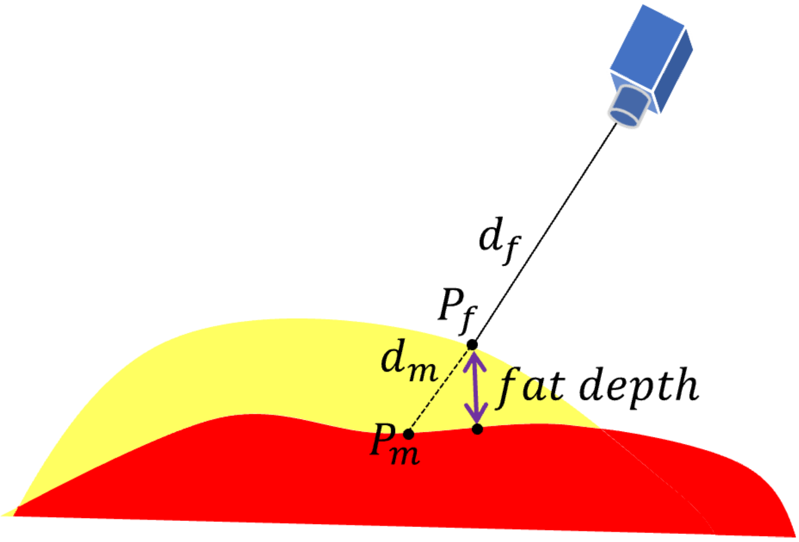


Figure 6.9: Ray casting from the hyperspectral camera to the fat (yellow) and muscle (red) CT meshes showing the intersection points and their corresponding distances. The actual fat depth is the distance highlighted by the purple arrow.

muscle resulting in the intersection points P_f and P_m. If the intersections are valid, the distances between the hyperspectral camera and the fat and muscle intersection points are determined as d_f and d_m respectively. There are five outcomes that can occur based on the combinations of intersections and the calculated distances. These are summarised in Table 6.1. For a ${}^{\mathcal{H}}\boldsymbol{P}$, where both muscle and fat intersections occur and $d_f < d_m$, the fat depth is calculated by performing a nearest-neighbour search on the muscle mesh at P_f as shown in Figure 6.9. The results of the ray casting are visualised on the meshes in Figure 6.10.

Figure 6.10: Visualisation of the ray casting on a lamb leg cut from the hyperspectral pixels to determine subcutaneous fat depth, using the Polyscope viewer [125].

6.5 Experiments

In this section, we evaluate the capability of four machine learning models for performing fat depth regression on the lamb leg reflectance dataset from the hyperspectral camera, and compare the regression models for three different reflectance estimation methods.

6.5.1 Evaluation Protocol

The evaluation uses the data of the two cuts of lamb shown in Figure 6.6 which is referred to as the lamb leg dataset. One cut is used for training and the other for testing. The different models are evaluated with the same pair of training and testing hypercubes. The hypercubes for each cut have a size of 320 × 599 × 164, where each has approximately 66000 valid pixels for which reflectance and fat depth are known.

Performance is measured using two metrics. First, we consider the mean absolute error (MAE), which quantifies the overall prediction error in the fat depth by averaging across all pixel. In addition, we measure the normalised area under the curve (AUC) of the fat depth prediction error with respect to the ground truth fat depth, which is computed by averaging the absolute error at different depth intervals. As such, the AUC quantifies the

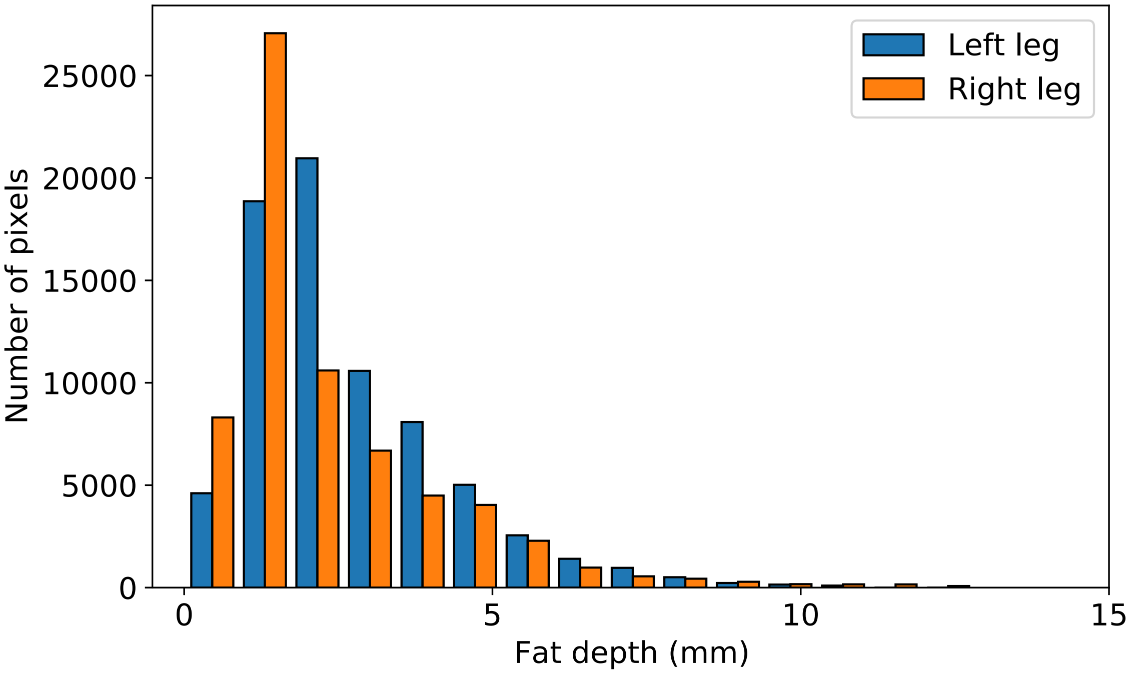


Figure 6.11: Histogram of fat depth in millimetres for both lamb leg samples showing the imbalance in the data.

error on binarised samples, which counteracts the imbalance in fat depth values across both samples as seen in the long-tail of the fat depth distribution shown in Figure 6.11.

6.5.2 Fat Depth Regression Models

Four different machine learning models of linear regression, multilayer perceptron (MLP), ERFNet[127], and MiniNet-v2 [120] are trained on the lamb leg dataset where their number of training parameters is summarised in Table 6.2. The simplest model is linear regression which we treat as a baseline. By examining Figure 6.11, it can be seen that a significant portion of the observed fat depth is very shallow (approximately 1–3 mm) with only a few regions exhibiting deeper fat content (up to 14 mm). It is expected that linear regression can model the depth well for the majority of shallow measurements, but at the expense of incorrect predictions for the deeper fat regions. A MLP is also evaluated as this provides higher capabilities to model the non-linearity of the deeper fat. The MLP consists of two hidden layers with 16 and 10 fully connected neurons, where each layer employs a rectified linear unit (ReLU) as activation function. Both linear regression and MLP learn from pixel reflectance inputs where the input sample is a λ-dimensional normalised feature vector.

Table 6.2: Number of training parameters for each machine learning model based on input hypercube dimensions of 320 × 599 × 20.

Machine learning model	Number of parameters
Linear regression	21
MLP	511
ERFNet [127]	2,095,165
MiniNet-v2 [120]	481,889

To process a full image and estimate the fat depth for each pixel, while leveraging local spatial correlations given limited data samples, we adopt two efficient lightweight CNN-based architectures of ERFNet [127] and MiniNet-v2 [120]. Their architectures are shown in Figure 6.12. ERFNet has an encoder-decoder architecture which is typically used for semantic segmentation of RGB images. Its novelty comes from sequential hidden layers that are comprised of skip connections, to mitigate the vanishing gradient problem and improve model performance in regards to accuracy [128], and convolutions with 1D kernels that significantly reduce computation compared to 2D kernels allowing for real-time performance. Although, the original ERFNet model is not designed for higher dimensional hyperspectral images, slight modifications to its architecture have allowed for training of higher dimensional images with comparative performance [129]. MiniNet-v2 similar to ERFNet was originally developed for semantic segmentation of RGB images, and has been shown to perform well on autonomous driving datasets with a significant reduction in parameters compared to other state-of-the-art architectures. It makes use of multi-dilation depth-wise convolutional layers and an additional convolutional branch rather than skip connections, which offer a good trade-off between accuracy and efficiency. Recently, MiniNet-v2 was applied to hyperspectral images for a remote sensing dataset [121]. For performing fat depth regression using ERFNet and MiniNet-v2, we modify the architectures by altering the final 2D convolutional upsample layer to produce a single value for each pixel of the image.

6.5.3 Dimensionality Reduction

Due to the high dimensionality of hyperspectral data, it is common to apply band selection or reduction techniques. These methods reduce its overall size, but can also remove redundancies that are detrimental when training machine learning algorithms. In line

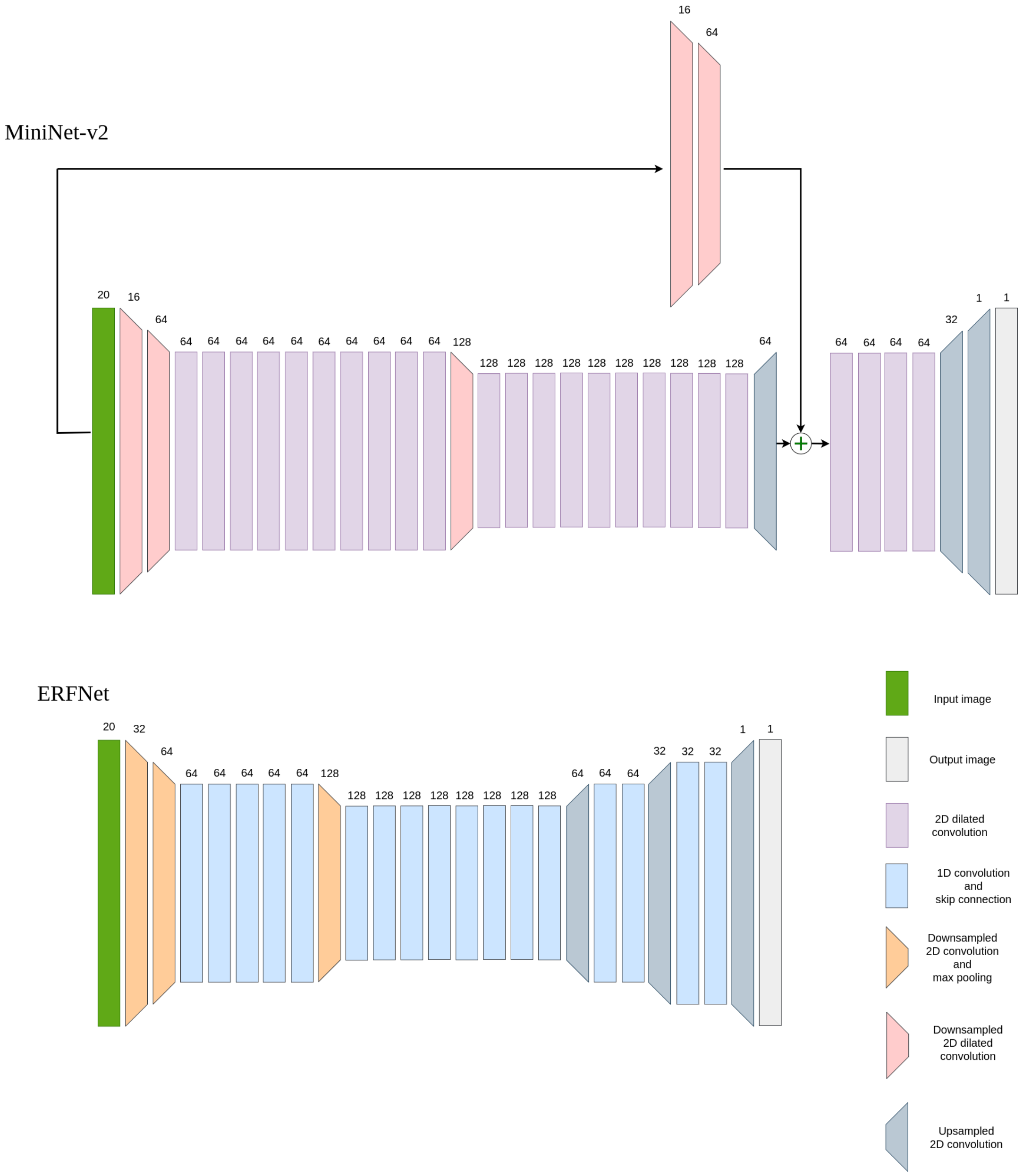


Figure 6.12: The deep learning architecture diagrams for ERFNet [127] and MiniNet-v2 [120] which are the CNN models trained using lamb leg dataset. The number above each block represents the feature size of the output hypercube for that block. For both models, the input hypercube has feature size of 20 and the output has feature size of 1 corresponding to the predicted fat depth for each pixel.

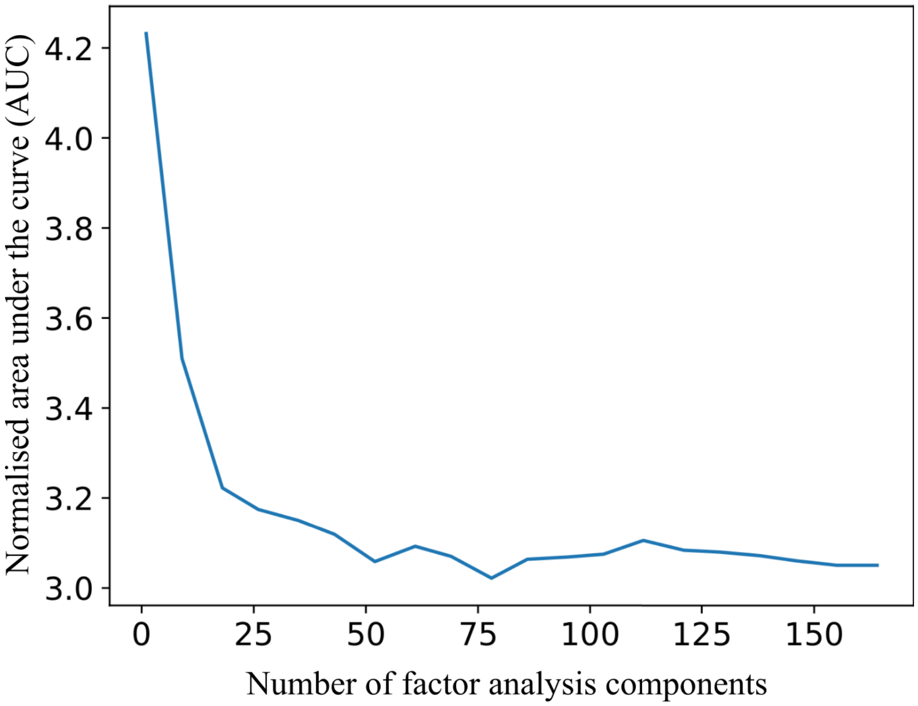


Figure 6.13: The normalised area under the curve (AUC) of linear regression fat depth model, trained using estimated reflectance from the method in Chapter 5, with a varying number of factor analysis components.

with [121], this work applies factor analysis to reduce the data. Figure 6.13 shows the performance of the linear regression model trained with a varying number of components. It is clear the performance gain beyond 20 components is small, thus we use this value for all experiments as it is a good trade-off between performance and data size.

6.5.4 Data Augmentation

MiniNet-v2 and ERFNet both take the reduced hypercube after dimensionality reduction as input. Since only one hypercube is available for training, we apply data augmentation to aid in model generalisation and improve accuracy [130]. The list of augmentation transformations are summarised in Table 6.3. During each epoch, the augmentations are first applied to the reduced hypercube, and then the augmented hypercube was used for training the models as shown in Figure 6.14.

Table 6.3: Data augmentation transformations applied to lamb leg dataset for training ERFNet [127] and MiniNet-v2 [120] hyperspectral fat depth regression models.

Augmentation type	Parameter	Value
Random rotation	Angle	$[-180°, 180°]$
	Interpolation	Bilinear
	Fill	Zeros
Random crop and resize	Crop scale	$[0.2, 1.0]$
	Interpolation	Bilinear
Random vertical flip	Probability	0.5
Random horizontal flip	Probability	0.5

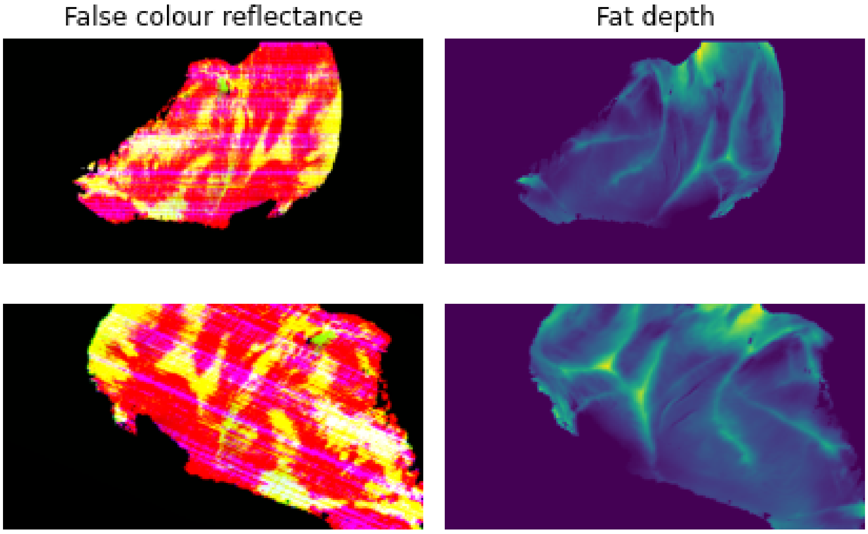


Figure 6.14: Visualisation of the augmentations applied to the training data of the lamb leg dataset for training ERFNet and MiniNet-v2. The left column shows the false colour reflectance image, where the first three factor analysis components have been mapped to RGB. The second column shows the fat depth in millimetres. The top row is the original data, and the bottom row represents one instance of applied augmentations.

6.5.5 Fat Depth Model Results

The process of training the fat depth models first involved normalising the training data between 0 and 1, where the transformation was recorded such that it could be applied to the testing data. Then dimensionality reduction was carried out as described in the previous section which reduced the spectral dimensions from 164 to 20. The MLP, ERFNet and MiniNet-v2 were all trained using the stochastic gradient descent (SGD) algorithm with the L2 loss criterion. The hyperparameters, which include the learning rate, momentum and weight decay, were all tuned using the Ray Tune library [131]. This tuning was

Table 6.4: The mean absolute error (MAE) in predicted fat depth measured in millimetres of the lamb leg dataset. The rows from top to bottom represent three reflectance methods: calibrated, Krebs [87], and the proposed method from Chapter 5. The columns from left to right are four regression models: linear regression, multilayer perceptron (MLP), ERFNet [127], and MiniNet-v2 [120]. Highlighted in bold is the smallest value of the metric for each regression method.

Reflectance method	**Linear regression**	**MLP**	**ERFNet** [127]	**MiniNet-v2** [120]
Calibrated	1.13	1.30	1.40	0.98
Krebs [87]	0.99	1.25	0.98	1.47
Proposed (Chapter 5)	**0.68**	**0.83**	**0.79**	**0.89**

carried out by splitting the training hypercube image using a 8:2 split ratio into training and validation datasets, and then resizing the images back to the original size. The best set of hyperparameters for each model was used to train on the fat depth regression data for 500 epochs. The batch size of the MLP was fixed at 200 samples. The linear regression model was fit using ordinary least squares.

Tables 6.4 and 6.5 present quantitative MAE and AUC results respectively for fat depth regression on the tested lamb leg for each regression model and each reflectance method. The absolute error in predicted fat depth across the lamb leg was shown in Figure 6.16 for each regression model and each reflectance method. The distribution of the MAE was plotted against the ground truth fat depth with 1-σ STD bounds as shown in Figure 6.15 for each regression model and each reflectance method. In general using the reflectance estimation proposed in Chapter 5 showed the least error in predicted fat depth regardless of the regression model. This is expected as it incorporates both lighting and shape information to estimate the reflectance. The deep learning-based approaches were shown to have the least error regardless of the reflectance data, and were able to best capture the regions with high fat depth, given the limited data samples. The performance of MiniNet-v2 was comparable to ERFNet with approximately a quarter of the training parameters (as given in Table 6.2). Interestingly, as shown in Figure 6.15 linear regression and MLP models performed well in regions with low fat depth (1 to 4mm) where there were significantly more samples for training, and generally had an increasing linear error in predicted fat depth in higher fat regions, highlighting their inability to capture the non-linear interactions between the spectral features.

Table 6.5: The normalised area under the curve (AUC) in predicted fat depth measured in millimetres of the lamb leg dataset. The rows from top to bottom represent three reflectance methods: calibrated, Krebs [87], and the proposed method from Chapter 5. The columns from left to right are four regression models: linear regression, multilayer perceptron (MLP), ERFNet [127], and MiniNet-v2 [120]. Highlighted in bold is the smallest value of the metric for each regression method.

Reflectance method	**Linear regression**	**MLP**	**ERFNet** [127]	**MiniNet-v2** [120]
Calibrated	3.41	3.24	2.02	2.59
Krebs [87]	2.72	3.82	1.99	2.95
Proposed (Chapter 5)	**2.23**	**2.28**	**1.65**	**1.61**

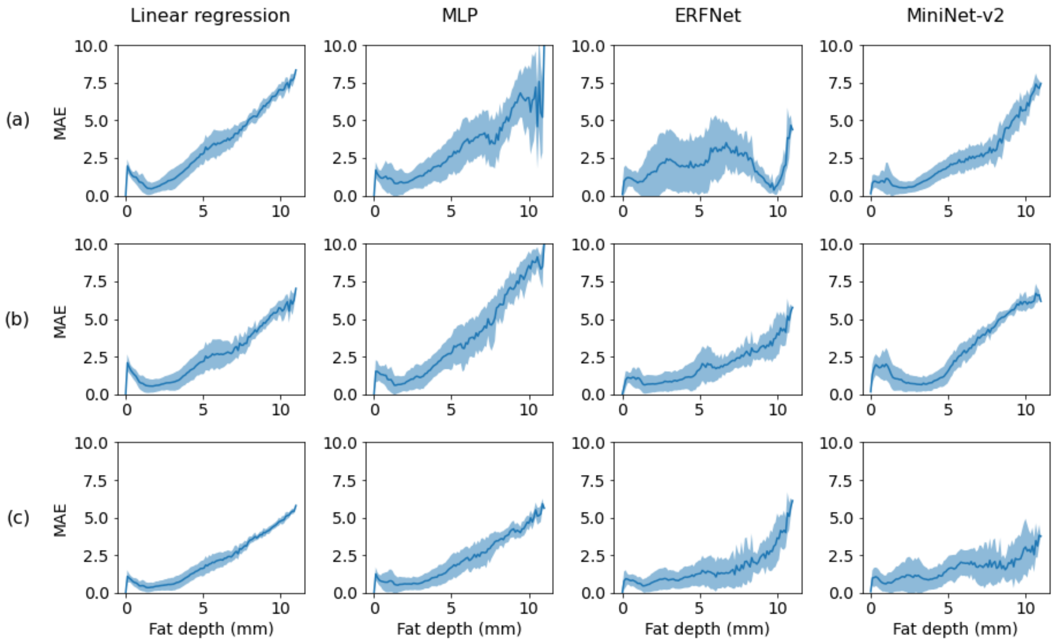


Figure 6.15: The plots of the MAE in predicted fat depth vs true fat depth with 1-σ STD bounds. The rows from top to bottom represent three reflectance methods: calibrated, Krebs [87], and the proposed method from Chapter 5. The columns from left to right are four regression models: linear regression, multilayer perceptron (MLP), ERFNet [127], and MiniNet-v2 [120].

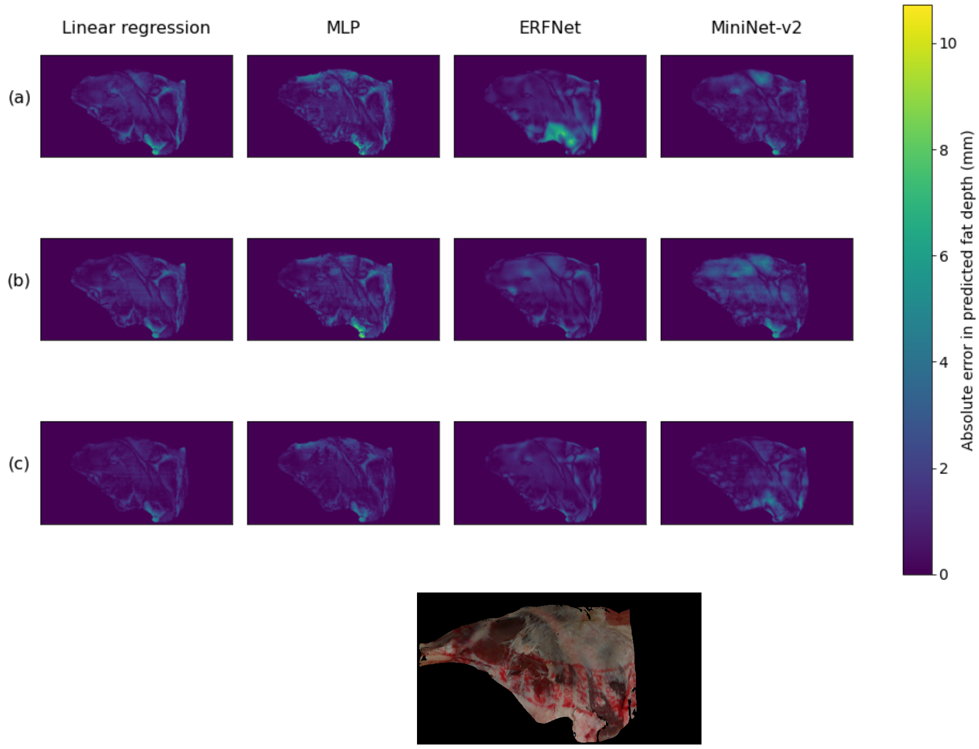


Figure 6.16: The absolute error in the predicted fat depth for the lamb leg dataset. The rows from top to bottom represent three reflectance methods: calibrated, Krebs [87], and the proposed method from Chapter 5. The columns from left to right are four regression models: linear regression, multilayer perceptron (MLP), ERFNet [127], and MiniNet-v2 [120]. A colour image of the tested lamb leg re-projected to the hyperspectral image is included for visualisation of fat and muscle. Note that a mask is applied to all hypercubes to remove the background or regions with no fat, therefore not all pixels are used for training and testing.

6.6 Conclusion

A case study investigates the use of hyperspectral imaging to build regression models of subcutaneous fat depth for lamb leg samples. The fat depth is acquired from CT scans. A single hypercube was used for training of a dataset largely imbalanced in terms of fat depth. The reflectance is estimated using three methods: one involves radiance normalisation, and the other two are optimisation-based that assume the material reflections can be described by the DRM, where one of the methods incorporates shape and lighting information as discussed in Chapter 5. For each hyperspectral reflectance pixel, the fat depth ground truth is obtained by ray casting on meshes reconstructed from CT scans that are non-rigidly aligned to a depth reconstruction. After applying a dimensionality reduction to the reflectance data, four fat depth regression models of linear regression, multilayer perceptron (MLP), and two deep learning-based CNNs are trained on a single lamb leg hypercube and tested on a separate lamb leg. It was found that, in general, the deep learning-based models best captured the imbalanced fat depth distribution. Furthermore, the optimisation-based reflectance method from Chapter 5, has the least error for fat depth estimation.

The case study demonstrates that subsurface material estimation in a near-field environment, here being subcutaneous fat depth of lamb cuts, is possible using hyperspectral imaging. Recovering the material property of reflectance attempts to overcome the variation in surface shape and irradiance across the tested samples. The method which is used to acquire reflectance will affect the performance of the trained models. The models are best trained using reflectance which is acquired through the knowledge of shape and lighting as per Chapter 5.

Chapter 7

Conclusions

This thesis presented an approach to hyperspectral imaging for near-field applications, which focuses on estimating an intrinsic material property from the measured hypercube data through prior modelling of the hyperspectral camera, light source and surface shape. An emphasis is placed on probabilistic modelling that propagates measurement uncertainty. The material property considered is diffuse reflectance which only varies over the spectral dimension. Much of the efforts are placed on acquiring reflectance that is truly invariant to surface shape and incident illumination. The property is assumed to be capable of distinguishing one material from another. Therefore, it is used to spatial model a desired attribute of an object over its surface.

In this final chapter, we summarise each of the four contributions, highlighting their significance relating to this research and the potential research questions that could be addressed.

7.1 Contributions

Active Line-scan Frame Camera System Calibration

In the calibration of a line-scan frame camera system, uncertainty propagation is shown to be an important element to this work as noisy pixel measurements are used throughout the calibration procedure. The active calibration algorithm looks to reduce the uncertainty in the final calibration parameters, by only considering measurements that will improve parameter estimation through the analysis of observability in the non-linear optimisation. This effectively removes images that are of similar pose to ones before it, reducing redundant measurements to avoid issues of insufficient parallax. The algorithm also introduces a strategy to terminate the calibration once observability is seen to reach a steady-state.

Working with the line-scan hyperspectral camera is challenging mainly due to the line-scan scanning method. Unlike 2D images from a frame camera, it is difficult to understand the scene through a single spatial dimension. Calibrating the line-scan hyperspectral with a frame camera made it possible to locate the view-line that is seen by the line-scan camera, which could then be easily re-projected onto a frame image to better comprehend its 3D position. Effectively the line-scan hyperspectral measurements can be considered as another modality of the frame camera, similar to how colour depth channels (RGB-D) cameras operate where depth maps are aligned to an red, green, and blue colour channels (RGB) image. This approach could be used for other unconventionally sensors to aid in understanding what is being captured

Given the difficultly to work with line-scan cameras, the active calibration algorithm provides some reassurance to filter images that are not useful for calibration. Such scenarios are common when capturing many images of the calibration board with the line-scan camera. The results of the experiments which compare the active algorithm with the naive approach that uses all images, show that a subset of images that are considered to have sufficient parallax, better estimate parameters than using all images. The algorithm highlights the importance of gathering proper image measurements for calibration, which is generally overlooked in existing calibration methods for any camera.

Probabilistic Light Source Modelling

The light source modelling focuses on capturing the irradiance distribution of the light source used with the hyperspectral camera in a near-field environment. A disk source with a non-isotropic and asymmetric radiant intensity distribution (RID) are used to model the irradiance of the light source using a data-driven Gaussian process (GP) model. This GP has a unique mean function that expresses the symmetrical properties of the light source through the disk source model, whilst the data-driven component is able to capture the inherent asymmetries that are present in the light source. Discrete measurements of a known surface are captured with the calibrated camera system in order to build the light source model. The GP model is shown to have the least error in its estimated irradiance whilst successfully capturing asymmetric features when compared to full parametric least squares fit and full data-driven GP with zero-mean function fit.

The RID of real light sources is asymmetric. Existing methods can model them with complex parametric functions, but would always assume radial symmetry. To capture the unknown asymmetry in the RID using a parametric model, the asymmetry must first be located, which requires the modelled RID. Rather than assume a fixed function, the GP model with the unique mean function provides a flexible distribution which uses the measurements to define the asymmetries in the RID, and the mean function to encapsulate the underlying properties of the disk source model. The experiments help to explain this further as measurements are captured in the limited viewing frustum of the hyperspectral line-scan camera. Outside this region, the GP model still provides reasonable irradiance values as it extrapolates using the mean function, but the variance is seen to be larger due to their being no nearby measurements. A standard data-driven GP with a zero-mean function cannot capture the properties of the source given the limited viewing frustum, and thus cannot be used to estimate the irradiance outside the viewing frustum.

The geometry of the light source causes variations in the light ray directions, and the asymmetries in the distribution cause variations in the irradiance arriving at a surface. The effects are highly dependent on the distance between the light source and the surface. Therefore, near-field experiments with the hyperspectral camera should consider these effects on the surface, especially when there are also variations in the surface shape, as the measured radiance by the camera depends on the incident irradiance.

Reflectance Estimation with Light and Shape Information

The reflectance estimation method, improves the estimation of diffuse reflectance from a hyperspectral radiance measurements by including terms for known surface normals and surface light ray directions to quadratic optimisation formulations. It assumes that the materials of an object being observed follow the dichromatic reflectance model (DRM), where radiance is reflected in both a diffuse and specular manner. The irradiance arriving at the surface is estimated from the GP light source model. The addition of the surface normals and surface light ray directions aid in constraining the estimated reflectance, when compared to existing optimisation methods that do not include it, which is demonstrated in experiments using a spectral renderer.

Recovering reflectance from the DRM is an underdetermined inverse problem which requires constraints. Existing methods place strict smoothness constraints on reflectance, surface shape and shading factor, which comes from diffuse radiance and describes the interaction of the light source with the surface shape. Smoothness of reflectance and surface shape is not guaranteed in general objects that are comprised of a mixture of materials. The shading factor can generally be considered smooth as long as there is no sudden change in surface normals, such as edges or corners. In the proposed method, the addition of the surface normals and light source direction vectors allow for soft smoothness constraints on the shading factor. Therefore, in the optimisation, sudden changes in surface shape are accounted by the shading factor and not confused with a sudden change in material. In terms of the near-field experiment with hyperspectral imaging, similar reflectance can be estimated for the same object regardless of the light source position and direction, and surface shape.

Case Study: Subcutaneous Fat Depth Estimation of Lamb Cuts

The use of hyperspectral imaging to estimate subcutaneous fat depth on lamb cuts is investigated. This involves building regression models of fat depth, which is determined from computed tomography (CT) scans, with estimated reflectance of the hyperspectral radiance measurements. This work relies on the previous contributions of light source modelling and reflectance estimation. The fat depth for each reflectance pixel is acquired through ray casting 3D reconstructed CT models that have been non-rigidly aligned to a depth reconstructed model. Different reflectance methods, including the proposed method of this thesis, were tested with different regression models. By training on one cut and testing on another, it was found that a deep learning-based model gave the best fit using the proposed estimated reflectance that incorporated light and shape information.

The results of the case study show that fat depth estimation is possible with hyperspectral imaging, and more importantly, the proposed light source modelling and proposed reflectance estimation methods improved the performance of the regression models. The case study highlights one of the major difficulties with using hyperspectral imaging, which is the limited datasets. In this work, the regression models were trained with pixels from only one lamb leg, which also had a large imbalance in fat depth samples. Regardless, the deep learning-based models that train on the entire hypercube are able to best capture the non-linearity in the regression problem when compared to the pixel-wise training models.

7.2 Future Questions

Active Line-scan Frame Camera System Calibration

A major component of this work is uncertainty estimation. Although experiments showed that the approximate uncertainty for the resulting calibration parameters were valid, improvements could be made by using higher-order error propagation terms.

The active calibration algorithm is performed offline with images that have been captured. It would be more useful to have the algorithm running during image collection. Such an implementation could be used to actively guide the poses of the pattern with an aim to reduce the uncertainty in the calibration parameters online.

Probabilistic Light Source Modelling

The light source modelling has assumed only a single light source is used with the hyperspectral camera, but ideally, multiple sources should be used to sufficiently illuminate an object. The light sources could be modelled individually using the proposed GP model, but a more interesting approach would consider modelling all light sources at once, where the measured radiance at the camera would be treated as a sum of the individual radiance from each light source. Comparing the difference between individual models would highlight the effectiveness of the all at once modelling approach.

For each measured spectral band of the hyperspectral camera, a separate light source model is fit for the same light source. The irradiance arriving at a surface is continuous and smooth in the spectral dimension, but simply adding another dimension to the GP would greatly increase the training time and storage complexity. The mean function would also need to be adapted to capture the smooth variation over the spectral dimension.

Reflectance Estimation with Light and Shape Information

The DRM is relied upon to describe the interactions of a material in order to estimate reflectance. It is still relatively simplistic in nature and does not model subsurface scattering, which limits its ability to generalise to a range of materials. Therefore, a more complex model that captures both surface and subsurface scattering could be used. The aim would be to formulate a similar optimisation, that uses smoothness constraints and pixel similarity in order to recover reflectance.

Case Study: Subcutaneous Fat Depth Estimation of Lamb Cuts

Deep learning provides a powerful tool, that when given enough data, can learn most functions. The reflectance is estimated from the radiance measurements to remove the effects of lighting and surface shape. It may be possible to learn invariance to these factors using a deep learning model with synthetic radiance measurements. The reflectance would still need to be estimated at least once with the methods from this thesis in order to create the synthetic radiance measurements. It may be of interest to determine which variable, from either lighting or surface shape, is more difficult to learn.

For hyperspectral imaging, the thesis only focuses on variations caused by lighting and shape to the radiance measurements. The viewer direction is overlooked as generally the hyperspectral sensor is fixed within a rig, and it is the object of interest which is moved. If the hyperspectral camera was instead attached to a robot arm, then there could be both translations and rotations relative to a surface. It would be interesting to consider the effects on the estimated reflectance on the same surface from different viewing directions.

Bibliography

[1] J. Mehami, T. Vidal-Calleja, and A. Alempijevic. Observability driven Multi-modal Line-scan Camera Calibration. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2020-Septe:285–290, 2020. doi: 10.1109/MFI49285.2020.9235226.

[2] J. Mehami, R. Falque, T. Vidal-Calleja, and A. Alempijevic. Multi-modal Non-Isotropic Light Source Modelling for Reflectance Estimation in Hyperspectral Imaging. *IEEE Robotics and Automation Letters*, 2022. ISSN 2377-3766.

[3] F. Peña, J. Mehami, R. Guenot-Falque, T. Patten, A. Alempijevic, and T. Vidal Calleja. Subcutaneous fat depth regression using hyperspectral and depth imaging. In *Australasian Conference on Robotics and Automation*. ARAA, 2022.

[4] M. Usayiwevu, C. Le Gentil, J. Mehami, C. Yoo, R. Fitch, and T. Vidal-Calleja. Information driven self-calibration for lidar-inertial systems. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9961–9967. IEEE, 2020.

[5] N. Koenig and A. Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 3, pages 2149–2154. IEEE, 2004.

[6] H. Carroll. *Read this if you want to take great photographs.* Laurence King Publishing, 2014.

[7] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision.* Cambridge university press, 2003. ISBN 0521540518.

[8] T. D. Barfoot. State estimation for robotics. *State Estimation for Robotics*, pages 1–368, 2017. doi: 10.1017/9781316671528.

[9] J. Sola. *Towards visual localization, mapping and moving objects tracking by a mobile robot: a geometric and probabilistic approach.* PhD thesis, Institut National Polytechnique de Toulouse-INPT, 2007.

[10] D. Scaramuzza and K. Ikeuchi. Omnidirectional camera. *Zurich Open Repository and Archive*, 2014.

[11] D. C. Brown. Close-range camera calibration, photogrammetric engineering. *Engineering and Remote Sensing*, 37(8):855–66, 1971.

[12] J. A. Sanz, A. M. Fernandes, E. Barrenechea, S. Silva, V. Santos, N. Gonçalves, D. Paternain, A. Jurio, and P. Melo-Pinto. Lamb muscle discrimination using hyperspectral imaging: Comparison of various machine learning algorithms. *Journal of Food Engineering*, 174:92–100, 2016. ISSN 0260-8774.

[13] S. Ghosh, P. Mishra, S. N. H. Mohamad, R. M. de Santos, B. D. Iglesias, and P. B. Elorza. Discrimination of peanuts from bulk cereals and nuts by near infrared reflectance spectroscopy. *Biosystems Engineering*, 151:178–186, 2016. ISSN 15375110. doi: 10.1016/j.biosystemseng.2016.09.008.

[14] D. Wu and D. W. Sun. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review - Part I: Fundamentals. *Innovative Food Science and Emerging Technologies*, 19:1–14, 2013. ISSN 14668564. doi: 10.1016/j.ifset.2013.04.014.

[15] D. Wu and D. W. Sun. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review - Part II: Applications. *Innovative Food Science and Emerging Technologies*, 19:15–28, 2013. ISSN 14668564. doi: 10.1016/j.ifset.2013.04.016.

[16] L. Gevaux. *3D-hyperspectral imaging and optical analysis of skin for the human face*. PhD thesis, Lyon, 2019.

[17] C. H. Feng, Y. Makino, S. Oshita, and J. F. García Martín. Hyperspectral imaging and multispectral imaging as the novel techniques for detecting defects in raw and processed meat products: Current state-of-the-art research advances. *Food Control*, 84:165–176, 2018. ISSN 09567135. doi: 10.1016/j.foodcont.2017.07.013.

[18] S. Lohumi, S. Lee, H. Lee, M. S. Kim, W. H. Lee, and B. K. Cho. Application of hyperspectral imaging for characterization of intramuscular fat distribution in beef. *Infrared Physics and Technology*, 74:1–10, 2016. ISSN 13504495. doi: 10.1016/j.infrared.2015.11.004.

[19] A. Siedliska, P. Baranowski, M. Zubik, W. Mazurek, and B. Sosnowska. Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. *Postharvest Biology and Technology*, 139(February):115–126, 2018. ISSN 09255214. doi: 10.1016/j.postharvbio.2018.01.018.

[20] X. Zheng, Y. Li, W. Wei, and Y. Peng. Detection of adulteration with duck meat in minced lamb meat by using visible near-infrared hyperspectral imaging. *Meat Science*, 149(March 2018):55–62, 2019. ISSN 03091740. doi: 10.1016/j.meatsci.2018.11.005.

[21] P. N. Schaare and D. G. Fraser. Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (Actinidia chinensis). *Postharvest Biology and Technology*, 20(2):175–184, 2000. ISSN 09255214. doi: 10.1016/S0925-5214(00)00130-7.

[22] F. Tao and M. Ngadi. Recent advances in rapid and nondestructive determination of fat content and fatty acids composition of muscle foods. *Critical Reviews in Food Science and Nutrition*, 58(9):1565–1593, 2018. ISSN 15497852. doi: 10.1080/10408398.2016.1261332.

[23] L. M. Kandpal, J. Lee, H. Bae, M. S. Kim, I. Baek, and B.-K. Cho. Near-infrared transmittance spectral imaging for nondestructive measurement of internal disorder in korean ginseng. *Sensors*, 20(1):273, 2020.

[24] P. Gou, E. Santos-Garcés, M. Høy, J. Wold, K. Liland, and E. Fulladosa. Feasibility of nir interactance hyperspectral imaging for on-line measurement of crude composition in vacuum packed dry-cured ham slices. *Meat Science*, 95(2):250–255, 2013. ISSN 0309-1740. doi: https://doi.org/10.1016/j.meatsci.2013.05.013.

[25] B. Cho, M. S. Kim, K. Chao, K. Lawrence, B. Park, and K. Kim. Detection of fecal residue on poultry carcasses by laser-induced fluorescence imaging. *Journal of food science*, 74(3):E154–E159, 2009.

[26] C. P. Huynh and A. Robles-Kelly. A solution of the dichromatic model for multispectral photometric invariance. *International Journal of Computer Vision*, 90(1): 1–27, 2010. ISSN 0920-5691.

[27] B. Boom, S. Orts-Escolano, X. Ning, S. McDonagh, P. Sandilands, and R. B. Fisher. Point light source estimation based on scenes recorded by a rgb-d camera. In *BMVC*, 2013.

[28] W. S. Lee and J. Blasco. Sensors i: Color imaging and basics of image processing. *Fundamentals of Agricultural and Field Robotics*, pages 13–37, 2021.

[29] T. Young. I. the bakerian lecture. experiments and calculations relative to physical optics. *Philosophical transactions of the Royal Society of London*, 94:1–16, 1804.

[30] H. von Helmholtz. *Treatise on physiological optics*, volume 3. Courier Corporation, 2013.

[31] G. Hu, S. Huang, L. Zhao, A. Alempijevic, and G. Dissanayake. A robust RGB-D slam algorithm. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1714–1719. IEEE, 2012.

[32] P. Ozog and R. M. Eustice. On the importance of modeling camera calibration uncertainty in visual SLAM. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 3777–3784, 2013. ISSN 10504729. doi: 10.1109/ICRA.2013.6631108.

[33] B. Sun, J. Zhu, L. Yang, S. Yang, and Z. Niu. Calibration of line-scan cameras for precision measurement. *Applied optics*, 55(25):6836–6843, 2016.

[34] D. Li, G. Wen, B. Wei Hui, S. Qiu, and W. Wang. Cross-ratio invariant based line scan camera geometric calibration with static linear data. *Optics and Lasers in Engineering*, 62:119–125, 2014. ISSN 01438166. doi: 10.1016/j.optlaseng.2014.03.004.

[35] C. A. Luna, M. Mazo, J. L. Lázaro, and J. F. Vázquez. Calibration of line-scan cameras. *IEEE Transactions on Instrumentation and Measurement*, 59(8):2185–2190, 2010. ISSN 00189456. doi: 10.1109/TIM.2009.2031344.

[36] D. Su, A. Bender, and S. Sukkarieh. Improved cross-ratio invariant-based intrinsic calibration of a hyperspectral line-scan camera. *Sensors (Switzerland)*, 18(6), 2018. ISSN 14248220. doi: 10.3390/s18061885.

[37] Z. Zhang. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1330–1334, dec 2000.

[38] D. Li, G. Wen, and S. Qiu. Cross-ratio–based line scan camera calibration using a planar pattern. *Optical Engineering*, 55(1):014104, 2016. ISSN 0091-3286. doi: 10.1117/1.oe.55.1.014104.

[39] B. Hui, G. Wen, P. Zhang, and D. Li. A novel line scan camera calibration technique with an auxiliary frame camera. *IEEE Transactions on Instrumentation and Measurement*, 62(9):2567–2575, 2013. ISSN 00189456. doi: 10.1109/TIM.2013.2256815.

[40] G. Di Leo, C. Liguori, A. Pietrosanto, and R. Lengu. Uncertainty of line camera image based measurements. *I2MTC 2017 - 2017 IEEE International Instrumentation and Measurement Technology Conference, Proceedings*, pages 1–6, 2017. doi: 10.1109/I2MTC.2017.7969698.

[41] A. Wendel and J. Underwood. Extrinsic parameter calibration for line scanning cameras on ground vehicles with navigation systems using a calibration pattern. *Sensors*, 17(11):2491, 2017.

[42] S. Dong, X. Shao, X. Kang, F. Yang, and X. He. Extrinsic calibration of a non-overlapping camera network based on close-range photogrammetry. *Applied Optics*, 55(23):6363, 2016. ISSN 0003-6935. doi: 10.1364/ao.55.006363.

[43] B. Sun, J. Zhu, L. Yang, Y. Guo, and J. Lin. Stereo line-scan sensor calibration for 3D shape measurement. *Applied Optics*, 56(28):7905, 2017. ISSN 1559-128X. doi: 10.1364/ao.56.007905.

[44] G. Kamberova and R. Bajcsy. Sensor errors and the uncertainties in stereo reconstruction. *Empirical Evaluation Techniques in Computer Vision*, pages 96–116, 1998.

[45] R. M. Haralick. Propagating covariance in computer vision. In *Performance Characterization in Computer Vision*, pages 95–114. Springer, 2000.

[46] L. Zhu, H. Luo, and X. Zhang. Uncertainty and sensitivity analysis for camera calibration. *Industrial Robot*, 36(3):238–243, 2009. ISSN 0143991X. doi: 10.1108/01439910910950496.

[47] R. Galego, A. Ortega, R. Ferreira, A. Bernardino, J. Andrade-Cetto, and J. Gaspar. Uncertainty analysis of the DLT-Lines calibration algorithm for cameras with radial distortion. *Computer Vision and Image Understanding*, 140:115–126, nov 2015. ISSN 1090235X. doi: 10.1016/j.cviu.2015.05.015.

[48] S. Fang, X. Xia, and Y. Xiao. A calibration method of lens distortion for line scan cameras. *Optik*, 124(24):6749–6751, 2013. ISSN 00304026. doi: 10.1016/j.ijleo.2013.05.084.

[49] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280–2292, 2014.

[50] E. A. Wan and R. Van Der Merwe. The unscented Kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, pages 153–158. Ieee, 2000. ISBN 0780358007.

[51] Z. Wang and G. Dissanayake. Observability analysis of slam using fisher information matrix. In *2008 10th International Conference on Control, Automation, Robotics and Vision*, pages 1242–1247. IEEE, 2008.

[52] D. Su, T. Vidal-Calleja, and J. V. Miro. Asynchronous microphone arrays calibration and sound source tracking. *Autonomous Robots*, 44(2):183–204, 2020.

[53] F. Nobre and C. Heckman. Learning to calibrate: Reinforcement learning for guided calibration of visual–inertial rigs. *The International Journal of Robotics Research*, 38(12-13):1388–1402, 2019.

[54] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan. Estimation with applications to tracking and navigation. In *Estimation with Applications to Tracking and Navigation*, chapter 3, pages 164–165. John Wiley & Sons, 2001. ISBN 0471465216. doi: 10.1002/0471221279.

[55] D. Azinović, T.-M. Li, A. Kaplanyan, and M. Nießner. Inverse Path Tracing for Joint Material and Lighting Estimation. *arXiv preprint arXiv:1903.07145*, 2019.

[56] M. Kasper, N. Keivan, G. Sibley, and C. Heckman. Light source estimation with analytical path-tracing. *arXiv preprint arXiv:1701.04101*, 2017.

[57] P. C. Kuo, H. Y. Huang, and S. H. Lai. Lighting estimation from a single image containing. *Proceedings of the 7th International Conference on Multimedia Systems, MMSys 2016*, pages 242–250, 2016. doi: 10.1145/2910017.2910613.

[58] H. Chen and X. Liu. Lighting and material of halo 3. In *ACM SIGGRAPH 2008 Games*, pages 1–22. 2008.

[59] M. Jongewaard. Guide to selecting the appropriate type of light source model. In *Modeling and Characterization of Light Sources*, volume 4775, pages 86–98. SPIE, 2002.

[60] J. Park, S. N. Sinha, Y. Matsushita, Y.-W. Tai, and I. So Kweon. Calibrating a non-isotropic near point light source using a plane. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2259–2266, 2014.

[61] L. Ma, J. Liu, X. Pei, Y. Hu, and F. Sun. Calibration of position and orientation for point light source synchronously with single image in photometric stereo. *Optics Express*, 27(4):4024, 2019. ISSN 1094-4087. doi: 10.1364/oe.27.004024.

[62] M. W. Powell, S. Sarkar, and D. Goldgof. A simple strategy for calibrating the geometry of light sources. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):1022–1027, 2001. ISSN 0162-8828.

[63] J. Ackermann, S. Fuhrmann, and M. Goesele. Geometric point light source calibration. *18th International Workshop on Vision, Modeling and Visualization, VMV 2013*, pages 161–168, 2013. doi: 10.2312/PE.VMV.VMV13.161-168.

[64] T. Takai, A. Maki, K. Niinuma, and T. Matsuyama. Difference sphere: An approach to near light source estimation. *Computer Vision and Image Understanding*, 113(9): 966–978, 2009. ISSN 10773142. doi: 10.1016/j.cviu.2009.03.017.

[65] Y. Nie, Z. Song, M. Ji, and L. Zhu. A novel calibration method for the photometric stereo system with non-isotropic LED lamps. *2016 IEEE International Conference on Real-Time Computing and Robotics, RCAR 2016*, pages 289–294, 2016. doi: 10.1109/RCAR.2016.7784041.

[66] M. Visentini-Scarzanella and H. Kawasaki. Simultaneous camera, light position and radiant intensity distribution calibration. In *Image and Video Technology*, pages 557–571. Springer, 2015.

[67] T. Aoto, T. Taketomi, T. Sato, Y. Mukaigawa, and N. Yokoya. Position estimation of near point light sources using a clear hollow sphere. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 3721–3724. IEEE, 2012.

[68] C. E. Rasmussen. Gaussian processes in machine learning. In *Summer School on Machine Learning*, pages 63–71. Springer, 2003.

[69] N. Wahlstrom, M. Kok, T. B. Schon, and F. Gustafsson. Modeling magnetic fields using Gaussian processes. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 3522–3526, 2013. ISSN 15206149. doi: 10.1109/ICASSP.2013.6638313.

[70] K. M. B. Lee, C. Yoo, B. Hollings, S. Anstee, S. Huang, and R. Fitch. Online estimation of ocean current from sparse GPS data for underwater vehicles. *Proceedings -*

IEEE International Conference on Robotics and Automation, 2019-May:3443–3449, 2019. ISSN 10504729. doi: 10.1109/ICRA.2019.8794308.

[71] J. Yan, H. Liu, W. Zhao, and Y. Su. Led filament standard lamp for total luminous flux with uniform spatial distribution. In *Journal of Physics: Conference Series*, volume 1345, page 032046. IOP Publishing, 2019.

[72] PassMark Software. 5W GU10 CW Downlight Review, 2013. URL `http://www.ledbenchmark.com`.

[73] M. López, K. Bredemeier, N. Rohrbeck, C. Véron, F. Schmidt, and A. Sperling. LED near-field goniophotometer at PTB. *Metrologia*, 49(2):S141, 2012. ISSN 0026-1394.

[74] G. Leschhorn, R. Young, et al. *Handbook of LED and SSL Metrology*. Pro Business, 2017.

[75] C. Yuksel. Point Light Attenuation Without Singularity. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks*, pages 1–2, 2020.

[76] D. Ginsburg, B. Purnomo, D. Shreiner, and A. Munshi. *OpenGL ES 3.0 programming guide*. Addison-Wesley Professional, 2014. ISBN 0133440125.

[77] S. Lagarde and C. de Rousiers. Moving frostbite to physically based rendering. In *SIGGRAPH 2014 Conference, Vancouver*, pages 45–46, 2014.

[78] M. D. Grossberg and S. K. Nayar. What is the space of camera response functions? *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2, 2003. ISSN 10636919. doi: 10.1109/cvpr.2003.1211522.

[79] W. Martens, Y. Poffet, P. R. Soria, R. Fitch, and S. Sukkarieh. Geometric priors for Gaussian process implicit surfaces. *IEEE Robotics and Automation Letters*, 2(2): 373–380, 2016. ISSN 2377-3766.

[80] Avian Technologies. Avian-b white reflectance coating, 2020. URL `https://aviantechnologies.com/product/avian-b-white-reflectance-coating/`.

[81] Spectralon® diffuse reflectance standards - diffuse reflectance, 2022. URL `https://www.labsphere.com/product/spectralon-diffuse-reflectance-standards/`.

[82] S. Rahman, P. Quin, T. Walsh, T. Vidal-Calleja, M. J. McPhee, E. Toohey, and A. Alempijevic. Preliminary estimation of fat depth in the lamb short loin using a hyperspectral camera. *Animal Production Science*, 58(8):1488–1496, 2018. ISSN 1836-5787.

[83] A. Robles-Kelly and C. P. Huynh. *Imaging spectroscopy for scene analysis*. Springer Science & Business Media, 2012. ISBN 1447146522.

[84] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2014. ISSN 0162-8828.

[85] S. A. Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.

[86] A. Krebs, Y. Benezeth, and F. Marzani. Quadratic objective functions for dichromatic model parameters estimation. In *2017 International Conference on Digital Image Computing: Techniques and Applications*, pages 1–8. IEEE, 2017. ISBN 1538628392.

[87] A. Krebs, Y. Benezeth, and F. Marzani. Intrinsic rgb and multispectral images recovery by independent quadratic programming. *PeerJ Computer Science*, 6, 2020.

[88] S. Rahman and A. Robles-Kelly. An optimisation approach to the recovery of reflection parameters from a single hyperspectral image. *Computer vision and image understanding*, 117(12):1672–1688, 2013.

[89] O. De Carvalho Jr and P. Meneses. Spectral correlation mapper: An improvement on the spectral angle mapper. In *Summaries of the 9th Airborne Earth Science Workshop*, pages 1–9, 2000.

[90] W. Jakob. Mitsuba Renderer Documentation, 2010. URL `http://www.mitsuba-renderer.org`.

[91] G. ElMasry, D. W. Sun, and P. Allen. Chemical-free assessment and mapping of major constituents in beef using hyperspectral imaging. *Journal of Food Engineering*, 117(2):235–246, 2013. ISSN 02608774. doi: 10.1016/j.jfoodeng.2013.02.016.

[92] M. Kamruzzaman, Y. Makino, and S. Oshita. Non-invasive analytical technology for the detection of contamination, adulteration, and authenticity of meat, poultry, and fish: A review. *Analytica Chimica Acta*, 853(1):19–29, 2015. ISSN 18734324. doi: 10.1016/j.aca.2014.08.043.

[93] Y. Yang, O. O. Soyemi, M. R. Landry, and B. R. Soller. Influence of a fat layer on the near infrared spectra of human muscle: quantitative analysis based on two-layered monte carlo simulations and phantom experiments. *Optics express*, 13(5):1570–1579, 2005.

[94] F. H. Mustafa, P. W. Jones, and A. L. McEwan. Near infrared spectroscopy for body fat sensing in neonates: quantitative analysis by gamos simulations. *Biomedical engineering online*, 16(1):1–17, 2017.

[95] D.-S. Ho, E.-H. Kim, I.-D. Hwang, K.-S. Shin, J.-T. Oh, and B.-M. Kim. Optical skin-fat thickness measurement using miniaturized chip leds: a preliminary human study. *Journal of the Optical Society of Korea*, 13(3):304–309, 2009.

[96] Y. Wang, Z. Yang, D. Hao, S. Zhang, Y. Yang, and Y. Zeng. Measurement of subcutaneous adipose tissue thickness by near-infrared. *Australasian physical & engineering sciences in medicine*, 36(2):201–208, 2013.

[97] A. Weber, J. Salau, J. H. Haas, W. Junge, U. Bauer, J. Harms, O. Suhr, K. Schönrock, H. Rothfuß, S. Bieletzki, and G. Thaller. Estimation of backfat thickness using extracted traits from an automatic 3D optical system in lactating Holstein-Friesian cows. *Livestock Science*, 165(1):129–137, 2014. ISSN 18711413. doi: 10.1016/j.livsci.2014.03.022.

[98] J. Rodríguez Alvarez, M. Arroqui, P. Mangudo, J. Toloza, D. Jatip, J. M. Rodríguez, A. Teyseyre, C. Sanz, A. Zunino, C. Machado, and C. Mateos. Body condition estimation on cows from depth images using Convolutional Neural Networks. *Computers and Electronics in Agriculture*, 155(October):12–22, 2018. ISSN 01681699. doi: 10.1016/j.compag.2018.09.039.

[99] S. R. Silva, R. Payan-Carreira, M. Quaresma, C. M. Guedes, and A. S. Santos. Relationships between body condition score and ultrasound skin-associated subcutaneous fat depth in equids. *Acta Veterinaria Scandinavica*, 58(1):37–42, 2016. ISSN 0044605X. doi: 10.1186/s13028-016-0243-2.

[100] A. Kirton and D. Johnson. Interrelationships between gr and other lamb carcass fatness measurements. *Proceedings of the New Zealand Society of Animal Production (New Zealand)*, 1979.

[101] K. Stanford, S. Jones, and M. Price. Methods of predicting lamb carcass composition: A review. *Small Ruminant Research*, 29(3):241–254, 1998.

[102] A. Williams, F. Anderson, J. Siddell, D. Pethick, J. H. Edwards, and G. Gardner. Predicting lamb carcase composition from carcase weight and gr tissue depth. *Nurturing Locally, Growing Globally*, page 729, 2017.

[103] D. Hopkins, E. Ponnampalam, and R. Warner. Predicting the composition of lamb carcases using alternative fat and muscle depth measures. *Meat Science*, 78(4):400–405, 2008.

[104] P. Kenyon, S. Maloney, and D. Blache. Review of sheep body condition score in relation to production characteristics. *New Zealand Journal of Agricultural Research*, 57(1):38–64, 2014.

[105] C. Shands, B. McLeod, M. Lollback, G. Duddy, S. Hatcher, and W. O'halloran. Comparison of manual assessments of ewe fat reserves for on-farm use. *Animal Production Science*, 49(7):630–636, 2009.

[106] L. Bünger, C. A. Glasbey, G. Simm, J. Conington, J. M. Macfarlane, K. A. McLean, K. Moore, and N. R. Lambe. *Use of X-ray computed tomography (CT) in UK sheep production and breeding.* INTECH Open Access Publisher, 2011. ISBN 9533079436.

[107] D. L. Hopkins, K. L. Pirlot, A. H. K. Roberts, and A. S. Beattie. Changes in fat depths and muscle dimensions in growing lambs as measured by real-time ultrasound. *Australian Journal of Experimental Agriculture*, 33(6):707–712, 1993.

[108] J. Maxa, E. Norberg, P. Berg, and M. Milerski. Genetic parameters for body weight, longissimus muscle depth and fat depth for Suffolk sheep in the Czech Republic. *Small Ruminant Research*, 72(2-3):87–91, 2007. ISSN 09214488. doi: 10.1016/j.smallrumres.2006.04.018.

[109] N. Morsy and D. W. Sun. Robust linear and non-linear models of NIR spectroscopy for detection and quantification of adulterants in fresh and frozen-thawed minced beef. *Meat Science*, 93(2):292–302, 2013. ISSN 03091740. doi: 10.1016/j.meatsci.2012.09.005.

[110] Z. Xiong, D. W. Sun, H. Pu, Z. Zhu, and M. Luo. Combination of spectra and texture data of hyperspectral imaging for differentiating between free-range and broiler chicken meats. *LWT - Food Science and Technology*, 60(2):649–655, 2015. ISSN 00236438. doi: 10.1016/j.lwt.2014.10.021.

[111] B. G. Logan, D. L. Hopkins, L. M. Schmidtke, and S. M. Fowler. Analysis of Raman spectra for the verification of Australian grass-and grain-fed beef using principal component analysis and partial least square models. *Journal of Raman Spectroscopy*, 51(11):2338–2346, 2020. ISSN 0377-0486.

[112] G. ElMasry, D.-W. Sun, and P. Allen. Near-infrared hyperspectral imaging for predicting colour, ph and tenderness of fresh beef. *Journal of Food Engineering*, 110(1):127–140, 2012. ISSN 0260-8774. doi: https://doi.org/10.1016/j.jfoodeng.2011.11.028.

[113] Y. N. Chen, D. W. Sun, J. H. Cheng, and W. H. Gao. Recent Advances for Rapid Identification of Chemical Information of Muscle Foods by Hyperspectral Imaging Analysis. *Food Engineering Reviews*, 8(3):336–350, 2016. ISSN 18667929. doi: 10.1007/s12393-016-9139-1.

[114] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[115] R. Huang, J. Pedoeem, and C. Chen. Yolo-lite: a real-time object detection algorithm optimized for non-gpu computers. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 2503–2510. IEEE, 2018.

[116] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):37–78, 2017. ISSN 2168-6831.

[117] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251, 2016.

[118] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:279–317, 2019. ISSN 09242716. doi: 10.1016/j.isprsjprs.2019.09.006.

[119] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709, 2019. ISSN 0196-2892.

[120] I. Alonso, L. Riazuelo, and A. C. Murillo. MiniNet: An efficient semantic segmentation convnet for real-time robotic applications. *IEEE Transactions on Robotics*, 36 (4):1340–1347, 2020.

[121] F. Peña, P. V. Aguilar, D. S. Gracia, and A. C. Murillo. Efficient semantic segmentation with hyperspectral images. In *Proceedings of the Iberian Robotics Conference*, 2022. (to appear).

[122] G. A. Borkan, S. G. Gerzof, A. H. Robbins, D. Hults, C. Silbert, and J. Silbert. Assessment of abdominal fat content by computed tomography. *The American journal of clinical nutrition*, 36(1):172–177, 1982.

[123] B. H. Goodpaster, F. L. Thaete, and D. E. Kelley. Composition of skeletal muscle evaluated with computed tomography. *Annals of the New York Academy of Sciences*, 904(1):18–24, 2000.

[124] S. Adrian. DicomToMesh, 2020. https://github.com/AOT-AG/DicomToMesh.

[125] N. Sharp et al. Polyscope, 2019. www.polyscope.run.

[126] R. W. Sumner, J. Schmid, and M. Pauly. Embedded deformation for shape manipulation. In *ACM siggraph 2007 papers*, pages 80–es, 2007.

[127] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 19(1):263–272, 2017.

[128] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[129] Y. Sun, W. Zuo, and M. Liu. Rtfnet: Rgb-thermal fusion network for semantic segmentation of urban scenes. *IEEE Robotics and Automation Letters*, 4(3):2576–2583, 2019.

[130] W. Li, C. Chen, M. Zhang, H. Li, and Q. Du. Data augmentation for hyperspectral image classification with deep cnn. *IEEE Geoscience and Remote Sensing Letters*, 16(4):593–597, 2018.

[131] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica. Tune: A research platform for distributed model selection and training. *arXiv preprint arXiv:1807.05118*, 2018.