Natural Illumination Invariant Imaging

by

Andrew Ulrich Rutgers

B.A.Sc., Queen’s University, 2001

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in
The Faculty of Graduate Studies
(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)
April, 2011
© Andrew Ulrich Rutgers 2011
Abstract

Shadows confound many machine vision algorithms and strong ambient illumination confounds active imaging, especially in outdoor automation applications such as surface mining. Natural Illumination Invariant Imaging takes images that appear to be illuminated by only an intentional flash, even with a flash only 1/37000th the power of the ambient light. This work combines hardware techniques, including brief pulses and filtering to reduce the apparent intensity of ambient light, with subtracting a reference image of the scene under only ambient illumination. Most existing techniques simply subtract an image of the scene taken at a previous time, or use several cameras to estimate the reference image. This work explores creating an accurate reference image using two cameras. The improved reference image for subtraction allows ambient shadows to be substantially removed from an image.

Four experimental methods: reference estimation, segmentation, shadow removal and depth correlation; were used to evaluate the performance of ten techniques. Estimates of flash-free reference images created with the proposed techniques were compared to real images taken without a flash showed up to 63% reduction in error over existing techniques. The techniques were applied to three image processing methods: segmentation, shadow removal and depth correlation. They were applied to video for segmentation, and demonstrated segmenting an object 55% more accurately, though the measurement is scene dependent. Next, shadow ratio experiments examined the magnitude of the shadow remaining after subtraction, showing up to 70% shadow intensity removal. Depth correlation experiments showed over a 60% increase in the image area recognized using an efficient logical correlation algorithm. The theoretical performance of the techniques was examined. In summary, illumination invariant imaging was explored theoretically and experimentally, showing significant improvements for some important methods.
# Table of Contents

Abstract ................................................................. ii

Table of Contents ....................................................... iii

List of Tables .......................................................... vi

List of Figures .......................................................... vii

List of Algorithms ...................................................... ix

Nomenclature ........................................................... x

Acknowledgments ......................................................... xvii

Dedication ............................................................... xviii

1 Introduction .......................................................... 1
  1.1 Overview .......................................................... 3
  1.2 Machine Vision Methods Assisted by Illumination Invariance .... 4
  1.3 General Applications of Illumination Invariance .................... 7
  1.4 Selected Specific Applications of Illumination Invariance ......... 9
  1.5 Previous Work .................................................... 11
  1.6 Objectives of This Thesis ........................................ 16
  1.7 Organization of This Thesis ..................................... 17

2 Approach ............................................................. 18
  2.1 Filtering & Flash ................................................ 18
  2.2 Subtraction ....................................................... 19
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3 Imaging Model</td>
<td>20</td>
</tr>
<tr>
<td>2.4 Ratio Based Estimation</td>
<td>21</td>
</tr>
<tr>
<td>2.5 Offset Ratio Based Estimation</td>
<td>27</td>
</tr>
<tr>
<td>2.6 Central Ratio Based Estimation</td>
<td>29</td>
</tr>
<tr>
<td>2.7 Time Selection</td>
<td>36</td>
</tr>
<tr>
<td>3 Ambient-Illumination Only Reference Image Estimate Accuracy</td>
<td>37</td>
</tr>
<tr>
<td>3.1 Accuracy Evaluation</td>
<td>37</td>
</tr>
<tr>
<td>3.2 Dual Camera Apparatus</td>
<td>38</td>
</tr>
<tr>
<td>3.3 Experiment</td>
<td>43</td>
</tr>
<tr>
<td>3.4 Results</td>
<td>45</td>
</tr>
<tr>
<td>3.5 Summary</td>
<td>49</td>
</tr>
<tr>
<td>4 Segmentation</td>
<td>52</td>
</tr>
<tr>
<td>4.1 Experimental Setup</td>
<td>52</td>
</tr>
<tr>
<td>4.2 Object Segmentation</td>
<td>57</td>
</tr>
<tr>
<td>4.3 Noise Analysis</td>
<td>60</td>
</tr>
<tr>
<td>5 Shadow Removal</td>
<td>75</td>
</tr>
<tr>
<td>5.1 Apparatus</td>
<td>75</td>
</tr>
<tr>
<td>5.2 Qualitative Evaluation</td>
<td>76</td>
</tr>
<tr>
<td>5.3 Shadow Ratio</td>
<td>77</td>
</tr>
<tr>
<td>5.4 Results</td>
<td>83</td>
</tr>
<tr>
<td>6 Depth Correlation</td>
<td>84</td>
</tr>
<tr>
<td>6.1 Apparatus</td>
<td>85</td>
</tr>
<tr>
<td>6.2 Image Processing</td>
<td>91</td>
</tr>
<tr>
<td>6.3 Results</td>
<td>96</td>
</tr>
<tr>
<td>6.4 Computational Feasibility of Higher Depth Resolution</td>
<td>98</td>
</tr>
<tr>
<td>6.5 Summary</td>
<td>99</td>
</tr>
<tr>
<td>7 Bounds on Ultimate Performance</td>
<td>103</td>
</tr>
<tr>
<td>7.1 Wavelength Selection</td>
<td>103</td>
</tr>
<tr>
<td>7.2 Optical Safety</td>
<td>108</td>
</tr>
</tbody>
</table>
List of Tables

1  Notations Conventions Used for Images .................. xiv
2  Notation for Images ........................................ xv
3  Notation for Partial Estimates ............................... xv
4  Summary of Estimation Techniques ........................... xvi
3.1 NIII Technique Performance ................................. 46
4.1 Confusion Matrix for Bar Sequence ......................... 61
4.2 Confusion Matrix for Funnel Sequence ...................... 61
4.3 Monte Carlo Noise Estimates for Different NIII Techniques ... 73
5.1 Shadow Ratio .................................................. 83
6.1 Depth Categorization Accuracy .............................. 98
7.1 Typical Optical Material Ranges ............................ 105
7.2 Typical Optical Detector Sensitivity Ranges ................. 106
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Sample Scene Under Different Lighting Conditions</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Three Key Complications that Shadows Cause in Machine Vision</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Sample Scene Showing False Edges Created by Shadows</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>Comparison of Padilla and Subtraction</td>
<td>13</td>
</tr>
<tr>
<td>2.1</td>
<td>RBE Apparatus Diagram</td>
<td>26</td>
</tr>
<tr>
<td>2.2</td>
<td>The Four Weighting Techniques</td>
<td>32</td>
</tr>
<tr>
<td>3.1</td>
<td>Rendering of Dual Camera System</td>
<td>39</td>
</tr>
<tr>
<td>3.2</td>
<td>Image of the Completed Dual Camera System</td>
<td>39</td>
</tr>
<tr>
<td>3.3</td>
<td>Alignment Calibration Image</td>
<td>44</td>
</tr>
<tr>
<td>3.4</td>
<td>False-Color Outdoor Scene Video Examples</td>
<td>45</td>
</tr>
<tr>
<td>3.5</td>
<td>Offset Ratio Based Estimation Performance v. $q$</td>
<td>47</td>
</tr>
<tr>
<td>3.6</td>
<td>Variation in ORBE Performance</td>
<td>48</td>
</tr>
<tr>
<td>3.7</td>
<td>Mean SRBE Absolute Performance vs. $\alpha$</td>
<td>49</td>
</tr>
<tr>
<td>3.8</td>
<td>SRBE Performance Variation</td>
<td>50</td>
</tr>
<tr>
<td>4.1</td>
<td>Segmentation Experimental Setup</td>
<td>53</td>
</tr>
<tr>
<td>4.2</td>
<td>RBE/GUI Main Window</td>
<td>56</td>
</tr>
<tr>
<td>4.3</td>
<td>RBE/GUI Image Viewer Window</td>
<td>56</td>
</tr>
<tr>
<td>4.4</td>
<td>Shadow Removal Video Frame - Bar Sequence</td>
<td>58</td>
</tr>
<tr>
<td>4.5</td>
<td>Shadow Removal Video Frame - Funnel Sequence</td>
<td>59</td>
</tr>
<tr>
<td>4.6</td>
<td>Image SNR Examples</td>
<td>63</td>
</tr>
<tr>
<td>4.7</td>
<td>Segmentation SNR Error Calculation Example</td>
<td>64</td>
</tr>
<tr>
<td>4.8</td>
<td>Segmentation Error Rate v. SNR</td>
<td>65</td>
</tr>
<tr>
<td>4.9</td>
<td>Example Segmentation Image used for Noise Analysis</td>
<td>66</td>
</tr>
</tbody>
</table>
# List of Algorithms

2.1 PDIR: Past-Only Direct Estimation ............................................. 20  
2.2 PRBE: Past-Only Ratio Based Estimation ................................. 27  
2.3 ORBE: Offset Ratio Based Estimation ................................. 28  
2.4 ARBE: Average Ratio Based Estimation ................................. 34  
2.5 ADIR: Average Direct Estimation ............................................. 34  
2.6 LRBE: Linear Ratio Based Estimation ............................................. 34  
2.7 LDIR: Linear Direct Estimation ............................................. 35  
2.8 SRBE: Sigmoid Ratio Based Estimation ............................................. 35  
2.9 SDIR: Sigmoid Direct Estimation ............................................. 35  
2.10 TSEL: Time Selection ......................................................... 36
Nomenclature

ADC  Analog to Digital Converter

ADIR  Average Direct Estimation. See Table 4

\( \alpha \)  A coefficient which sets the width of the sigmoid function in Sigmoid-Weighting techniques (SRBE, SDIR) according to (2.28).

ARBE  Average Ratio Based Estimation. See Table 4

BRDF  Bidirectional Reflectance Distribution Function

CRBE  Central Ratio Based Estimation. This encompasses RBE techniques where past and future images are used to create the estimate, including ARBE, LRBE and SRBE.

DIR  Direct Estimation - where an image (or weighted combination of images) at the same wavelength is used as the no-flash estimate.

DLP  Digital Light Processor, the Texas Instruments trade name for a micromirror array.

DOE  Diffractive Optical Element

\( \varepsilon \)  The rms residual, a measure of how close an estimate is to a real image details in Section 3.1.1 and defined in (3.2).

FPN  Fixed Pattern Noise

fps  frames per second
FWHM  Full-Width Half-Maximum is a measure of filter bandwidth. It measures the bandwidth of the passband at 50% of the maximum transmission of the filter.

\( \gamma \) A coefficient calculated for each pixel in Sigmoid Weighting (for SRBE, SDIR techniques) using (2.29).

GUI  Graphical User Interface

\( I \) Pixel Intensity

IEC  International Electrotechnical Commission

IPP  The Intel Performance Primitives is a library of mathematical functions optimized for their processors, including using multiple cores on large operations. It includes thousands of standard mathematical and image processing functions [37].

\( \kappa \) The shadow ratio is the ratio between the shadow intensity (\( \zeta \)) and the light intensity (\( \xi \)).

KPI  Known-Pattern Image

\( \lambda_A \) The wavelength range matched to the flash wavelength.

\( \lambda_B \) The wavelength range used for a reference, which does not include the wavelengths the flash produces.

LCD  Liquid Crystal Displays are a common form of display technology where an electrical impulse can change an element to transmit or absorb light.

LDIR  Linear Direct Estimation. See Table 4

lm  lumen - a measure of light flux weighted by wavelength to represent the apparent intensity to a human eye.

LRBE  Linear Ratio Based Estimation. See Table 4

M  Number of rows in an image
m  Row index of an image. The number of rows is M.

MPE  Maximum Permissible Exposure is the highest light intensity deemed safe by standards.

N  Number of columns in an image

n  Column index of an image

Nd-YAG  Neodymium - Yttrium Aluminum Garnet. A type of laser which operates at 1064nm and is frequently used for industrial applications due to its high power.

NIII  Natural Illumination Invariant Imaging

NIR  Near InfraRed light has a wavelength between 800nm and 1200nm

$N_Q$  Noise due to Quantization

$N_{read}$  Noise due to Readout Error

$N_T$  Total Noise

ORBE  Offset Ratio Based Estimation. See Table 4

OSRAM  A wholly owned subsidiary of Siemens which produces lighting products, including the LED arrays used in some of this work.

PDIR  Past-Only Direct Estimation. See Table 4

PE  Partial Estimation

PRBE  Past-Only Ratio Based Estimation. See Table 4

PRNU  Photon-Response Non-Uniformity

Q  Smallest Quantization Step

$q$  An offset added to terms in RBE to reduce the impact of quantization noise.
Qt  An open source graphical user interface library developed by Nokia. The
library is implemented on Linux, Mac, Windows and some cell phone OSes
allowing easy porting between platforms [68].

RBE  Ratio Based Estimation

\( \rho \)  A weighting coefficient used to weight two partial estimates from different
image pairs.

ROI  Region of Interest

SBRDF  Spectral Bidirectional Reflectance Function

SDIR  Sigmoid Direct Estimation. See Table 4

SL  Structured Light. A method of 3D depth vision where a pattern is projected
on a scene. Using angles from the camera and projector the depth can be
triangulated.

\( S N_{\text{dark}} \)  Dark Current Shot Noise

\( S N_{\text{ph}} \)  Photon Shot Noise

SRBE  Sigmoid Ratio Based Estimation. See Table 4

\( \tau \)  The attenuation of light due to the geometry of a scene, including incident
angles and distance.

TSEL  Time Selection Estimation. See Table 4

UV  Ultraviolet light has a wavelength <400nm

\( \xi \)  The intensity of the shadow over a region of interest in an image.

xml  Extensible Markup Language

\( \zeta \)  The light intensity over a region of interest of an image.
<table>
<thead>
<tr>
<th>Math Notation</th>
<th>Algorithm Notation</th>
<th>Wavelength</th>
<th>Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>$\lambda_A$</td>
<td>Image taken with the camera narrow bandpass filtered to the same wavelength range as the flash</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>$\lambda_B$</td>
<td>Image taken with the camera narrow bandpass filtered to a wavelength range which excludes the flash</td>
</tr>
<tr>
<td>$\hat{A}$</td>
<td>$\text{Aest}$</td>
<td>$\lambda_A$</td>
<td>Estimate of an Image taken with the camera filtered to the same wavelength as the flash</td>
</tr>
<tr>
<td>$\bar{A}$</td>
<td>$\text{Amean}$</td>
<td>$\lambda_A$</td>
<td>Mean Image of only the background created by averaging images of an unchanging scene.</td>
</tr>
</tbody>
</table>

(a) Notation Convention for the Wavelength an image is taken

<table>
<thead>
<tr>
<th>Math Notation</th>
<th>Algorithm Notation</th>
<th>Time</th>
<th>Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>$'$</td>
<td>1</td>
<td>$t_1$</td>
<td>Always without flash</td>
</tr>
<tr>
<td>''</td>
<td>2</td>
<td>$t_2$</td>
<td>With flash at wavelength $\lambda_A$, except for estimate accuracy experiments</td>
</tr>
<tr>
<td>'''</td>
<td>3</td>
<td>$t_3$</td>
<td>Always without flash</td>
</tr>
</tbody>
</table>

(b) Notation Convention for the Time when an image is taken

<table>
<thead>
<tr>
<th>Math Notation (Subscript)</th>
<th>Algorithm Notation</th>
<th>Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>Ambient-Illumination Only</td>
</tr>
<tr>
<td>F</td>
<td>f</td>
<td>Flash-Illumination Only</td>
</tr>
<tr>
<td>$AF$</td>
<td>af</td>
<td>Ambient and Flash Illumination</td>
</tr>
</tbody>
</table>

(c) Notation Convention for the Scene Illumination when an image is taken

Table 1: Notations Conventions Used for Images
<table>
<thead>
<tr>
<th>Math Notation</th>
<th>Algorithm Notation</th>
<th>Time</th>
<th>Wavelength</th>
<th>Illumination</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A'_A$</td>
<td>A1a</td>
<td>$t_1$</td>
<td>$\lambda_A$</td>
<td>Ambient</td>
<td>Image</td>
</tr>
<tr>
<td>$A''_A$</td>
<td>A2a</td>
<td>$t_2$</td>
<td>$\lambda_A$</td>
<td>Ambient</td>
<td>Image</td>
</tr>
<tr>
<td>$A''_{AF}$</td>
<td>A2af</td>
<td>$t_2$</td>
<td>$\lambda_A$</td>
<td>Ambient + Flash</td>
<td>Image</td>
</tr>
<tr>
<td>$A''_f$</td>
<td>A2f</td>
<td>$t_2$</td>
<td>$\lambda_A$</td>
<td>Flash</td>
<td>Image</td>
</tr>
<tr>
<td>$A''_A$</td>
<td>A3a</td>
<td>$t_3$</td>
<td>$\lambda_A$</td>
<td>Ambient</td>
<td>Image</td>
</tr>
<tr>
<td>$B'_A$</td>
<td>B1a</td>
<td>$t_1$</td>
<td>$\lambda_B$</td>
<td>Ambient</td>
<td>Image</td>
</tr>
<tr>
<td>$B''_A$</td>
<td>B2a</td>
<td>$t_2$</td>
<td>$\lambda_B$</td>
<td>Ambient</td>
<td>Image</td>
</tr>
<tr>
<td>$B''_A$</td>
<td>B3a</td>
<td>$t_3$</td>
<td>$\lambda_B$</td>
<td>Ambient</td>
<td>Image</td>
</tr>
<tr>
<td>$A''_{Aest}$</td>
<td>Aest2a</td>
<td>$t_2$</td>
<td>$\lambda_A$</td>
<td>Ambient</td>
<td>Estimate</td>
</tr>
<tr>
<td>$A''_{f}$</td>
<td>Aest2f</td>
<td>$t_2$</td>
<td>$\lambda_A$</td>
<td>Flash</td>
<td>Estimate</td>
</tr>
<tr>
<td>$\bar{A}_A$</td>
<td>AmeanA</td>
<td>n/a</td>
<td>$\lambda_A$</td>
<td>Ambient</td>
<td>Mean</td>
</tr>
<tr>
<td>$\bar{A}_F$</td>
<td>AmeanF</td>
<td>n/a</td>
<td>$\lambda_A$</td>
<td>Flash</td>
<td>Mean</td>
</tr>
</tbody>
</table>

Table 2: Notation for Images. Mean images don’t have a time since they are a mean of many images in a sequence.

<table>
<thead>
<tr>
<th>Math Notation</th>
<th>Algorithm Notation</th>
<th>Source Information Times</th>
<th>Technique</th>
<th>Equation</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{A}'_{12}^{dir}$</td>
<td>Aest2dir12</td>
<td>$t_1, t_2$</td>
<td>DIR</td>
<td>2.21</td>
<td>30</td>
</tr>
<tr>
<td>$\hat{A}'_{12}^{dir}$</td>
<td>Aest2dir32</td>
<td>$t_2, t_3$</td>
<td>DIR</td>
<td>2.22</td>
<td>30</td>
</tr>
<tr>
<td>$\hat{A}'_{12}^{rbe}$</td>
<td>Aest2rbe12</td>
<td>$t_1, t_2$</td>
<td>RBE</td>
<td>2.23</td>
<td>30</td>
</tr>
<tr>
<td>$\hat{A}'_{12}^{rbe}$</td>
<td>Aest2rbe32</td>
<td>$t_2, t_3$</td>
<td>RBE</td>
<td>2.24</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 3: Notation for Partial Estimates, discussed in Subsection 2.6.1 on page 29. All partial estimates are at time $t_2$ and wavelength $\lambda_A$. 
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Name</th>
<th>Section</th>
<th>Page</th>
<th>Equation(s)</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDIR</td>
<td>Past-Only Direct</td>
<td>2.2</td>
<td>19</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>PRBE</td>
<td>Past-Only Ratio Based Estimation</td>
<td>2.4</td>
<td>21</td>
<td>2.12, 2.18</td>
<td>2.2</td>
</tr>
<tr>
<td>ORBE</td>
<td>Offset Ratio Based Estimation</td>
<td>2.5</td>
<td>27</td>
<td>2.19, 2.18</td>
<td>2.3</td>
</tr>
<tr>
<td>ADIR</td>
<td>Average Weighted Direct</td>
<td>2.6</td>
<td>29</td>
<td>2.21, 2.22, 2.25, ρ = 0.5</td>
<td>2.5</td>
</tr>
<tr>
<td>ARBE</td>
<td>Average Weighted Ratio Based Estimation</td>
<td>2.6</td>
<td>29</td>
<td>2.23, 2.24, 2.25, ρ = 0.5</td>
<td>2.4</td>
</tr>
<tr>
<td>LDIR</td>
<td>Linear Weighted Direct</td>
<td>2.6</td>
<td>29</td>
<td>2.21, 2.22, 2.25, 2.26, 2.27</td>
<td>2.7</td>
</tr>
<tr>
<td>LRBE</td>
<td>Linear Weighted Ratio Based Estimation</td>
<td>2.6</td>
<td>29</td>
<td>2.23, 2.24, 2.25, 2.26, 2.27</td>
<td>2.6</td>
</tr>
<tr>
<td>SDIR</td>
<td>Sigmoid Weighted Direct</td>
<td>2.6</td>
<td>29</td>
<td>2.21, 2.22, 2.25, 2.28, 2.29</td>
<td>2.9</td>
</tr>
<tr>
<td>SRBE</td>
<td>Sigmoid Weighted Ratio Based Estimation</td>
<td>2.6</td>
<td>29</td>
<td>2.23, 2.24, 2.25, 2.28, 2.29</td>
<td>2.8</td>
</tr>
<tr>
<td>TSEL</td>
<td>Time Selection</td>
<td>2.7</td>
<td>36</td>
<td>2.30</td>
<td>2.10</td>
</tr>
</tbody>
</table>

Table 4: Summary of Estimation Techniques
Acknowledgments

My research would have been impossible without the support of my advisers, Dr. Peter Lawrence and Dr. Robert Hall and their continuous support, insight and encouragement.

This work was financially supported by the BC Innovation Council - Industrial Innovation Scholarship, NSERC grant STPGP 321813 to R. Hall (PI), P. Lawrence and S. Salcudean, and the UBC Ph.D. Tuition Fee Award. Westgrid high performance computer resources were used for some of the analysis.

Many people contributed resources, equipment, their time and trust to this research including Daan, Nick, Thio, Kristie, Don, Neil Dave F. and others. Darren, a summer intern, helped researching the structured light algorithms. Their support helped this research run smoothly.

My lab mates, Ali, Nick, Nima, James, Mike, Joel and Hamid were invaluable for inspiration, ideas and the occasional welcome distraction. My friends and fellow students, Scott, Tom and Kalev provided welcome fresh perspectives over countless curries.
Dedication

To my family and friends, for their support throughout my studies,

and

To all those who inspired and encouraged me towards engineering in my youth.
Chapter 1

Introduction

Illumination poses a major challenge for machine vision applications in uncontrolled environments. On farms, on roads, in open pit mines and in forests the lighting cannot be controlled, creating unpredictable shadows and a wide image intensity range. Natural Illumination Invariant Imaging (NIII) aims to produce images which are invariant to the ambient lighting conditions using a fixed, predictable flash and subtracting out ambient light.

Machine vision has many industrial applications, within the confines of structured environments where lighting can be controlled. Factories often use automated inspection to ensure their products are made properly before sending them to the consumer. These applications are possible because of a carefully controlled environment, with consistent lighting.

Lighting can dramatically affect machine vision performance. Figure 1.1 shows a simple example of the impact of illumination on scene interpretation. Both images on the top (Figures 1.1a, 1.1b) were taken of the same scene. Figure 1.1a was taken under a diffuse ambient scene lighting, while Figure 1.1b was taken with an additional powerful spot light located to the right of the scene. Both the top images are converted to black and white using a 50% intensity threshold, producing the bottom images. Figure 1.1c shows a reasonable outline of the objects while Figure 1.1d is badly distorted. In many situations, turning off the spotlight is impossible and other solutions, like NIII, are needed. Successful NIII will allow a poorly lit scene, such as Figure 1.1b, to appear well lit, like Figure 1.1a.

Shadows create three key complications for machine vision, illustrated in Figure 1.2. First, false edges are created by the shadow border, which is easily mistaken for a feature. Second, discontinuities appear in a feature when the feature crosses from a lit region to a shaded region. Finally, lowered contrast in shaded regions reduces the visibility of features. Since saturation occurs when the camera
pixel is overwhelmed with light and reads a fixed maximum value, adjusting the camera so the lit regions are not saturated may reduce the contrast so much there is no useful information in shaded regions. Adjusting the camera for good contrast in shaded regions may result in the camera saturating in the lit regions. It may not be possible to take an image with usable contrast in both the lit and shaded regions of a scene. Shadows complicate machine vision, NIII aims to reduce shadows and improve machine vision.

This thesis explores techniques for removing the appearance of ambient light on a scene from an image, leaving only the intended artificial light, even in cases like daylight with very strong ambient light. NIII removes unintended shadows and illumination from an image, improving the performance of other machine vision
Figure 1.2: Three Key Complications that Shadows Cause in Machine Vision; illustrated with a shadow cast over a black line feature.

1.1 Overview

Natural illumination invariance is produced by excluding natural illumination, while imaging with intentional illumination. Natural illumination can rarely be completely removed, so the objective is to make the intentional illumination as strong as possible compared to the remaining natural illumination so the image appears to be only illuminated by intentional illumination. There are two basic techniques, a) excluding as much natural light as possible from the image, and b) removing the natural illumination contribution to the image using software processing.

Three physical properties of light can be used to separate ambient and intentional light: the wavelength, the time it arrives at the camera and the polarization. The wavelength range can be selected using a narrow wavelength filter and a light source with a similar wavelength range. Natural light typically has a very broad bandwidth (e.g. sunlight), so a bandpass filter can exclude most of the natural
light, while passing most of the intentional light. The arrival time can be used to separate natural from intentional light using the camera shutter. Most natural light sources are continuous, whereas the intentional light can be a flash synchronized to a camera. Using a shutter and a flash reduces the average intentional illumination power. Most natural light, such as sunlight, has a range of polarizations. However, an intentional light source can have a uniform polarization. Using a polarization filter will typically only remove half of ambient illumination. Inevitably some of the intentional illumination will not maintain the original polarization, reducing the intentional illumination imaged. The change in the image appearance due to selective reflection by polarization and the comparatively poor exclusion of ambient light means polarization is not as useful for separating ambient and intentional illumination as wavelength and time. Thus, this research focuses on using wavelength and time to separate the natural illumination from the intentional, flash, illumination.

In most cases some natural illumination will inevitably be imaged with the flash illumination. If the natural illumination contribution to the image can be found, it can be subtracted out, leaving just the intentional flash illumination contribution to the image. Simultaneously taking images at the same wavelength both with and without a flash is impossible, consequently an ambient-illumination-only reference image needs to be estimated. The estimated reference image can then be subtracted from the actual image taken with both flash and ambient illumination, leaving an image of just the flash contribution to the image. This research develops techniques to estimate the ambient-illumination-only reference image, in the presence of a simultaneous flash, and evaluates their effectiveness in various machine vision methods.

1.2 Machine Vision Methods Assisted by Illumination Invariance

Natural-illumination invariant imaging is a building block for other technologies. It rarely allows a user to do something completely new, rather it allows other technologies to operate in more challenging conditions, or with lower power intentional
light sources. Some of the end user applications are explored in Section 1.3. Machine vision problems which can benefit from NIII are discussed here.

Three key machine vision applications for NIII are explored: segmentation, shadow removal and depth correlation. Segmentation is aimed at separating objects to help identify areas of interest in a scene. It can also be used to identify objects or to identify regions which can be ignored or replaced. Shadow removal improves images for other machine vision processing algorithms by reducing or eliminating the apparent shadows. Depth correlation images a pattern projected on an object, and correlates the image with a database of images of the pattern projected on a flat object at known depths from the camera to find the depth of the object. These NIII applications are explored further in the experimental chapters.

1.2.1 Structured Light

Structured Light (SL) is used to get depth points from a scene. It works by projecting a pattern on the scene, locating points in the projected pattern, and using them for triangulation. Structured light only works if the projected points can be correctly identified, which gets more difficult as ambient illumination increases compared to intentional illumination. Additionally, using Near Infrared (NIR) NIII could allow SL to operate simultaneously without being perceptible to people or regular color cameras.

1.2.2 Shadow-Length Depth Measurement

Recent work [21] uses the lengths of the shadows cast by an object under several flashes from different positions to determine the depth of objects. NIII techniques could help this technique operate in a broader range of environments including outdoor moving scenes.

1.2.3 Depth Correlation

Depth Correlation operates by projecting a pattern on a scene which varies with depth. As discussed in Chapter 6, the depth of each image region is found by matching image regions to previously known images of the pattern projected on a
1.2.4 Segmentation

Segmentation is a key machine vision method that separates objects in images to aid in their recognition. Figure 1.1 shows segmentation using a very simple method (50% threshold), which is substantially improved with diffuse lighting as in Figure 1.1a. A myriad of segmentation methods have been developed including using differences in intensity, color or edges. Illumination is typically one of the biggest challenges in segmentation. NIII can improve segmentation for these applications as shown in Chapter 4.

1.2.5 Edge Detection

The three key complications of shadows to machine vision are clearly apparent with edge detection: false edges due to shadows, shadows disrupting real edges and a wide range of image intensity. These issues are demonstrated in Figure 1.3. The image in Figure 1.3a is processed with a Canny edge detector [11] (implemented by edge, [57]), to produce Figure 1.3b. The shadow boundary is identified as an edge, even though there is no object edge there. The different light intensity on the background texture shows no edges in the shadow region, while showing many edges in the lit region. These issues are due to illumination, and can be mitigated with NIII.

1.2.6 Point Identification

Point identification, such as using a corner measure such as [33], has similar problems due to shadows as edge detection. Using regional maxima of the corner measure, as opposed to a global threshold, can help identify points even in darker regions. However, if the darker regions are too dark, since the camera is adjusted for the brightly light regions, there may not be enough information to detect points. NIII can help create a more uniformly illuminated image with identifiable points.
even in regions shaded by the ambient illumination.

1.3 General Applications of Illumination Invariance

The value of NIII is in improving the performance of other machine vision algorithms in outdoor applications. Several existing applications which rely on machine vision and illumination techniques in controlled environments or over limited distances are explored in this section.

1.3.1 Inspection

Visual inspection is frequently used in industrial environments for a range of applications to assess production quality. Printing, labeling and bottle filling are all routinely checked with machine vision. These methods are often limited to operating in controlled environments and are challenged by outdoor environments. NIII can improve outdoor performance for applications in agriculture, silviculture and mining.

1.3.2 Control

Much research related to illumination invariance for control applications has focused on autonomous driving applications with structured light. Outdoor laser line
stripers and pattern projectors have been proposed to help control Mars rovers, [59], buses, [61], self driving vehicles, [60] and [89], and hospital or home robots, [48] and [36]. All of these can only operate within a limited distance, which NIII can improve.

1.3.3 Security

Shadows confound both area video security and facial recognition. Segmenting foreground regions of interest can help security and monitoring applications identify relevant changes [49]. Shadow removal would reduce the challenge from variable illumination in facial recognition [95]. NIII has a variety of applications for security, including improving identification and tracking.

1.3.4 Human-Computer Interfaces

Several human-computer interface technologies, especially in the gaming industry, are using optical systems to understand the positions of people near the device and use their position as an interface method. The Sony PlayStation Move uses a camera looking at colored balls attached to the controllers to determine their positions [51]. The Nintendo Wii uses a camera in the controller imaging LEDs attached to the screen to determine the controller position [79, 66]. The Microsoft X-Box Kinect (formerly Project Natal) uses depth correlation, where it projects a depth-varying pattern on the scene which is then correlated with known reference patterns at different distances to determine distance [28]. All of these technologies require imaging a scene, typically indoors, and identifying an intended light source from ambient patterns. NIII would allow all of these to operate with less power or in more difficult conditions, such as outdoors or in a sunlit room. The Sony and Nintendo solutions both rely on detecting the light sources directly, while the Microsoft solution requires detecting reflections off objects in the scene; which NIII could improve most. NIII-assisted depth segmentation and depth correlation would allow these applications to work in a wider range of conditions.
1.3.5 Filming

Segmentation, or keying, allows people in the foreground to be cut out, and made to appear as though they are in front of a different background. Chroma keying, where the person stands in front of a monochromatic background has been used for decades. It requires a monochromatic background for filming, and the subject cannot have any of the background color on them. Several techniques allowing the subject to stand in front of any background, and using only the depth for segmentation have been proposed including using a second invisible light source and camera in [92], using a time-of-flight camera in [32, 13] and projecting a specially created pattern on the background to make it monochromatic in [31]. NIII based segmentation could be used as a keying method, allowing the subject to be in a wider range of scenes.

1.4 Selected Specific Applications of Illumination Invariance

Automation for open-pit surface mining motivated this research, and there are several specific applications where Illumination Invariance could assist mining processes. Experimenting with these methods in the harsh mining environment, such as extreme temperatures, high vibration levels and substantial dust would have required substantially ruggedizing the apparatus. Consequently the experiments were conducted indoors (Segmentation, Chapter 4; Depth Correlation, Chapter 6) or outdoors (Estimate Accuracy, Chapter 3; Shadow Removal, Chapter 5), in less challenging conditions.

1.4.1 Tooth Detection Systems

Mining shovels have replaceable teeth, which occasionally fall off into the aggregate. The teeth are substantially harder than the rest of the rock aggregate, and can cause substantial damage if they are accidentally fed into a crusher. Promptly detecting if a tooth has fallen off a shovel is an important machine vision application in mining [52]. NIII, used with segmentation (Chapter 4) or Depth Correlation
(Chapter 6) could help create a clear segmentation between the shovel and the background, improving the ability for machine vision algorithms to detect missing teeth. Improved tooth-loss detection could reduce costly equipment damage and production delays.

1.4.2 Aggregate Measurement

Aggregate is the fractured rock resulting after explosives break apart a rock face. It is shoveled into haul trucks and taken to a crusher to be further broken up for processing. The size of the aggregate particles is important to mine efficiency. If particles are too big they can be difficult for the crushers. If they are too small then more explosives than necessary were used. Measuring the aggregate size can improve mine efficiency by optimizing the amount of explosives used.

Machine vision has been applied to aggregate size measurement [53], but it is a challenging problem complicated by shadows cast by passing machines. Natural illumination invariant imaging aims to remove the shadows, as shown in Chapter 5. NIII shadow-free images would be uniformly illuminated, irrespective of the time of day. They would also have a uniform intensity and uniform, or no shadows (depending on the flash configuration) making it easier to segment the aggregate particles. The improved aggregate measurement would reduce the amount of explosives used, and thus costs.

1.4.3 Collision Warning Systems

Collisions between equipment, and other equipment or personnel are hazards of surface mining, as they are in many other environments. Accurately detecting nearby operators and other equipment enables warning the operator or stopping equipment before a collision. NIII techniques applied to segmentation (Chapter 4) or depth correlation (Chapter 6) could improve the detection of nearby objects, helping to reduce the hazards.
1.5 Previous Work

NIII solutions exist for specific problems, with constraints specific to the problem. Generalized NIII for outdoor applications, researched here, is challenging because of scene motion, the lack of prior scene information, the need for fast algorithms for real time applications and the intensity of sunlight. Many previous solutions address elements of these objectives, but none address all of them effectively. The previous work on NIII, and related problems, can be grouped into three categories: illumination and filtering, computational techniques and depth imaging.

1.5.1 Illumination and Filtering

Narrow wavelength range illumination, and filtering to the same wavelength range, has been used to reduce the impact of sunlight. A number of researchers have proposed subtraction and/or filtering. Le Moigne and Waxman [47] is one of the earliest works using a filtered camera and a structured light projector. Subtraction with a single scanned laser point was explored in [102]. Subtraction of a reference ambient-only image from an image of flash and ambient light has been proposed for industrial applications, such as characterizing weld pools, in [15]. Outdoor structured light was proposed for Mars rovers in [59], including lasers and filtered cameras. However the weaker Martian sunlight and the application’s lack of humans potentially harmed by optical radiation, reduced the complexity. More recently [60, 4] implemented a single laser line stripe projector to detect curbs. The recent papers discuss background subtraction, but did not implement it. Using infrared, as opposed to visible light, for structured light applications was discussed in [25] and used for tracking head and eye movement in vehicles using filtered cameras in [42] and [10]. Techniques to reduce the amount of ambient light imaged, and simple subtraction of a background have been implemented, but not the ambient estimation techniques proposed here.

High speed cameras have recently been coupled with DLP projectors in [65] allowing specific projected pixel intensity values to be recognized by the time dithering pattern the projector uses. Recognizing unique pixels intensities is useful for structured light. The technique has not been demonstrated in the presence of strong outside light, where the temporal patterns would have lower relative intensity and
be harder to identify.

**Three-Wavelength Interpolation**

Padilla [72] proposed using a monochromatic flash, one camera bandpass-filtered at the wavelength of the flash, and two cameras, one bandpass-filtered at higher and one bandpass-filtered at a lower wavelengths than the flash. All cameras took images simultaneously, then an estimate of the center wavelength image without the flash was created by interpolating each pixel between the two extreme wavelength images. The interpolation used one set of linear coefficients for the entire image (as opposed to the methods proposed in Chapter 2 which operate pixel-by-pixel), calculated using an image taken without the flash. The estimate of the image without the flash was subtracted from the image taken with the flash to produce an image that appears to be illuminated by only the flash, excluding ambient illumination. Padilla thus created a flash-only image using three simultaneous images taken at different wavelengths, and a set of coefficients found from an earlier image.

Padilla’s method relies on the assumption that the object reflectance does not change substantially between the three wavelengths. (e.g., the scene does not have strong colors.) Figure 1.4 illustrates the errors that result from this assumption. Figure 1.4a shows an image of the scene taken without a flash with a Canon T1i DSLR camera. Figure 1.4b shows a synthetic image of the scene taken with ambient light and a green flash. The synthetic image was created by combining the green channel information from an image taken with a white flash with the red and blue channels from an image taken without a flash (Figure 1.4a). Figure 1.4c shows the flash component of the image generated by subtracting the green channel of 1.4b from the green channel of Figure 1.4a. Figure 1.4d shows the flash-only image estimate generated using Padilla’s method.

Ideally the flash-only image created by Padilla’s method would look like Figure 1.4c. The center green jug is substantially brighter in Figure 1.4d than in Figure 1.4c since it is strongly colored green and thus Padilla’s method does not give an accurate intensity estimate. The brightness of the center jug in Figure 1.4d compared to Figure 1.4c illustrates the limitations of Padilla’s method when the ambient light center frequency channel (i.e. green) can not be accurately estimated.
from the intensities of channels at frequencies (i.e. red and blue) on either side of
the green channel.
A typical color camera (Canon EOS Rebel T1i) with red, green and blue chan-
nels was used for this demonstration for experimental convenience. A custom
implemented apparatus could use narrow wavelength filters with less separation
between wavelengths. The examples in Figure 1.4 illustrate a marked example of
the basic problem and the technical issues with the approach.

1.5.2 Computational Techniques

A range of computational techniques have been proposed to remove shadows from
images. These techniques are strictly useful for object segmentation, recognition
applications or artistic editing. Without any projected light they are not suitable for
depth correlation or reflected flash intensity based depth segmentation. Computa-
tional techniques are based on several principles, including shadow color, sequence processing and quotient imaging.

**Shadow Characteristics**

Outdoor shadows have a different color pallet than sunlit areas. Shadows are not totally dark, but illuminated by diffuse ambient bluish light from the sky. Meanwhile sunlit areas are lit by both direct, more yellowish, sunlight and the diffuse blue light from the sky. This difference can be used to detect shadow boundaries where the change in intensity and color is consistent with a single object under these two different lighting conditions. A number of papers have explored these techniques including [55]. Others such as [50, 22, 23, 24, 26] include relighting the images to appear as though there is no shadow. For specific applications with a well defined scene, such as lane marker detection, intensity thresholding and adjustment can be applied as in [43]. A known background can be adjusted using shadow color to separate foreground objects from their shadows on a known background [35]. These approaches are computationally intensive and depend on a well characterized illuminant which may not be the case in an industrial scene with a variety of artificial lights and sunlight. They are also not useful for active vision techniques such as depth correlation or structured light applications.

**Sequence Processing**

Surveillance or traffic footage is a key application for this technique where the problem is identifying relevant change in the scene (e.g. vehicle or foot traffic) as opposed to irrelevant (e.g. shadows changing as the sun moves across the sky).

Yoon [101] used a sequence of images taken with the scene illuminated from different angles to remove shadows. Wang [98] and Matsushita [58] (based on techniques in [99]) created a background image reference of a traffic scene by extracting constant segments of image sequences, then used this to identify changing objects. The shadows were identified using a model of the scene illumination.

Image sequences taken with deliberately different flash or other illumination can be used to enhance the appearance of photos, or provide limited 3D information. Higher quality images can be produced in low light by combining the large
scale color and intensity information from a no-flash low light image with the details from a flash image, reducing the apparent image noise [74].

Agrawal [1] used flash and no flash images to remove the flash artifacts (e.g. specular highlights) from the flash image, improving the aesthetic quality of the image. Sun [87] used a flash and no-flash pair to segment the foreground, though the optic flow techniques used are computationally intensive.

These techniques improved the appearance of images, or segmented them, though did not specifically remove the shadows.

**Quotient Imaging**

Quotient imaging has been applied principally to facial recognition [96]. It divides the image by a previous model image of the scene. Self quotient imaging uses a smoothed version of the image as the illumination model [97, 67]. These techniques are useful for scenes where the relevant information is high frequency, and the shadows are not, but may not be useful for arbitrary scenes.

**1.5.3 Depth Imaging**

Depth acquisition, using segmentation, depth correlation or structured light are potential applications for NIII. Existing techniques for these problems include range gating, time-of-flight, interferometry and stereo vision.

**Active Depth Imaging**

Depth Imaging techniques provide precise depth information over the image. There are several surveys of optical range imaging techniques, such as [6, 2, 8]. Many of the techniques such as time-of-flight require complex electronics for each measurement. Single point implementations need to scan to cover a scene. There are implementations of time-of-flight cameras such as [45], which are now available commercially, though they have limited range.

Range gating, gates the camera a precise time after emitting a flash [9, 93, 54, 83]. The camera is only exposed for a very brief (typically ns) period corresponding to when the flash would be expected to return from a specified distance. It has
been evaluated for excluding fog or haze from images by not exposing the camera when the flash would be reflecting off haze. The technique can be extended with precise gating to get an approximate depth measure for each pixel over short ranges by measuring the reflected intensity received during the gated period as in [32]. It requires cameras with very fast gating times and powerful pulse light sources.

Time of flight detectors use a modulated light source, then determine distance from the observed phase shift at the receiver requiring complex receiver electronics to determine the relative phase of the modulated signal. Time-of-flight cameras, employing arrays of detectors, have been implemented in [45], [86] and [64], they are now available in higher resolutions commercially, though with limited range.

Interferometry operates similar to time-of-flight cameras, but uses the frequency of light itself. Interferometry provides very high accuracy, but can confuse distances which are integer multiples of the wavelength. Marron [56] has addressed the confusion by using a range of wavelengths. Interferometry is usually used on small distances.

**Stereo Vision**

Stereo vision uses the disparity between two views of the same point to calculate the depth. Mathematically it works similar to structured light, however without a projected pattern, existing points on the scene need to be identified and matched between the views. There has been considerable work on stereo vision, surveyed in [41], includes using nearby points to assist finding correspondences, as in [88]. Gordon [30] combines range (from stereo) and color information to segment the background. Scenes with few easily identifiable point correspondences such as smooth objects, or scenes with easily confused points can challenge stereo vision.

### 1.6 Objectives of This Thesis

The objectives of this thesis are to:

- Examine techniques for improving the reference image used in the image subtraction process.
- Compare the proposed techniques with existing techniques.
• Examine the performance of the proposed techniques in important machine vision applications that can suffer variability due to ambient illumination.

• Examine the bounds on performance of the proposed techniques.

1.7 Organization of This Thesis

This thesis presents ten techniques for NIII in Chapter 2. Chapters 3 to 6 present four experiments comparing the techniques. Chapter 3 evaluates the accuracy of the estimated reference image created using videos without a flash. Chapter 4 evaluates NIII for segmentation applications. Chapter 5 demonstrates NIII for shadow removal. Chapter 6 applies NIII techniques to depth correlation. Chapter 7 analyzes the performance of the techniques and the design issues of the apparatus. Section 8.2 discusses potential further improvements to NIII techniques and apparatus.
Chapter 2

Approach

Trying to overpower an intense ambient source, like sunlight, with artificial lighting continuously and over a wide spectrum would take an enormous illumination intensity (and be unsafe to the eyes of anyone near it). Le Moign and Waxman [47] proposed three elements, discussed further below, of an approach to reduce the apparent ambient intensity:

- Using a short period flash and a camera synchronized to the flash to reduce the cumulative illumination energy required.

- Using a flash with a narrow bandwidth and filtering the camera to the same wavelength.

- Subtracting a reference image, taken under only ambient illumination, at a previous time from the image taken with both the flash and ambient illumination to create a difference image which appears to be principally illuminated by the flash.

This work extends those first steps to improve performance in a moving scene by using a second camera and comparing ten different algorithms presented below and summarized in Table 4.

2.1 Filtering & Flash

Filtering and flash are the most basic ways to reduce the proportion of ambient light compared to intentional light in a scene. Sunlight is around 1000 W/m² on a clear summer day, so competing on sheer power is very difficult. Using a narrow bandwidth light source and filtering the camera to the same wavelength range means the intentional light source is only competing with the ambient source in that narrow
wavelength range. For example using a 10nm bandwidth filter at 500nm would let less than 4% of visible light through. Flashing the intentional light source simultaneously with exposing the camera reduces the average power required by the light source, since it is only active for the brief periods the camera is exposing. For example a 0.5ms exposure at 30Hz would only require 1.5% of the ambient continuous power. Combined, the flash would only require an average power of 0.6W/m², (~1/1500 of the average power of the sunlight) to match the intensity of sunlight in the brief, filtered exposures. The reduced average power improves safety, power consumption and thermal management of the light source.

2.2 Subtraction

While filtering and flash can reduce the average power required to compete with an ambient light source, neither can remove the ambient illumination that subtraction does. Subtracting a reference image taken at a previous time, without a flash, is the simplest computational technique for removing ambient illumination. Reference image subtraction, as proposed in [47] and [60], reduces the ambient lighting visible in the resulting difference image and thus increasing the visibility of the flash.

Notation

The notation used for the images in the equations throughout this thesis includes information for each image about:

- the wavelength - bold face letter: $\mathbf{A}$ for $\lambda_A$ and $\mathbf{B}$ for $\lambda_B$;
- the time - superscript: $\prime$ for $t_1$, $\prime\prime$ for $t_2$, $\prime\prime\prime$ for $t_3$;
- the lighting - subscript: $\mathbf{A}$ for ambient, $\mathbf{F}$ for flash, $\mathbf{AF}$ for ambient and flash; and
- if it is an estimate - $\hat{\cdot}$

All the equations, unless otherwise noted, are scalar operations applied pixel-by-pixel across the whole image. The bold (e.g. $\mathbf{A}$) notation for images is used, usually
Subtraction Algorithm

Direct subtraction operates as in

$$\hat{A}_F'' = A''_{AF} - A'_A$$

and in algorithm format is Algorithm 2.1. Using a reference image simply taken at a previous time, as opposed to estimated, is referred to throughout the thesis as Past-Only Direct or PDIR.

Subtraction, PDIR, is only effective with a static scene. If the scene changes between taking the reference image ($A'_A$) and the ambient plus flash image ($A''_{AF}$) then the difference image ($\hat{A}_F''$) will be distorted where the scene changed.

2.3 Imaging Model

A model of scene imaging is used to develop techniques for creating a better reference image than $A'_A$ to subtract in dynamic scenes where the scene may have changed slightly between $t_1$ and $t_2$. The imaging model is created by considering the path of a photon from a light source to the camera sensor. Imaging models have been proposed by [34, 22, 50] and others. At a given wavelength $\lambda$, a light source will emit a radiant intensity of light $i_\lambda$. The light will be attenuated, by spreading, over the distance from the source to a scene object, and by the angle of the object surface with respect to the light source, represented by the factor $\tau$. $\tau$ is a function of geometry, and is not wavelength-dependent. Light of irradiance $i_\lambda \tau$ then strikes an object with wavelength-dependent reflectance $r_\lambda$ giving a reflected light radiosity $i_\lambda \tau r_\lambda$. The reflectance, $r_\lambda$, is a function of the angles, the surface, and
wavelength. The radiant flux of the reflected light from a given object surface area will decrease with distance according to the inverse square law \([12]\). The object surface area imaged by a given pixel will increase with distance by a square law. The decrease in the radiant flux from a fixed surface area and the increase in surface area with a given pixel with distance cancel out for a large flat surface being imaged, so the imaged radiance is constant with distance. Next the light enters the camera, where the response will be given by the sensitivity of its detector chip, the filter and lens used; these are combined in the camera responsivity \(c_{\lambda}\). Illumination radiant intensity \((i_{\lambda})\), reflectance \((r_{\lambda})\) and camera response \((c_{\lambda})\) must be integrated over a range of wavelengths. The camera detector is substantially linear \([34]\), so the digital pixel value is a linear function of the incident light. Each pixel intensity will be the sum of the \(i_{\lambda}r_{\lambda}c_{\lambda}\) of the individual incident rays. The intensity of light hitting a given pixel \((A_{m,n},\text{in image A})\) can be expressed as

\[
A_{m,n} = \int i_{\lambda}r_{\lambda}c_{\lambda} d\lambda
\]  

(2.2)

2.4 Ratio Based Estimation

To improve the reference image estimate a second camera is added. Unlike the camera bandpass filtered to \(\lambda_A\) (allowing only flash and the ambient light in that bandwidth), the second camera is bandpass filtered at \(\lambda_B\) to produce an image \((B)\) which excludes the flash, but passes ambient light in a band adjacent to the flash frequency. Subtraction is used in Ratio Based Estimation (RBE), but instead of subtracting a reference image taken at a different time, an estimate of the reference image, mitigating changes in the scene, is subtracted. The cameras are narrow bandpass filtered around either \(\lambda_A\) or \(\lambda_B\) (denoted by replacing the subscript \(\lambda\) with A or B) as in:

\[
A_{m,n} = i_{\lambda}r_{\lambda}c_{\lambda}A
\]  

(2.3)

\[
B_{m,n} = i_{\lambda}r_{\lambda}c_{\lambda}B
\]  

(2.4)

Both cameras take images simultaneously at two times to create the estimated
reference image, \( t_1 \) (denoted with a \(^{'}\)) and \( t_2 \) (denoted with \(^{''}\)) giving the two camera images at \( t_1 \) as in

\[
A^{'}_{m,n} = i_A^{'} r_A^{'} c_A
\]  
(2.5)

\[
B^{'}_{m,n} = i_B^{'} r_B^{'} c_B
\]  
(2.6)

and \( t_2 \) as in

\[
A^{''}_{m,n} = i_A^{''} r_A^{''} c_A
\]  
(2.7)

\[
B^{''}_{m,n} = i_B^{''} r_B^{''} c_B
\]  
(2.8)

Ratio Based Estimation (RBE) requires finding an estimate of the reference image \( A^{''}_A \). Consider if the time between \( t_1 \) and \( t_2 \) is infinitesimal, or if the separation is not infinitesimal then there is no change in the scene between \( t_1 \) and \( t_2 \), then

\[
A^{'A} \simeq A^{''A}
\]  
(2.9)

and

\[
B^{'A} \simeq B^{''A}
\]  
(2.10)

In the case where there is no scene motion then subtraction, PDIR, is effective. The techniques proposed here are intended to be used on a moving scene.

Therefore there is a ratio \( R \) between the wavelengths where

\[
R = \frac{A^{'}_A}{B^{'}_A} = \frac{A^{''}_A}{B^{''}_A}
\]  
(2.11)

Which can be rearranged to produce an estimated reference image \( \hat{A}^{''}_A \) as in

\[
\hat{A}^{''}_A = \frac{A^{'}_A}{B^{'}_A} B^{''}_A.
\]  
(2.12)

Even as the time between \( t_1 \) and \( t_2 \) increases, (2.12) will still be mathematically
equivalent in some cases, discussed below. As above, the bold notation (e.g. A) is used to represent scalar operations over a whole image, omitting the indices. In the cases where (2.12) is not mathematically equivalent, it will provide an average improvement as shown in Section 3.3. Equations (2.5), (2.7), (2.6) and (2.8) are substituted into (2.12) to get

$$
\hat{A}'' = \hat{A}'' = \hat{A}'' = \hat{A}''
$$

(2.13)

A basic simplification is that the camera response (c) for a given pixel doesn’t change between $t_1$ and $t_2$ ($c_A' = c_A'' = \hat{c}_A', c_B' = c_B''$). This simplification is accurate unless the camera lens is changed or adjusted between images, allowing (2.13) to simplify to

$$
\hat{A}'' = \hat{A}'' = \hat{A}'' = \hat{A}''
$$

(2.14)

Cases

1. When a Lambertian surface rotates it results in only $\tau$ changing. This case occurs when the angle between the object surface and the light source changes, reducing the incident irradiance by the cosine of the angle between the object surface and the light source, changing the factor $\tau$. Changes in the angle between the surface and the camera do not change the radiant intensity (power per solid angle) incident on the camera since the $1/\cos \theta$ reduction in radiance due to the Lambertian [12] reflectance is canceled by the $1/\cos \theta$ increase in the object surface area per steradian. The light source is constant ($i_A' = i_A'' = \hat{i}_A', i_B' = i_B''$), and the object is Lambertian and constant ($r_A' = r_A'' = \hat{r}_A', r_B' = r_B''$), but the distance or angle between object surface to the light source ($\tau' \neq \tau''$) changes, resulting in

$$
\hat{\tau}'' = \frac{\tau'}{\tau} \tau'' = \tau''
$$

(2.15)

In this case the estimate $\hat{\tau}''$ is mathematically equivalent to $\tau''$ and thus $\hat{A}'' = A''$.

2. Similarly, if the radiant intensity of the light source changes by factor $\beta$, but
not the spectrum, \( i_A' \neq i_A'', i_B' \neq i_B'', i_A' = \beta i_A'', i_B' = \beta i_B'' \), with a constant object reflectance, (2.14) reduces to

\[
\hat{i}_A'' \hat{\tau}'' = \beta i_A'' \tau'' \beta i_B'' \tau'' = \hat{i}_A'' \tau''
\]

and thus again \( \hat{A}'' = A'' \). This case will exist anytime the Spectral Bidirectional Reflectance Distribution Function (SBRDF, expressed as some function \( f(\theta, \phi, \theta_r, \phi_r, \lambda) \) of the incident, \( i \), and reflected, \( r \), spherical angles \( \theta \) and \( \phi \) and wavelength \( \lambda \) [12]) is separable into a spectral component and a Bidirectional Reflectance Function (BRDF) which is only geometrically dependant i.e. \( f(\lambda) g(\theta, \phi, \theta_r, \phi_r) \).

With other changes, especially changes in the object’s reflectance, the estimated reference is not mathematically equivalent to the actual image. In the limited cases discussed above, the RBE techniques create estimates which are mathematically equivalent to the actual images. Only some image regions will be represented by one of the two cases above. In the remainder of regions, where none of the assumptions hold, RBE techniques have been empirically shown to provide a statistical improvement as in Section 3.3.

The ratio between pixel intensities taken at different wavelengths \( \frac{A}{B} \) is not meaningful if the camera values are saturated. It is important to ensure the camera exposure and gain is set to avoid saturation when using RBE. This technique was presented in [77].

### 2.4.1 Apparatus

Ratio Based Estimation uses two simultaneously-triggered cameras with the same view of a scene, narrow bandpass filtered at two different wavelengths and a flash at one of the wavelengths as in Figure 2.1. The cameras were the same model, with matched lenses and aligned so they shared the same view. The physical alignment is imprecise, so software image registration was used to make each pixel of the images from both cameras represent the same point. The two cameras were triggered together to give two simultaneous images of a scene (A and B), with the same view, at different wavelengths \( \lambda_A \) and \( \lambda_B \) respectively. The pair of cameras were exposed
twice, once before the flash ($t_1$ producing $A'_A$ at $\lambda_A$ and $B'_A$ at $\lambda_B$) and once with the flash ($t_2$ producing $A''_A$ with both ambient and flash and $B''_A$ excluding the flash). The three flash-free ($A'_A$, $B'_A$ & $B''_A$) images can then be combined to create the estimated reference image as in (2.12).

For the shadow removal and segmentation applications the flash should be located as close as possible to the camera line of sight to minimize visible shadows cast by the flash. For structured light, and other applications such as using shadows for depth determination [21], a separation between the camera and flash is required. Placement of flash sources for machine vision applications requiring no shadows, or specific shadow characteristics, is discussed in [91]. [91] also proposes that shadows due to the flash could be eliminated by reflecting the flash through a semi-silvered mirror in front of the camera, similar to the $B$ camera, allowing the flash and camera to have the same view of the scene.

Smaller bandpass filter bandwidth allows less ambient light into the image. An ideal flash is monochromatic like a laser, to allow the filter bandwidth to be as small as possible. Excluding more ambient light, allows the camera to have a greater sensitivity, such as a longer exposure or higher gain, without saturating pixels. Higher camera sensitivity allows weaker flashes to be detected, thus allowing the apparatus to operate over a larger distance or use a less powerful flash for the same distance.

The effective range of the apparatus is limited by the radiance of the flash compared to the ambient illumination radiance within the filter pass band. The flash intensity effect on image quality is discussed in depth in Section 7.3.

**2.4.2 Estimation of a Flash-Only, NIII Image**

RBE creates an estimate of the reference image ($\hat{A}''_A$) using (2.12) as it would look with only ambient-illumination in band $\lambda_A$. The estimated reference image is then subtracted from the measurement $A''_{AF}$ (with ambient and flash) as in

$$\hat{A}''_F = A''_{AF} - \hat{A}''_A$$

similar to (2.1) to estimate the flash-only image $\hat{A}''_F$. RBE is computed pixel by pixel over the whole image.
Figure 2.1: RBE Apparatus Diagram
Algorithm 2.2 PRBE: Past-Only Ratio Based Estimation algorithm for each pixel.

1. \( A_{est2a} = \frac{(A1a/B1a) \times B2a}{B2a} \)  // create reference
2. \( A_{est2a} = \min(1, A_{est2a}) \)  // removes numbers over 1
3. \( A_{est2a} = \max(0, A_{est2a}) \)  // removes numbers below 0
4. \( A_{est2f} = A_{2af} - A_{est2a} \)  // calculate flash only

Since estimates (\( \hat{A}''_A \)) outside the pixel intensity range are unrealistic, the estimate is limited to the camera range, for example 0-255 for an 8-bit camera. In this work the camera measurements are normalized to the camera maximum intensity so the estimate range is limited to 0-1 as in

\[
\hat{A}''_A = \begin{cases} 
1 & 1 < \hat{A}''_A \\
\hat{A}''_A & 0 \leq \hat{A}''_A \leq 1 \\
0 & \hat{A}''_A < 0
\end{cases}
\] (2.18)

The PRBE algorithm for each pixel can be stated as Algorithm 2.2.

2.5 Offset Ratio Based Estimation

Cameras output discrete integer values for each pixel intensity. With small values (such as in shaded parts of the image) the change in the ratio between wavelengths (\( A/B \)) due to one integer step gets larger, leading to unreasonable estimates of the reference image. For example, if \( A = 3 \) and \( B = 3 \) the ratio \( \frac{A}{B} = \frac{3}{3} = 1 \). Changing \( B \) by one integer value, to \( B = 2 \), results in a very different ratio of \( \frac{A}{B} = \frac{3}{2} = 1.5 \) resulting in a very different estimate. Whereas if \( A \) and \( B \) are both large (e.g. \( A = B = 100 \Rightarrow \frac{A}{B} = \frac{100}{100} = 1 \)), a single integer change results in a small change in the ratio (e.g. \( A = 100, B = 99 \Rightarrow \frac{A}{B} = \frac{100}{99} = 1.010... \)). Offset Ratio Based Estimation (ORBE) attempts to reduce the impact of quantization by adding an offset (\( q \)) to each of the terms in (2.12) then subtracting the offset out of the result as in:

\[
\hat{A}''_A = \frac{A'_A + q}{B'_A + q} (B''_A + q) - q.
\] (2.19)
Algorithm 2.3 ORBE: Offset Ratio Based Estimation

1. \( \text{Aest2a} = \left( \frac{(A_{1a} + q)}{(B_{1a} + q)} \right) \times \left( \frac{(B_{2a} + q)}{(B_{2a} + q)} \right) - q \)  
   // creates reference
2. \( \text{Aest2a} = \min(1, \text{Aest2a}) \)  
   // removes numbers over 1
3. \( \text{Aest2a} = \max(0, \text{Aest2a}) \)  
   // removes numbers below 0
4. \( \text{Aest2f} = \text{A2af} - \text{Aest2a} \)  
   // calculate flash only

In the examples listed above, and offset of 53 (the 0-255 range integer value of the 0.21 normalized best performing value from Table 3.1) changes the ratio for small values \((A = 3, B = 2)\) to \(\frac{A}{B} = \frac{3 + 53}{2 + 53} = \frac{56}{55} = 1.018\)… and large values \((A = 100, B = 99)\) to \(\frac{A}{B} = \frac{100 + 53}{99 + 53} = \frac{153}{152} = 1.007\)… so the change in the ratio due to a one-value change, potentially due to quantization noise, is much smaller.

ORBE also uses (2.18) to limit the estimate range to the potential range of the camera, and can be implemented as Algorithm 2.3.

Using an offset is an attempt to reduce the impact of the quantization error, which statistically averages to a variance of \(Q^2/12\), where \(Q\) is the quantization step used [34]. The standard deviation of the error introduced due to quantization in (2.19) with respect to \(q\) is given by

\[
\tilde{A}_A'' = \sqrt{\frac{Q^2}{12} \left( \frac{1}{(B_A' + q)(B_A'' + q)} \right)^2 + \left( \frac{(A_A' + q)(B_A'' + q)}{(B_A + q)^2} \right)^2 + \left( \frac{(A_A' + q)}{(B_A + q)} \right)^2}
\]

(2.20)

A larger \(q\) always reduces the effect of quantization noise, but will also reduce the accuracy as it drives \(A_A''/B_A''\) towards unity. ORBE is only useful on scenes where the quantization error dominates the additional estimation error due to adding the offset. Quantization error is typically more substantial in dark images where most of the image intensity values are small. The value of \(q\) is selected by trying a range of values on a sample video and selecting the \(q\) value with the lowest estimate error as discussed in Section 3.3. The offset also tends to wash out the ratio between wavelengths \((A/B)\) towards unity, reducing extreme estimates. While most of the techniques presented can have Direct (-DIR) and RBE variations, Offset
Direct Estimation (ODIR) would be mathematically equivalent to PDIR, and is not examined separately.

2.6 Central Ratio Based Estimation

Central Ratio Based Estimation (CRBE) adds an additional pair of images taken after the flash to improve the estimated reference image. The first \((A', B')\) and third \((A'', B'')\) pairs are taken with only ambient-illumination without the flash \((A_A', B_A', A_A'', B_A'')\). Only the second \((t_2)\) pair is taken with the flash at \(\lambda_A\) only, adding to the existing ambient-illumination, producing images \(A_{AF}'\) (\(\lambda_A\) image with both ambient and flash), and \(B_A''\) (\(\lambda_B\) wavelength ambient-only image - the flash isn’t visible at \(\lambda_B\)). The third image pair increases the latency from the image being taken to the resulting flash-only NIII image estimate being available. The additional image pair provides a significant improvement in performance, as shown below.

CRBE consists of two parts: partial estimation and weighting, each of which is discussed in depth below. The two partial estimation (PE) techniques (direct and RBE) are used to create an estimate using data from two time pairs, either \(t_1t_2\) or \(t_2t_3\). These two estimates are then weighted and added together to come up with a single estimate of the reference image by one of four weighting techniques. The two PE and four weighting techniques can be combined to form six different variations which are evaluated here. The algorithms are abbreviated by specifying the weighting technique, followed by the PE technique. For example, ARBE for “Averaging RBE”.

2.6.1 Partial Estimation Techniques for Obtaining a Reference Image

Two partial estimation techniques are used, Direct (-DIR, is subtraction, based on (2.1)) and Ratio Based Estimation (-RBE, which is based on (2.12)). With three times, two partial estimates can be calculated, one from \(t_1t_2\) (Denoted by \(t_{12}\). The first subscript is the reference time, while the second subscript is the time the PE is created to estimate. ), the second partial estimate is created from \(t_2t_3\) (\(t_{32}\). DIR
and RBE partial estimation techniques have different advantages.

Direct estimation is computationally fast, but has poorer accuracy than RBE. Direct estimation with weighting may be useful for some computation-limited applications, even though it generally does not perform as well as RBE. In all the techniques examined, the performance will vary with the degree of change in the scene between image pairs.

The direct partial estimation technique is simply the $t_1$ and $t_3$ ambient-only images respectively. The previous time ($t_{12}$) and future time ($t_{32}$) partial estimates using DIR are:

$$\hat{A}_{t_{12} \text{DIR}}'' = A'$$

$$\hat{A}_{t_{32} \text{DIR}}'' = A'''$$

RBE, as in sec. 2.4.2, uses information from the B camera to calculate a better estimate using a ratio. CRBE uses the B image both for estimating and weighting the $\hat{A}''$ image. RBE is usually more accurate than Direct Estimation, but requires more computation. The previous time ($t_{12}$) and future time ($t_{32}$) estimates using RBE are:

$$\hat{A}_{t_{12} \text{RBE}}'' = \frac{A'}{B'}B''$$

$$\hat{A}_{t_{32} \text{RBE}}'' = \frac{A'''}{B'''}B''$$

### 2.6.2 Weighting Techniques for Obtaining a Reference Image

Weighting involves selecting which of the two partial estimates, either from $t_{12}$ or $t_{32}$ is more accurate and weighting it more heavily. Four weighting techniques are discussed and compared below: past-only (abbreviated by a “P-”), averaging (“A-”), linear weighting (“L-”) and sigmoid weighting (“S-”). Section 2.4 presented Past-only RBE with image pairs taken at only two times (before the flash, $t_1$ and with the flash, $t_2$). With additional time to get an image pair taken after the flash ($t_3$) either PE technique can be used twice, once with each of the two time pairs, using $t_{1,2}$, (denoted $t_{12}$) and $t_{2,3}$ ($t_{32}$) to create two possible PEs for the image at $t_2$. The challenge is weighting the two estimates ($t_{12,32}$) for the best overall estimate.

30
Weighting is finding the optimal coefficient $\rho$ on a pixel-by-pixel basis as in

$$\hat{A}'' = \rho \hat{A}_{t_{12}}'' + (1 - \rho) \hat{A}_{t_{32}}''$$  \hspace{1cm} (2.25)

where $\rho$ is a number between 0 and 1. Partial estimates from either PE technique can be weighted using (2.25).

Algorithms to find $\rho$ should satisfy several requirements. Since the $A''$ image has the additional flash illumination, $A$ camera images cannot be used in selecting the correct weighting. The $B$ camera images, which are taken at all three times, and don’t image the flash, are used to weight the estimates. The more likely estimate is found from whether $B''$ is more similar to $B'$ or $B'''$. The weighting technique should be symmetric, so that a change from light to dark and a change from dark to light are treated the same way. (i.e. equivalent behavior if $B' > B'''$ or $B' < B'''$)

The weighting technique must provide a value for $\rho$ over all possible values of $B', B''$ and $B'''$, including if $B' & B''' > B''$ and $B' & B''' < B''$. These requirements limit the choice of weighting, but ensure it will operate in all cases.

Four weighting techniques are examined here and are illustrated in Figure 2.2.

- Past-only (P) sets $\rho = 1$ and completely ignores the image pair taken at $t_3$. The past-only techniques (PDIR, PRBE) were presented in Section 2.4. PDIR is the simplest technique and is used as a performance baseline for other variations.

- Averaging (A) sets $\rho = 1/2$, which equally weights the $t_{12}$ and $t_{32}$ estimates. PDIR and ADIR do not require the $B$ camera for either weighting or partial estimation, saving on hardware complexity at the cost of reduced accuracy.

- Linear-Weighting (L) is where the value of $\rho$ is found with a straight line connecting $\rho = 1$ at $B'$ and $\rho = 0$ at $B'''$ as in

$$\bar{\rho} = \frac{B''' - B''}{B''' - B'}$$  \hspace{1cm} (2.26)
Figure 2.2: The four weighting techniques used, including the impact of a different $\alpha$ on the sigmoid weighting techniques. The x-axis is the value of a pixel in $B''$ normalized to the range of the same pixel from $B'$ and $B'''$. For example if $B' = B''$ then $\rho = 1$ for Linear or Past-Only.
then limiting $\bar{\rho}$ to 0 to 1 if $B''$ is outside the range of $B'$ to $B'''$ to get $\rho$ as in

$$
\rho = \begin{cases} 
1 & 1 < \bar{\rho} \\
\bar{\rho} & 0 \leq \bar{\rho} \leq 1 \\
0 & \bar{\rho} < 0
\end{cases}
$$

Equation (2.27)

- Sigmoid-Weighting (S) is the most complex technique it is given by

$$
\rho = \frac{1}{1 + \exp(-\gamma \alpha)}
$$

Equation (2.28)

The sigmoid function relies on two coefficients $\gamma$ and $\alpha$. The $\gamma$ term is calculated for each pixel using

$$
\gamma = \frac{B'' - \frac{1}{2}(B' + B''')}{B' - B'''}
$$

Equation (2.29)

Then it uses the difference between $B''$ and the mean of the $B'$ and $B'''$ to get the x-value for the sigmoid.

The second coefficient, $\alpha$, is an image-wide constant chosen to set the width of the function. A very high $\alpha$ value makes the sigmoid more like a step function, while a low $\alpha$ value reduces the slope, as shown in Figure 2.2. An initial estimate of $\alpha = 4.4$ was found by assuming that at $B'$, $\rho = 0.9$ . Experiments (see Section (3.4.2)) were run at a range of $\alpha$ values around this initial value to find the best performance, which will be scene dependent.

There is no need to limit the value of $\rho$ in a separate step since the sigmoid function cannot be less than 0 or greater than 1. It will never equal exactly 1 or 0, only approaching them asymptotically as the magnitude of $\gamma$ increases.

### 2.6.3 Combined Algorithms

The weighting techniques and partial estimation techniques are paired into eight algorithms, the first two of which (PDIR, PRBE) are discussed in the previous sections. Average RBE (ARBE), Algorithm 2.4, uses the average of past and future
Algorithm 2.4 ARBE: Average Ratio Based Estimation

1 Aest2a = (A1a/B1a + A3a/B3a) / 2 * B2a) // creates reference
2 Aest2a = min(1, Aest2a) // removes numbers over 1
3 Aest2a = max(0, Aest2a) // removes numbers below 0
4 Aest2f = A2af − Aest2a // calculate flash only

Algorithm 2.5 ADIR: Average Direct Estimation

1 Aest2a = (A1a + A3a) / 2 // creates reference
2 Aest2f = A2af − Aest2a // calculate flash only

Algorithm 2.6 LRBE: Linear Ratio Based Estimation

1 rho = (B3a−B2a)/(B3a−B1a) // creates rho
2 rho = min(1, rho) // removes rho values over 1
3 rho = max(0, rho) // removes rho values under 0
4 Aest2a = ((A1a/B1a) * rho + (A3a/B3a) * (1−rho)) * B2a) // creates reference
5 Aest2a = min(1, Aest2a) // removes numbers over 1
6 Aest2a = max(0, Aest2a) // removes numbers below 0
7 Aest2f = A2af − Aest2a // calculate flash only

Times; it is computationally simple. Average Direct (ADIR), Algorithm 2.5 takes advantage of images before and after, but does not require the λB camera; it can be implemented using only the λA camera. Linear RBE (LRBE) combines linear weighting with ratio based estimation as in Algorithm 2.6. Linear direct, LDIR, as in Algorithm 2.7, uses the linear weighting but without RBE. Replacing the linear weighting with sigmoid weighting results in sigmoid RBE (SRBE, Algorithm 2.8) and sigmoid direct (SDIR, Algorithm 2.9). The sigmoid implementations don’t need to limit the value of ρ (as in line 2,3 of Algorithm 2.6) since the sigmoid function can only produce ρ values between zero and 1. These are summarized in Table 4. The relative performance of these algorithms is explored in the experimental work Chapters 3-4.
Algorithm 2.7 LDIR: Linear Direct Estimation

1. \( \rho = \frac{(B_3a - B_2a)}{(B_3a - B_1a)} \) \hspace{1em} // creates \( \rho \)
2. \( \rho = \min(1, \rho) \) \hspace{1em} // removes \( \rho \) values over 1
3. \( \rho = \max(0, \rho) \) \hspace{1em} // removes \( \rho \) values under 0
4. \( A_{est2a} = A_1a \cdot \rho + A_3a \cdot (1 - \rho) \) \hspace{1em} // creates reference
5. \( A_{est2f} = A_{2af} - A_{est2a} \) \hspace{1em} // calculate flash only

Algorithm 2.8 SRBE: Sigmoid Ratio Based Estimation

1. \( \gamma = \frac{(B_2a - (B_1a + B_3a)/2)}{(B_1a - B_3a)} \) \hspace{1em} // calculate \( \gamma \)
2. \( \rho = 1/(1 + \exp(-\alpha \cdot \gamma)) \) \hspace{1em} // calculate \( \rho \)
3. \( A_{est2a} = ((A_1a/B_1a) \cdot \rho + (A_3a/B_3a) \cdot (1 - \rho)) \cdot B_2a \) \hspace{1em} // creates reference
4. \( A_{est2a} = \min(1, A_{est2a}) \) \hspace{1em} // removes numbers over 1
5. \( A_{est2a} = \max(0, A_{est2a}) \) \hspace{1em} // removes numbers below 0
6. \( A_{est2f} = A_{2af} - A_{est2a} \) \hspace{1em} // calculate flash only

Algorithm 2.9 SDIR: Sigmoid Direct Estimation

1. \( \gamma = \frac{(B_2a - (B_1a + B_3a)/2)}{(B_1a - B_3a)} \) \hspace{1em} // calculate \( \gamma \)
2. \( \rho = 1/(1 + \exp(-\alpha \cdot \gamma)) \) \hspace{1em} // calculate \( \rho \)
3. \( A_{est2a} = A_1a \cdot \rho + A_3a \cdot (1 - \rho) \) \hspace{1em} // creates reference
4. \( A_{est2f} = A_{2af} - A_{est2a} \) \hspace{1em} // calculate flash only
Algorithm 2.10 TSEL: Time Selection

1. \[ \text{if } \text{abs} \left( B_{2a} - B_{1a} \right) < \text{abs} \left( B_{2a} - B_{3a} \right) \]
2. \[ \quad / / \text{find which is closer in } B \]
3. \[ \text{A}_{\text{est}2a} = A_{1a} \quad / / \text{apply that to } A \]
4. \[ \text{else} \]
5. \[ \quad \text{A}_{\text{est}2a} = A_{3a} \]
6. \[ \text{A}_{\text{est}2f} = A_{2af} - \text{A}_{\text{est}2a} \quad / / \text{calculate flash only} \]

2.7 Time Selection

The Time Selection (TSEL) is the simplest technique using two cameras. It uses the \( \lambda_B \) camera to find which of the \( t_3 \) or \( t_1 \) is most similar to the \( t_2 \) image on a pixel by pixel basis, then applying that to the \( \lambda_A \) image as in

\[
\hat{A}_{\lambda_A}'' = \begin{cases} 
A'_{\lambda_A} & |B'_{\lambda_A} - B''_{\lambda_A}| < |B'''_{\lambda_A} - B''_{\lambda_A}| \\
A''_{\lambda_A} & |B'_{\lambda_A} - B''_{\lambda_A}| > |B'''_{\lambda_A} - B''_{\lambda_A}| 
\end{cases} \quad (2.30)
\]

This is represented in Algorithm 2.10. TSEL is computationally very simple.
Chapter 3

Ambient-Illumination Only
Reference Image Estimate
Accuracy

NIII techniques focus on creating an accurate estimate of what the reference image $A''_A$ looks like without the flash. The reference image estimate accuracy was quantitatively evaluated by comparing an actual image of $A''_A$ obtained without using the flash at time $t_2$ to the estimated reference image $\hat{A}''_A$ obtained by the proposed techniques. A dual-camera system with two, synchronized, aligned, narrow bandwidth cameras was built to collect videos. The camera system was used to collect 30 video data sets which were evaluated using all ten NIII techniques and sweeping a range of parameters for the techniques as required. All the proposed NIII techniques demonstrated an improvement over the baseline PDIR technique proposed in previous literature.

3.1 Accuracy Evaluation

Ideally the estimated reference image $\hat{A}''_A$ exactly matches the $A''_A$ image taken without flash and any difference is an error; evaluating this error was used to evaluate the technique performance. In normal operation, given by (2.17), NIII techniques take a real image at $t_2$ in wavelength $\lambda_A$ ($A''_{AF}$) with both flash and ambient illumination. The estimated reference image ($\hat{A}''_A$) is then subtracted to get the estimate of the flash only image ($\hat{A}''_F$) as in (2.1). In this Section however, the flash is turned off, and thus, in (2.1), the actual image should ideally be equal to the estimate of the reference image ($A''_A = \hat{A}''_A$). The difference between the two images should

37
ideally be zero ($\hat{A}''_F = 0$) in these experiments, and is given by

$$\hat{A}''_F = A''_A - \hat{A}''_A$$  \hspace{1cm} (3.1)

The difference between the images ($\hat{A}''_F$) is summed up over the image to get the rms residual ($\varepsilon$) as described below. This analysis can be applied to any multiwavelength (including RGB color) video sequence taken without a flash to assess the accuracy of a NIII technique.

### 3.1.1 Rms Residual, $\varepsilon$

The rms residual ($\varepsilon$) is used for comparison since it penalizes large residuals more than small ones. The rms residual between the $\hat{A}''_A$ and $A''_A$ images were summed over the image (over the total image rows $M$, and total columns $N$, with indices $m, n$ respectively) using

$$\varepsilon = \sqrt{\frac{\sum_{m=1}^{M} \sum_{n=1}^{N} |A''_{m,n} - \hat{A}''_{m,n}|^2}{MN}}$$  \hspace{1cm} (3.2)

Machine vision algorithms routinely deal with small errors due to inherent camera noise, which are not significant. Large residuals are more significant, and likely to disrupt any algorithms applied to the flash-only image output. Rms error is a typical noise measure for machine vision and imaging applications.

### 3.2 Dual Camera Apparatus

A combined dual-camera apparatus was constructed to capture the simultaneously triggered, aligned, narrow wavelength images to evaluate NIII techniques. The proposed illumination invariant imaging techniques require two simultaneous images in different wavelengths, requiring two cameras with different narrow wavelength filters. A semi-silvered mirror is used to give both cameras the same view. It is shown conceptually in Figure 2.1, as a rendering including the internal layout in Figure 3.1 and as constructed in Figure 3.2.

The semi-silvered mirror and filters are placed in front of the lenses, as opposed
Figure 3.1: Rendering of Dual Camera System

Figure 3.2: Image of the Completed Dual Camera System
to behind a single lens. Placing the semi-silvered mirror or the filter behind the lens would increase the back-focal length of the camera, significantly reducing the field of vision of the camera without lenses designed for a longer back-focal length. The combined apparatus provided simultaneous images of the same view, in different wavelengths.

### 3.2.1 Camera

Point Grey Research monochrome Grasshopper GRAS-14S3M-C [75] cameras were used. The key specifications are:

- Direct control over absolute gain, exposure and processing allows close control over camera functions and consistent settings between cameras.
- External triggering allows external control to synchronize camera exposure with external flash and each other.
- Image tagging allows the images to be marked with the gain, shutter and a digital line in indicating whether the flash fired to improve image handling.
- Pixel binning allows lower noise images at low resolution settings.
- 14-bit A/D converter allows for higher precision pixel measurement.
- Hardware image flipping for the image viewed through the mirror reduces the computational load on the host.

The cameras were operated in a binned mode where they output 640x480 images created in the camera by averaging 4 adjacent pixels from a 1384x1032 image cropped to 1280x960. The images were output from the camera and recorded using 16-bits for each pixel. The Grasshopper cameras synchronized their exposures over the Firewire bus for the estimate accuracy experiments in this chapter, while an external trigger (discussed in Section 4.1.2) was used for the experiments in Chapters 4-6.

Typical color cameras cannot be used for these NIII techniques since the techniques requires thicker interference filters, with a very narrow passband. A specially manufactured camera sensor with specialized filters could be used, as discussed in Section 8.2.7. Most color cameras use a single lens and a sensor with
tiny dye color filters, with a wider passband, over each pixel in an alternating pattern to take an image at three wavelengths. This, and the need to reconfigure easily for different experiments, drives the apparatus to use two separate cameras, each with its own filter, as opposed to a single camera with filters built onto the sensor. Three-chip color cameras, where a prism separates the image to three separate sensors, could be used if the filters were modified to the wavelengths required for NIII. Modifying a three-chip camera was deemed more complex than constructing the dual camera apparatus since the filters would need to be modified and most commercial camera software does not allow the low-level hardware control used in these experiments.

3.2.2 Lenses

Edmund Optics NT56-788 fixed focus 16mm lenses were used. They were selected for their large (f/1.4) aperture and 22.7° field of view with the Grasshopper cameras 1/2” sensor and C-mount attachment [18]. The lens aperture could be adjusted down to f/14, but was set fully open. Both lenses needed to be focused for the working distance independently, care was taken to ensure they were focused similarly so the images would match. Two methods were used to ensure the focus quality. First, manual inspection of the images was used. Second a software tool was created to display the highest difference between any adjacent pixels. If the image is out of focus, the blurring will result in a low gradient. Only a properly focused image will have high gradients between adjacent pixels.

3.2.3 Filters

Narrow band (typically 10nm FWHM) filters were used for the experiments. Interference filters, also referred to as Fabry-Perot Etalons, are a common type of narrow-bandpass filter. They work by creating destructive interference at wavelengths outside the target wavelength [100, 78]. They are created with three layers of different refractive index materials. The layers are separated by a distance \( d \) given in (3.3), where \( \lambda \) is the wavelength of the light in the center medium and \( n \) is any integer. The disadvantage of interference filters is they are sensitive to the incident angle of the light. If the incident angle of the light isn’t perpendicular, the
apparent distance between the layers changes, thus changing the wavelength of the light being filtered and limiting the field of view of the NIII system [5].

\[ d = \frac{\lambda n}{2} \]  

(3.3)

The peak transmission was very low, often only 50% of in-band light [16]. Newer filter models pass up to 85% [17].

The small amount of light transmitted by these filters required a longer exposure time (as high as 10ms) and higher camera gain, contributing to grainier images and more blurring due to motion. The narrow bandwidth filters were used for most experiments, but were dependent on the incident angle of the light, limiting the field of view and requiring longer exposure times.

For these tests the A camera filter was centered at 520nm (green) while the B camera was filter was centered at 580nm (yellow), both had a bandwidth of approximately 10nm and a peak transmission of approximately 50%. The wavelengths are separated to ensure the wide bandwidth flash was not visible in the \( \lambda_B \) camera.

### 3.2.4 Dual Camera Registration

The dual camera apparatus needed to produce two images, at different wavelengths, which were registered with each other. Two methods were used to ensure registration, physical design and software compensation. Ideally the physical design would provide good registration, practically software registration was also needed.

#### Hardware Alignment

The dual camera base was designed so that the cameras would have equal length optical paths and the same view. The cameras and lenses were a matched pair. The camera lenses were fixed focal length to prevent different fields of view. The camera looking through the mirror was aligned accounting for the refraction through the semi silvered mirror. The physical alignment proved inadequate, but software alignment was able to compensate.
Software Alignment

Software image alignment was required. Calibration images were taken of a set of alignment targets with both cameras and corresponding alignment points were manually selected in both images, as shown in Figure 3.3. The point correspondences were used to create an affine transform registering one image to the other. Software optimization was used to improve the point correspondences and the final transform. The affine transform was created using Matlab and applied in both Matlab experimental code and in the RBE/GUI software (explained in Section 4.1.3) using the IPP affine transform function. The registered images were then cropped to the region overlapped by both images.

The quality of the registration was verified both visually and numerically. A visual check was conducted by creating an image where the registered images at different wavelengths were used for different color channels. Any registration errors were visible as colored bars at edges. Numerical verification (part of the optimization) was done by normalizing the intensity range of each image (to adjust for different exposures), subtracting the two images, and summing the difference. Slight variations in the registration transform were tested to ensure the transform had the minimum difference. Using visual and numerical techniques ensured a good fit.

3.3 Experiment

A set of 30 video sequences (100 frames each, at 640x480, 8-bit per pixel, for each of A and B) was taken of different outdoor scenes using the dual camera apparatus, but with no flash. The camera was setup on a tripod with the two cameras running synchronized and filtered to different wavelengths. The sequences all included static elements, such a buildings, and moving elements such as pedestrians and cars at distances from a few meters to approximately 200m. The sequences were all taken under partly cloudy conditions. The images from the separate cameras were rectified together and cropped to the overlapping area. These sequences, such as the examples in Figure 3.4 and corresponding videos listed in Appendix A, were used to evaluate the different NIII techniques.
Figure 3.3: Alignment Calibration Image. The red arrows illustrate the shift between the reference points on the image shown, and the same points on the image being registered.
3.4 Results

Table 3.1 summarizes the residual results for each NIII technique. They are presented as the rms residual ($\varepsilon$, defined in Section 3.1.1) normalized to the camera range and as a percentage of PDIR residual. The addition of an extra image pair at $t_3$ (A,L & S weightings) improves the performance compared to the past-only (P) weightings. RBE techniques consistently outperform DIR techniques. The data shows that, for the same PE technique, the linear (L) and sigmoid (S) weighting techniques perform better than past-only (P) and averaging (A) weighting techniques. RBE techniques also provide a lower standard deviation ($\sigma$) of residuals, which indicates a more consistent performance over a range of scenes. TSEL offered an improvement over PDIR and ADIR, but did not perform as well as the slightly more complex LDIR or any of the RBE techniques. LRBE provides a good trade-off between computational demands (discussed in Section 7.4) and residual performance and is the primary algorithm considered further.

3.4.1 $q$ Selection for Offset Ratio Based Estimation

Figure 3.5 shows the ORBE rms residuals from the 30 videos using different values of $q$. A small $q$ slightly reduces the mean residual. Small values of $q$ had little
### Table 3.1: NIII Technique Performance.

<table>
<thead>
<tr>
<th>NIII Technique</th>
<th>Result ((\varepsilon), rms residual)</th>
<th>(% of PDIR)</th>
<th>(\sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past-only Direct (PDIR)</td>
<td>0.100</td>
<td>100%</td>
<td>0.068</td>
</tr>
<tr>
<td>Past-only RBE (PRBE)</td>
<td>0.0492</td>
<td>49%</td>
<td>0.024</td>
</tr>
<tr>
<td>Offset RBE (ORBE) (q = 0.21)</td>
<td>0.0485</td>
<td>48%</td>
<td>0.024</td>
</tr>
<tr>
<td>Average Direct (ADIR)</td>
<td>0.079</td>
<td>79%</td>
<td>0.054</td>
</tr>
<tr>
<td>Average RBE (ARBE)</td>
<td>0.042</td>
<td>42%</td>
<td>0.021</td>
</tr>
<tr>
<td>Linear Weight Direct (LDIR)</td>
<td>0.057</td>
<td>56%</td>
<td>0.036</td>
</tr>
<tr>
<td>Linear Weight RBE (LRBE)</td>
<td>0.038</td>
<td>38%</td>
<td>0.017</td>
</tr>
<tr>
<td>Sigmoid Weight Direct (SDIR) (\alpha = 5)</td>
<td>0.057</td>
<td>56%</td>
<td>0.037</td>
</tr>
<tr>
<td>Sigmoid Weight RBE (SRBE) (\alpha = 5)</td>
<td>0.037</td>
<td>37%</td>
<td>0.017</td>
</tr>
<tr>
<td>Time Select (TSEL)</td>
<td>0.061</td>
<td>61%</td>
<td>0.039</td>
</tr>
</tbody>
</table>

The residual, \(\varepsilon\), defined in Section 3.1.1, is also presented as a percentage of the PDIR residual level. The sigmoid weighting (SDIR, SRBE) and ORBE results presented are the \(\alpha\) and \(q\) with the lowest residual respectively in each technique. The equations and algorithms for techniques are summarized in Table 4.

Impact, while large values of \(q\) cause the ratio between wavelengths \(\frac{A_i}{B_i+q}\) to tend to unity. The offset performance and optimum \(q\) value are highly scene dependent as shown in Figure 3.6. ORBE isn’t explored further, due to the performance variability and the increased computing requirements.

#### 3.4.2 \(\alpha\) Selection for Sigmoid-Weighted Estimation

Figure 3.7 shows SRBE performance by different sigmoid coefficient, \(\alpha\), values. SRBE offers only a very small improvement over LRBE, but requires far more computing operations. Figure 3.8 shows the sigmoid weighted (SRBE) performance normalized to the linear weighted (LRBE) performance with different sequences selected for this example. The optimal \(\alpha\) choice for the sequences in Figure 3.8 varies, illustrating that \(\alpha\) must be carefully chosen for a specific scene. In some cases SRBE is always better than LRBE, in others it is significantly worse depending on \(\alpha\). High values of \(\alpha\) cause an abrupt change in the weighting. SRBE resembles TSEL as \(\alpha\) gets very high. One reason for the weak improvement using SRBE with low values of \(\alpha\) is the sigmoid function never reaches exactly 1
Figure 3.5: Offset Ratio Based Estimation Performance v. $q$
Figure 3.6: Variation in ORBE Performance. Each color shows the ratio of ORBE performance to PRBE performance for selected video frames. The frames were selected to illustrate the wide performance range of ORBE.
or 0. When the $t_2$ image exactly matches the $t_1$ or $t_3$ image, sigmoid weighting still includes the other dissimilar estimate at a low weight, often resulting in poorer performance. Similar to ORBE, SRBE can provide a small average improvement at a high computational cost and a risk of far worse performance due to the wrong value of $\alpha$, and is not examined further.

### 3.5 Summary

A dual-camera system allowing validation of NIII techniques was constructed. The NIII techniques examined allowed creating estimated reference images with as little as 37% of the residual noise compared to the baseline PDIR technique. RBE techniques always performed better than a similarly weighted DIR techniques by at least 18%. ORBE and SRBE both offered small (<2%) advantages over simpler
Figure 3.8: SRBE Performance Variation. Each color shows the SRBE residual/LRBE residual by $\alpha$ for selected frames. The frames were selected to illustrate the variation in performance.
(PRBE, LRBE respectively) algorithms, but at computational cost and requiring a scene-specific constant. LRBE was selected as the principal algorithm for further analysis, with PDIR evaluated as a baseline.
Chapter 4

Segmentation

The first application investigated for NIII is object segmentation of moving scenes. In object segmentation the challenge is separating the moving object from its own shadow. Segmentation could improve performance in applications such as tooth detection (Section 1.4.1) and collision warning (Section 1.4.3), among others. It is explored using a dark-colored moving object close to a light-colored background so the object casts a shadow on the background. Threshold segmentation performance using raw images, PDIR flash only-images and LDIR flash-only images, was evaluated analytically using a camera noise model and the results are presented in Section 4.3.

4.1 Experimental Setup

The segmentation experiments were conducted with the apparatus shown in Figure 4.1. The dual camera apparatus (a) with an LED flash light source (b) and trigger mechanism was used with the recording software. The scene was illuminated with four 500W halogen lights (c) at approximately 3m distance to simulate a strongly lit ambient scene. The scene consisted of a dark object (e) which casts a shadow (f) on a light background (d). A simple segmentation algorithm was used, and the results were compared with a manually segmented ground truth. A pendulum was used to provide a repeatable moving scene.

These experiments apply a simple segmentation algorithm to a challenging segmentation problem and compare the performance of raw and NIII images. Automatic segmentation was done by applying an intensity threshold to the image. The threshold level was automatically determined for each frame using Otsu’s method [71], which is implemented as greythresh in Matlab. The automatic segmentation was then compared to a manually segmented ground truth.
Figure 4.1: Segmentation Experimental Setup
4.1.1 Light Source

LEDs and lasers can be used for NIII. In order to match the narrow filter bandwidth (needed to exclude as much ambient light as possible, while allowing as much flash light as possible), the light source needs to have a similar peak wavelength and bandwidth to the filter. The light source needs to flash with short rise and fall times so the flash is at full power for the $t_2$ (flash+ambient) exposure, but off for the subsequent $t_3$ (ambient-only) exposure. LEDs meet both of these requirements with a narrow spectral bandwidth and short rise and fall times. Lasers could be used, as they have a desirable narrower bandwidth, and are able to flash very quickly, but they are more expensive. For these experiments LEDs are used.

The LED used was an OSRAM green LED array (OSRAM LE T H3A-KBMA-24, [70]) with a Mightex driver (Mightex SLC-SV04-5 [62]) emitting 481lm (approximately 1W) over a 2.1mm $\times$ 3.2mm emission area operating at 520nm (green) with a 44nm FWHM. The OSRAM green module was in part selected for its small size and emission area, allowing it to be integrated into the projector used in Chapter 6.

4.1.2 Trigger

A single trigger was required to trigger both cameras simultaneously, and the flash simultaneously on every third trigger. All the data sets were recorded with one flash every third trigger. NIII techniques not requiring images at all three times (e.g. PRBE, ORBE) simply ignored the extra data. The flash trigger could have been increased to every second exposure by using the $t_3$ exposure from the previous sequence as the $t_1$ image for the next sequence. The $t_3$ and $t_1$ images were not overlapped in these experiments to preclude mathematical correlation between sequences for validation. While these experiments used a flash trigger every third camera exposure, a deployed system could use the flash every alternate exposure with the proposed NIII methods.

An Altera DE1 development board with a Cyclone II FPGA (Field Programmable Gate Array) was used to generate the trigger signals for the cameras and flash. It provided excellent configurability, including the ability to change the timing of the trigger signals relative to each other to account for different rise and propagation
times. The FPGA was far more complex than required for the application, but was readily available and easy to program and integrate. Images were triggered at 24Hz, enabling NIII estimates at 8Hz.

4.1.3 Software

Operating software, named RBE/GUI, was developed to run the apparatus, record data and do real time analysis. The software was developed using the Qt GUI library on Ubuntu Linux 9.04. RBE/GUI features include:

- All camera, flash and pattern parameters were controlled via an xml configuration file.
- Images were buffered then saved in floating point TIFF files. Buffered recording was required since the hard drive couldn’t keep up with the data rate of the camera.
- Hardware device control including:
  - Control of the Mightex flash driver for the OSRAM LED was carried out via serial communications.
  - Control of the remote projector display computer for pattern selection was carried out via UDP (over ethernet). - *Used for depth correlation tests in Chapter 6.*
  - Control of the cameras, including mode 7, image tagging, shutter, gain and trigger settings was carried out via 1384b (Firewire) data bus.

Example screen shots are shown in Figure 4.2, of the main window and in Figure 4.3 of the image viewer window. The main window allows loading and saving xml configuration files for the apparatus, automatically setting the camera exposure settings, selecting the NIII algorithm used and triggering either immediate or delayed image recording. Delayed image recording was used to allow the operator to trigger the system, then start the scene motion. RBE/GUI integrated the apparatus and ensured experimental configurations were consistent and recorded.
Figure 4.2: RBE/GUI Main Window

Figure 4.3: RBE/GUI Image Viewer Window
4.2 Object Segmentation

Object segmentation experiments performed here attempted to identify a dark object on a light background; it was confounded by the object casting a dark shadow on the background. Object segmentation scenes were constructed with a light background and a dark object casting a shadow on the background, as in Figure 4.4a. Two objects were used, a wooden board and a plastic funnel; both were swung in a pendulum motion. The aim of the experiment is to correctly identify the object, while ignoring the shadow or background throughout a 50 frame NIII video.

The segmentation performance can be shown qualitatively. Figures 4.4 and 4.5 are examples from videos showing the performance of the automatic segmentation using LRBE. The top row of each figure shows the images created, while the bottom shows the same image automatically segmented and compared to a manual segmentation. On the bottom row red indicates an incorrectly segmented pixel, while green indicates a correctly segmented pixel. Each column represents a different imaging technique. The first column is the raw, unprocessed image taken with the flash at $t_2$ (image in (a) and segmented in (d)). The second frame is using PDIR (image in (b) and segmented in (e)); it is simply the $t_2$ image in the first column minus the image taken at $t_1$ (not shown). The final column is using LRBE (image in (c) and segmented in (f)). The videos show segmentation of LRBE images is visibly better than PDIR or raw images.

The improvements due to PDIR and LRBE can be clearly seen in Figure 4.4. In Figure 4.4a there is a strong moving shadow of the board and a static shadow on the right side. Both of which are segmented incorrectly as seen in Figure 4.4d. PDIR (Figure 4.4b) is effective at removing the static shadow to the right, but leaves ghosts of the previous position of the board resulting in poor segmentation (Figure 4.4e) since the shadow is more prominent than the object. Finally, using LRBE (Figure 4.4c) to segment (Figure 4.4f) both the static shadow on the right and the moving shadow with the board are substantially removed. While PDIR improves the object segmentation by removing static shadows, LRBE can also substantially remove moving shadows.

The segmentation performance can be expressed numerically using a confusion matrix as in Tables 4.1 and 4.2 produced from the video sequence. The elements
Figure 4.4: Shadow Removal Video Frame - Bar Sequence. The top row shows the images created using NIII different techniques, and the bottom row shows the accuracy of the automatic segmentation of the images (green indicates correct, red indicates incorrect). The full video is available as listed in Appendix A.
Figure 4.5: Shadow Removal Video Frame - Funnel Sequence. The top row shows the images created using different NIII techniques, and the bottom row shows the accuracy of the automatic segmentation of the images (green indicates correct, red indicates incorrect). The full video is available as listed in Appendix A.
of the confusion matrices represent the percent of the image pixels in each of four possible segmentation cases below. The color listed is used to indicate the case in the segmented images and videos.

1. **Top Left - Dark Green - 0,0** - The pixel was segmented as 0, or part of the object by both manual and automatic segmentation.

2. **Top Right - Light Red - 0,1** - The pixel was manually segmented as a 1, background, but automatically segmented as 0, part of the object.

3. **Bottom Left - Dark Red - 1,0** - The pixel was automatically segmented as 1, object, but manually segmented as 0, background.

4. **Bottom Right - Light Green - 1,1** - The pixel was segmented as 1, or background, by both manual and automatically.

Cases 1,4 are correct, while cases 2,3 are incorrect. The accuracy is the sum of the 0,0 and 1,1, terms. The four terms sum to 100%, except for rounding errors.

LRBE shows an absolute improvement of 4.4% in the bar sequence and 1.5% in the funnel sequence over PDIR and raw images. While the absolute improvement between PDIR and LRBE for the Funnel Sequence (Tables 4.2b and 4.2c respectively) is only 1.5%, the area of the segmented object is only 2.7%, so the relative improvement is 55% of the object area. These experiments show LRBE improves the segmentation performance of a simple algorithm by identifying an object 55% more accurately on a scene where the object and shadows are easily confused.

### 4.3 Noise Analysis

A key question for implementing NIII is: How intense does the flash need to be for an application? This question is answered in three parts:

1. **What image quality does the application require?** This question is answered in Section 4.3.1. Segmentation is examined since it is sufficiently simple that the performance can be analytically related to image characteristics, specifically Signal-to-Noise Ratio \((SNR = \frac{signal}{noise})\). The SNR is found from the next two questions.
Table 4.1: Confusion Matrix for Bar Sequence

<table>
<thead>
<tr>
<th></th>
<th>Manual</th>
<th></th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Auto</td>
<td>0</td>
<td>6.0%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.0%</td>
<td>76.7%</td>
</tr>
</tbody>
</table>

(a) Raw Performance, 82.7% Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Auto</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

(c) LRBE Performance, 98.0% Accuracy

Table 4.2: Confusion Matrix for Funnel Sequence

<table>
<thead>
<tr>
<th></th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Auto</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

(a) Raw Performance, 80.4% Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Auto</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

(b) PDIR Performance, 97.8% Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Auto</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

(c) LRBE Performance, 99.3% Accuracy
2. **What is the image noise level?** Noise is a result of the camera imaging process (discussed in Section 4.3.2) and the NIII algorithm processing the camera images (discussed in Section 4.3.3).

3. **What is the image signal level?** The factors influencing the signal level are discussed in Chapter 7.

These questions are answered partially in this section, and partially in Chapter 7, culminating in a predicted operational range in Section 7.3.

### 4.3.1 Signal to Noise Ratio Required for Segmentation

The required image quality, measured as SNR, will depend on the application and the scene. SNR is first shown qualitatively, to provide a visual perception of the SNR. Next, the impact of SNR on segmentation performance is evaluated in an ideal case. Finally, the impact of SNR in a practical case is considered.

#### Qualitative SNR

Figure 4.6 shows synthetic example images with a range of signal to noise ratios to provide a qualitative understanding of SNR. The images were created starting with a base image of a light square on a dark background and adding Gaussian-distributed intensity noise to the image. The intensity difference between the light square and the dark background was 1, then random noise with a standard deviation equal to $1/SNR$ was added to the image.

#### Ideal SNR

The SNR can be used to find the segmentation error rate - how frequently pixels are incorrectly segmented. In a simple case, segmentation classifies each pixel between 1 (light background) and 0 (dark object) groups based on whether they are above or below a threshold set half-way (0.5) between the levels ($I_{thresh}$). Segmentation errors will occur when a pixel that should be 1, i.e. background, has a value below 0.5 due to noise, and is classified as 0, i.e. object; or the reverse. Given the standard deviation of the noise, the number of incorrectly segmented pixels can be calculated assuming a normal distribution.
Figure 4.6: Image Signal to Noise Ratio Examples. (n.b. Image compression algorithms usually remove some high frequency components of images assuming it is noise and will distort this image, reducing the apparent noise. Care has been taken to avoid this, but it will depend on the viewing or printing software.)

Figure 4.7 illustrates this analysis; it assumes an SNR of 3, the background mean intensity ($\mu_{\text{Background}}$) is 2 the object mean intensity ($\mu_{\text{Object}}$) is 1 and therefore the segmentation threshold ($I_{\text{Thresh}}$) is 1.5. A threshold set half way between the object and background intensities, is used since there are no assumptions about the scene. The threshold may be set differently if assumptions about the scene are added. With a signal intensity of 1, and an SNR of 3, the noise standard deviation ($\sigma$) is $1/3$. This assumes that the object being segmented and background are perfectly uniform. The background and object pixels have normally distributed noise added to them, creating the normal distributions shown. Some of the pixels have
so much noise they cross the segmentation threshold, as shown by the shaded areas, and are incorrectly classified. The red shaded area are background pixels that are incorrectly classified as object, while the blue are object pixels classified as background.

The probability of pixel measurements being over the segmentation threshold, causing the pixel to be misclassified, can be found from the cumulative distribution function ($CDF$) of a normal distribution. The error rate is given by

$$\text{Error Rate} = (1 - CDF(\mu_{Obj}, \sigma, I_{Thresh})) + CDF(\mu_{Background}, \sigma, I_{Thresh}) \quad (4.1)$$

The ideal-case segmentation error rate, as compared to SNR is shown in Figure 4.8.
Figure 4.8: Segmentation Error Rate v. SNR
Practical SNR

In practice, the scene will have some variance of its own, due to shading and variations in reflectance. The object and background noise needs to be added to the camera noise to find the effective required SNR of the system. Figure 4.10 shows the combination of noise from a real image (shown in Figure 4.9) in the object and background added to the camera noise. The additional noise due to the background and object reduces the margin for the camera noise, giving a realistic segmentation error rate v. SNR plot, in Figure 4.11. Figure 4.11 cannot be found analytically using 4.1 like Figure 4.8. Instead, the histograms obtained from 4.9 were numerically convolved with a normal distribution, as in Figure 4.10, and the probabilities beyond the threshold were summed then divided by the number of pixels. There is a small region on the object which is bright due to a specular highlight; it will always be incorrectly classified as background irrespective of SNR. This analysis is highly scene dependent.

4.3.2 Camera Noise

Next, the camera noise needs to be examined. The camera noise comes from several sources in the imaging process, discussed below. Irie [40] and others [63, 90] studied the noise in CCD cameras. (Camera noise terms are characterized by the standard deviation of the measured camera pixel intensity $I$, presented here with a
Figure 4.10: Segmentation SNR Calculation using a Non-Ideal Sample Image (shown in Figure 4.9). The Solid line is the image histogram; while the dotted line is the image histogram with a simulated SNR of 5, creating the smooth appearance. The signal intensity is the distance between the two peaks (red and blue dashed lines). The shaded areas show the pixels which are incorrectly categorized using the threshold. The raw image had camera noise, though it was substantially lower than the added noise for the SNR of 5 example.
Figure 4.11: Segmentation Error Rate v. SNR for Non-Ideal Case
normalized range) assuming a normal distribution.

**Noise Terms**

The noise terms can be grouped into two broad categories, spatially varying (different pixels measure the same signal differently) and temporally varying (one pixel makes different measurements of the same signal). *PRNU*, photon-response non-uniformity, and *FPN*, fixed pattern noise, are spatially varying, but constant over time for each pixel. The NIII techniques proposed are performed pixel-by-pixel, without using neighboring pixels, so the spatially varying noise terms are unaffected by NIII. Spatially varying noise can be measured for a specific camera, and usually can be removed from images in processing. Demosaicing (*N_D*) and filter noise (*N_filt*) are the products of signal processing, which does not take place when using raw data from a monochrome camera.

The shot noise, dark current, read noise and quantization noise terms are significant to evaluating NIII. \( SN_{ph}(I) \) is photon shot noise due to the statistical uncertainty in the number of photons actually striking a given pixel. It is dependent on the intensity of the observed pixel and follows a Poisson distribution [34]. \( SN_{dark} \) is dark current shot noise caused by random collection of stray electrons, unrelated to photons, even while the CCD is not exposed. Dark current is temperature dependent, but this dependence was neglected since the camera was operated in a narrow temperature range. \( N_{read} \) is readout error due to electronic noise in the camera electronics. \( N_Q \) is quantization noise from converting continuous values to discrete values; it can be shown analytically to be \( Q^2/12 \) where \( Q \) is the smallest quantization step. The Grasshopper camera used has a 14-bit A/D converter, so the quantization step is \( 1/2^{14} \approx 6.1 \times 10^{-5} \), giving a fixed quantization noise of \( 3.1 \times 10^{-10} \). Of these terms, photon shot noise is usually the most significant.

**Simplified Temporal Noise Model**

NIII techniques can be evaluated using a simplified, three-term, intensity-dependent empirical temporal camera noise model, demonstrated in [40].

\[
N_\Sigma = c_o + c_1I + c_2\sqrt{I} \tag{4.2}
\]

69
where \( N \) is the total noise, \( c_0, c_1, c_2 \) are coefficients found for a given gain, temperature and exposure setting of the camera and \( I \) is the measured pixel intensity.

**Camera Noise Measurement**

The camera noise was found by taking a sequence of images of a static scene. The measured value for each pixel should be constant throughout the sequence, so the difference, measured by the standard deviation, between the same pixel in different images is due to the temporal camera noise. The standard deviation was averaged for pixels with similar mean intensities to get the average noise at that intensity. The average standard deviation at each intensity was plotted and fit to find the coefficients. Figure 4.12 shows the data and the regression fit for the Grasshopper camera used at 0dB gain. The coefficients found were \( c_0 = 3.01 \times 10^{-4}, c_1 = 5.80 \times 10^{-4}, c_2 = 5.82 \times 10^{-3} \). Analyzing a sequence of images of a static scene provided the coefficients for this simple camera noise model.

**4.3.3 NIII Technique Image Noise**

The proposed NIII techniques use input from up to six images. Each of these images contributes to the noise of the NIII flash-only image. The camera noise from the six individual images needs to be propagated through the NIII techniques to find the output flash-only image (\( \hat{A}''_F \)) noise. It is possible to analytically propagate a noise distribution through linear operations, but it becomes more difficult with a saturation limited range. CRBE techniques saturate \( \rho \), part way through the algorithm, limiting the utility of analytical methods. Monte Carlo analysis [7] can propagate noise distributions through non-linear systems by running a large number of random trials, where the distributions of the random inputs follows the expected distribution of real inputs. The resulting output distribution will be more representative as the number of iterations increases. Monte Carlo analysis was used to compare the impact of different NIII techniques on the camera noise.

The expected NIII flash-only image noise will depend on both the intensity and the noise of each of the six input images (\( A', A'', A''' \), \( B', B'', B''' \)), creating a twelve dimensional input, which is too complex to present meaningfully. For each image, the noise can be predicted from the intensity as modeled in Section 4.3.2, reduc-
Figure 4.12: Camera Noise Measurements and Fit for Point Grey Grasshopper used at 0dB Gain.
ing the input to six image intensities. The analysis can be further constrained by assuming the scene is unchanged between the three times. While this will not be true in a moving scene, some regions will get darker while others will get lighter, so assuming the average image intensity is similar between image times isn’t unreasonable. With this assumption the first and last $\lambda_A$ images should be the same ($A' = A'''$), the $\lambda_B$ images should be the same ($B' = B'' = B'''$) and the $t_2 \lambda_A$ image should be brighter than the other times due to the flash ($A'' > A'$). This reduces the analysis to three dimensions, a $\lambda_A$ intensity, $\lambda_B$ intensity and a flash intensity.

Using the model created in Section 4.3.2 and simple assumptions, a test case can be created with only an ambient intensity and a flash intensity.

A test case with $\lambda_A$ and $\lambda_B$ ambient intensities of 0.5 and a flash intensity (signal) of 0.1 ($B' = B'' = B''' = A' = A''' = 0.5, A'' = 0.6$) was evaluated each NIII technique using Monte Carlo analysis, and the results shown in Table 4.3. The second column lists the absolute noise in the output flash-only image. The third column is the ratio between the output noise of the technique and the input noise of a single image (with 0.5 intensity), giving the factor by which the technique increases camera noise. The PDIR noise can be calculated analytically since the noise from subtracting two values. ADIR and LDIR noise is actually lower than PDIR since LDIR and ADIR average two time values. Averaging two values doubles the signal by adding the two values while increasing the noise by $\sqrt{2}$, then dividing by the two values gives the same signal strength as the original but with $\sqrt{2}/2$ of the noise. All the techniques can produce results at 1/2 the input frame rate by using the $t_3$ image from one sequence as the $t_1$ image for the next sequence. The Monte Carlo analysis only addresses the camera noise, not the effectiveness of the technique at removing ambient illumination.

Figure 4.13 shows the variation in output camera noise from LRBE at given background and flash intensities. The variation is principally due to the higher input noise at greater intensities. The surface is not smooth since Monte Carlo analysis will give slightly different results with each run, converging on a value as the number of iterations approaches infinity.
Conclusion

NIII techniques substantially improve the visibility of the flash (signal) by reducing the ambient illumination (as shown in Chapters 3 and 5) but at a cost of increased camera noise. In many scenes the reduction in ambient intensity will outweigh the additional camera noise added by using NIII techniques. The NIII technique with the lowest noise is not necessarily the optimal choice, since it may not remove the ambient light as effectively. The choice of NIII technique will depend on the cameras and scene.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Absolute Noise</th>
<th>Technique Noise/Single Image Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDIR</td>
<td>$4.495 \times 10^{-4}$</td>
<td>$\sqrt{2}$</td>
</tr>
<tr>
<td>PRBE</td>
<td>$6.363 \times 10^{-4}$</td>
<td>2.0</td>
</tr>
<tr>
<td>ORBE</td>
<td>$6.363 \times 10^{-4}$</td>
<td>2.0</td>
</tr>
<tr>
<td>ADIR</td>
<td>$3.890 \times 10^{-4}$</td>
<td>1.3</td>
</tr>
<tr>
<td>ARBE</td>
<td>$5.498 \times 10^{-4}$</td>
<td>1.7</td>
</tr>
<tr>
<td>LDIR</td>
<td>$4.348 \times 10^{-4}$</td>
<td>1.4</td>
</tr>
<tr>
<td>LRBE</td>
<td>$5.178 \times 10^{-4}$</td>
<td>1.6</td>
</tr>
<tr>
<td>SDIR</td>
<td>$4.435 \times 10^{-4}$</td>
<td>1.4</td>
</tr>
<tr>
<td>SRBE</td>
<td>$5.304 \times 10^{-4}$</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table 4.3: Monte Carlo Noise Estimates for Different Techniques assuming ambient intensities of 0.5 and a flash intensity (signal) of 0.1 ($B' = B'' = B''' = A' = A'' = A''' = 0.5, A'' = 0.6$). PDIR is just a subtraction, so the analytical solution is $\sqrt{2}$. 

73
Figure 4.13: Output Noise of LRBE with scene intensity in $x$ and flash intensity in $y$ ($x = B' = B'' = B''' = A' = A'''$, $A'' = x + y$).
Chapter 5

Shadow Removal

Shadow removal in moving scenes is another application investigated for NIII, and could be applied to mining applications such as aggregate measurement (Section 1.4.2). Two evaluations were used for shadow removal: qualitative and shadow ratio. Qualitative evaluation was simply visual examination. Shadow Ratio evaluated the reconstruction of the background where the shadow appears, comparing the NIII image where an ambient shadow was removed to a known shadow-free background. The results of these two experimental methods are presented below.

5.1 Apparatus

Videos were created using the full dual-camera system, with a high powered LED as shown in Figure 5.1. The OSRAM green LED module was replaced with a more powerful Enfis Quattro-Mini air-cooled violet LED module (ENMQ VIO-LET LE KIT DOM, [20]) with 200W Enfis driver producing 13.6W at 405nm (violet) over 4cm² for shadow removal experiments. The Enfis module has a larger emission area, which is incompatible with the projector optics (discussed in 6.1.1) and hence couldn’t be used for the depth correlation investigation (in Chapter 6), but is considerably more powerful than the OSRAM green module. The $\lambda_A$ camera used a 0.5ms exposure with its narrow (10nm FWHM, 85% peak transmittance) bandwidth 405nm filter, while the $\lambda_B$ camera used a broad bandwidth dichroic filter allowing a faster 0.05ms exposure. The mounting was modified from apparatus originally designed for the projector and is discussed in depth in Section 6.1.2. The scene ambient illumination was full (mid-day November in Vancouver) sunlight. Objects were manually moved back and forth in the scene, approximately 2.5m from the stationary camera.
5.2 Qualitative Evaluation

NIII algorithm performance was demonstrated qualitatively using videos of scenes taken outdoors, as shown in Figure 5.2. Figure 5.3 shows images taken with an outdoor sequence set against a south facing wall with the flash at full power. Figure 5.3a shows the scene with only ambient light, just before the flash image is taken. Shadows are clearly visible, cast from a model tractor on a white board. Almost the same scene is shown, this time with the flash, in Figure 5.3b. Using PDIR, subtracting Figure 5.3a from Figure 5.3b, produces Figure 5.3c. Strong light and dark bands are visible in the PDIR image caused by the tractor motion. The final Figure, 5.3d, using LRBE, has removed most of the light and dark bands around the tractor, giving a far more accurate illumination invariant image. There is a small shadow to the left of the tractor in the NIII image due to the flash being
located to the right of the camera. The illumination invariant image was created with approximately 1/11000th the average continuous illumination power of the ambient sunlight. Figure 5.4 shows similar examples to Figure 5.3, but taken with a reduced flash power. In this case, the flash is 1/37000th the average intensity of the sunlight. These qualitative experiments show the practical potential of NIII for improving images for machine vision even in challenging conditions.

5.3 Shadow Ratio

NIII shadow removal was quantitatively evaluated by measuring the accuracy of the shadow-free region using shadow ratio ($\kappa$). The shadow ratio is the difference between the NIII created flash-only image and a shadow-free average image created without the objects (and thus their shadows) in the scene, divided by the intensity
Figure 5.3: Tractor Example of Outdoor Performance with Full Power Flash. There is a shadow in the NIII images due to the LED light separation from the camera, but the ambient shadow is substantially removed. The full video is available in Appendix A. The contrast and brightness is the same for all images and has been adjusted for clarity.
Figure 5.4: Tractor Example of Outdoor Performance with 20% Flash. The full video is available in Appendix A. The contrast for 5.4b and 5.4a is the same as Figure 5.3, the contrast is doubled for clarity in 5.4c and 5.4d. The brightness has been uniformly increased for clarity.
of the shadow-free average image. In order to exclude the object itself, the shadow region is only computed over a manually segmented region-of-interest (ROI). Figure 5.5 shows example images of the steps of shadow ratio analysis; the sub-figures are described in detail below. This section describes how the shadow-free average images and region of interest images are created, then how the shadow ratio is calculated both for NIII images and for ambient images for comparison. The shadow ratio evaluates the intensity of the remaining shadows, unlike quantitative shadow measures proposed by [76] which only measure the area accurately classified as shadow. The shadow ratio will vary with the scene, especially with changes in the size of the shadow region.

5.3.1 Shadow-Free Average Image Creation

In order to compute a shadow ratio, average shadow-free images approximating an ideal, object-free, and thus shadow-free, image were calculated. Shadow-free average images of both the ambient-only case (\(\bar{A}_A\), e.g. Figure 5.5a), and the ambient-and-flash case (\(\bar{A}_{AF}\)) were created by averaging together video frames taken of the scene without the object. Averaging the 50 images in the sequence reduced the noise by a factor of \(1/\sqrt{50}\). A flash-only average (\(\bar{A}_F\)) image was created by subtracting the ambient image from the ambient-and-flash average images as in

\[
\bar{A}_F = \bar{A}_{AF} - \bar{A}_A
\]

5.3.2 Region of Interest

The video sequences, e.g. Figure 5.5b, included the object itself, and its shadow. The object was manually segmented out, by creating a Region Of Interest (ROI) mask as in Figure 5.5c, so the analysis would evaluate only the shadows and not the object. The ROI was conservatively segmented to ensure that none of the differences are due to the object itself. The ROI mask image is referred to as \(S\), where a value of 1 (white in Figure 5.5c) represents a pixel that is part of the ROI and 0 (black in Figure 5.5c) represents a pixel which is part of the object and excluded from the analysis.
Figure 5.5: Example Images of Shadow Ratio Analysis. The d,e,f images are bidirectional and adjusted to the same contrast level, with a 0 value represented by 50% intensity gray.
5.3.3 Flash Shadow Ratio

The shadow ratio is the rms of the difference between the measured flash-only image and the average image over all pixels in the image (over the total image rows \( M \), and total columns \( N \), with indices \( m, n \) respectively), within the ROI (\( S \)) divided by the mean intensity of the average image within the ROI. The intensity of the shadow \( (\xi) \) is evaluated as the difference between the NIII flash-only image \( (\hat{A}''_F) \) and the shadow-free, flash-only reference image \( (\bar{A}_F) \) over the region of interest (where \( S_{m,n} = 1 \), as opposed to outside the ROI where \( S_{m,n} = 0 \)) as in

\[
\xi = \sqrt{\frac{\sum_{m\in1:M} \sum_{n\in1:N} |\hat{A}''_F(m,n) - \bar{A}_F(m,n)|^2 S_{m,n}}{\sum_{m\in1:M} \sum_{n\in1:N} S_{m,n}}} \quad (5.1)
\]

The intensity of the light \( (\zeta) \) is the intensity of the shadow-free, flash-only reference image \( (\bar{A}_F) \) over the ROI (\( S \)) as in

\[
\zeta = \frac{\sum_{m\in1:M} \sum_{n\in1:N} \bar{A}_F(m,n) S_{m,n}}{\sum_{m\in1:M} \sum_{n\in1:N} S_{m,n}} \quad (5.2)
\]

The shadow ratio \( (\kappa) \), the ratio of the shadow intensity to the light intensity is then given by

\[
\kappa = \frac{\xi}{\zeta}
\]

The minimum possible shadow ratio will be when \( \xi \) is due to only camera noise.

5.3.4 Ambient Shadow Ratio

The shadow ratio of the scene without using NIII (i.e. Raw) is also calculated for comparison. Since the images without NIII don’t include a flash, the ratio is based on only ambient illumination. To calculate the shadow intensity the raw ambient-only image \( (A'_A) \) and mean ambient illumination \( (\bar{A}_A) \) replaced the NIII estimate \( (\hat{A}''_F(m,n)) \) and mean flash illumination \( (\bar{A}_F) \) respectively in (5.1) for
\[ \xi = \sqrt{\frac{\sum_{m \in 1: M} \sum_{n \in 1: N} |A'_{A(m,n)} - \tilde{A}_{A(m,n)}|^2 S_{m,n}}{\sum_{m \in 1: M} \sum_{n \in 1: N} S_{m,n}}} \]  

(5.3)

and in (5.2) for

\[ \xi = \frac{\sum_{m \in 1: M} \sum_{n \in 1: N} \tilde{A}_{A(m,n)} S_{m,n}}{\sum_{m \in 1: M} \sum_{n \in 1: N} S_{m,n}} \]  

(5.4)

## 5.4 Results

The shadow ratios are presented in Table 5.1. The first column gives the apparent intensity of the flash as a proportion of the ambient light (in this experiment, sunlight). LRBE (e.g. Figure 5.5f) is effective at reducing shadows even at lower flash intensities, however higher flash intensities provide lower shadow ratios. PDIR (e.g. Figure 5.5e) performs worse than Raw images (e.g. Figure 5.5b) when the light levels are low and the PDIR subtraction introduces artifacts of greater magnitude than the shadows it removes. The shadow ratio analysis numerically demonstrates the performance improvement of up to 70% compared to raw or 65% compared to PDIR from using NIII for shadow removal.

<table>
<thead>
<tr>
<th>Flash Intensity Compared to Ambient</th>
<th>Shadow Ratio, ( \kappa )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw</td>
</tr>
<tr>
<td>13.6%</td>
<td>0.327</td>
</tr>
<tr>
<td>27.7%</td>
<td>0.362</td>
</tr>
<tr>
<td>35.6%</td>
<td>0.303</td>
</tr>
<tr>
<td>44.7%</td>
<td>0.327</td>
</tr>
</tbody>
</table>

Table 5.1: Shadow Ratio
Chapter 6

Depth Correlation

Depth Correlation is an active imaging technique for determining depth in an image, illustrated in Figure 6.1. Depth correlation is believed to be the technique behind the Microsoft Kinect [28] and has been studied in [27, 80, 14]. A pattern, the appearance of which varies with depth, is projected on a scene. The recorded images are then correlated with a previously generated set of known pattern images taken of the projection on a plane at different depths. The known pattern image with the highest correlation to a recorded image region identifies the depth of that region. This technique is computationally intensive, requiring the recorded image to be correlated with a set of known pattern images to find the best match. NIII techniques were applied to improve the visibility of the projected pattern in the image, allowing the image to be thresholded and converted to logical. Correlation with logical (1-bit) images is far less computationally demanding than with arithmetic (e.g. 8-bit) images, potentially allowing depth correlation techniques to be used in power sensitive mobile applications. Improved depth correlation could improve performance in surface mining applications such as tooth detection (Section 1.4.1) and collision warning (Section 1.4.3), among others.

In this chapter, depth correlation using NIII images was implemented and tested under different lighting conditions. A projector was modified to use an LED flash to project a pattern on a scene with the triggering and wavelength required for NIII (Section 6.1.1) and mounted with the dual-camera system (Section 6.1.2). A simple scene was created (Section 6.1.3), of a paddle swinging in front of a background, was used to create a scene with two depths (background and foreground) which could be measured by depth correlation. A depth correlation algorithm was implemented (Section 6.2). The experiments showed a significant improvement in depth correlation accuracy using NIII on a scene with bright illumination (Section 6.3). Finally, the computing operations required for arithmetic versus logical
6.1 Apparatus

The depth correlation experimental apparatus included the dual-camera (see Section 3.2), light source (see Section 4.1.1), trigger mechanism (see Section 4.1.2) and software (see Section 4.1.3) from earlier experiments. A projector, mount and scene (all discussed below) were constructed for the depth correlation experiments. Figure 6.2 diagrams the experimental setup, with the dual-camera, projector and lights focused on a simple scene. The assembled projector and camera apparatus is shown in Figure 6.3. On the left side is the projector with the modified electronics using the green LED module. Below it there is the LED driver and the Altera board used for triggering. The trigger unit simultaneously triggers the cameras, and every third trigger also triggers the flash as required. The flash could be used every second trigger, as discussed in Section 4.1.2. The dual-camera system, previously correlation is examined (Section 6.4).
explained in Section 3.2, is on the top right. The mounting holds the projector and dual-camera in fixed relative positions required for 3D depth acquisition. The software (presented in Section 4.1.3) drives all the devices and both records data for later analysis and performs real time analysis. This section outlines the design of the apparatus to investigate depth correlation using NIII.

![Diagram](image.png)

**Figure 6.2: Depth Correlation Setup**

### 6.1.1 Projector

For the depth correlation investigation a patterned light projector was built. In the cited papers a laser with diffractive optics, which alter the phase of parts of the laser beam creating a depth dependent speckle pattern, was used. The visible pattern created by diffraction changes with the distance between the source and the
object. This change allows the source and camera to be located close together, minimizing the image area where objects visibly occlude the projection from objects behind them. The projector setup was simplified for these experiments by using a standard projector with an LED source. The projector was angled with respect to the camera so the projected pattern appears to shift with the depth of the object. The disadvantage of this approach is objects may occlude the projected pattern from some regions within the camera view. This disadvantage was determined to be acceptable for these experiments. An existing commercial projector was modified to provide a patterned monochromatic flash.

Two types of projectors are commonly manufactured, LCD and micro-mirror. LCD based projectors use RGB colored switching elements. These do not allow the color of the projection to be easily changed as the color is integral to the switching element. Micro-mirror projectors (often known by their Texas Instruments trade name - Digital Light Processor, DLP) use an array of micro-mirrors to selectively reflect each pixel of light either to the outgoing projection lens or astray. By modu-
lating how much time each mirror directs the light out through the lens, each pixel intensity can be adjusted. Color is created in micro-mirror projectors using a color wheel which rotates at a multiple of the refresh rate. The color wheel has segments colored in each of red, green and blue (many projectors also have white, some use other colors). The image in each color is shown on the micro-mirror array as the corresponding color passes on the color wheel. For our experiments, micro-mirror projectors are more suitable as they allow the color control to be separated from the modulation of each pixel.

A commercial DLP-based projector, the Sharp XR-10, was modified for the apparatus; as shown in Figure 6.4. It was selected for the quality of available documentation [82] to assist in the modifications. The projector was modified to allow it to operate with an LED based light source. Figure 6.5 shows the modified projector and related equipment. The LED light source (green OSRAM, discussed in Section 4.1.1) was fitted in place of the incandescent light the projector normally uses. The mixing rod (part of the original projector) internally reflects the incoming light several times mixing incident rays from different parts of the input thus making the light pattern more homogeneous, so a filament or individual LED chips are not visible in the projection. The alignment mirror (also part of the original projector) was used to direct the light from the source onto the DLP array. The power supply control and monitoring lines for the high voltage light driver were run to an external micro-controller which provided fake responses to the projector CPU to indicate the light was running, even without the light. The color wheel was also removed and the external micro-controller provided fake feedback signals to the projector CPU to indicate the color wheel was working. The pattern the projector used was provided through the VGA interface from a computer. The modified projector, coupled with the OSRAM LED module and associated Mightex driver was able to project monochromatic patterns when triggered.

A pattern was selected to project on the scene based on the structured light pattern proposed in [94]. It is shown, as seen at the background depth, in Figure 6.6. The large scale checkerboard pattern makes phase shifts easily apparent for illustration. The projector is positioned to the right of the camera and rotated so the camera and projector fields-of-view overlap, causing the pattern to principally shift right to left with increasing depth. The right to left pattern variation helps separate
between cycles of the checkerboard pattern. Arbitrary patterns can be used with depth correlation, typically line or speckle patterns are used, the selected pattern helps clearly illustrate the process.

6.1.2 Mounting

The mount provided a fixed distance and orientation between the camera and projector so the projected pattern was consistent. A rendering of the mount is shown in Figure 6.7. It was constructed of T-slot stock (80/20 Inc.) to allow easy adjustment. The projector (c) and dual-camera (d) each sit on a shelf with an adjustable pitch angle about the pivots (e) on the top of the three front supports. The two shelves pitch differently since the projector lens axis is directed upwards with respect to its housing base (so it can be used sitting on a table), while the camera lens axis is parallel to its housing base. The two shelves are connected with the vertical connector (b), then only one of the shelves is connected to the vertical support post (a), allowing the two to be adjusted simultaneously without changing the calibra-
Figure 6.5: Projector Diagram. The control lines are shown in black and the green dashed lines show the light path.

6.1.3 Scene

A simple scene was used to clearly demonstrate depth correlation, shown in Figure 6.2. The dual-camera (a) and projector (b) and four 500W halogen work lights (c) are aimed at the scene. The scene is composed of a white wall in the background (d) and a swinging wooden pendulum (e). A dual-camera image of the scene with the work-lights, but no pattern projected, is shown in Figure 6.8.

The scene intentionally includes common challenges for machine vision applications. The position of the four work lights creates a pattern of shadows behind
the paddle as illustrated (f) in Figure 6.9, these shadows are visible in Figure 6.8. The shadows will change as the paddle moves, complicating scene interpretation. Figure 6.10 illustrates how the paddle occludes part of the projected pattern from the background within the view of the camera. The algorithm can’t detect any pattern in that region, and should mark the depth as unknown. The projector used had to be separated from the camera so the pattern would appear to shift with object depth, as viewed by the camera. If a diffractive optical element and a laser, which can project a pattern which changes diffraction pattern with depth, were used they could be located very close to the camera reducing the occluded area.

### 6.2 Image Processing

Depth correlation operates by comparing an image taken of the scene with a pattern projected on it with with a set of known pattern images of what the pattern looks like at a range of depths. The depth of the known pattern image which correlates most closely with the image identifies the depth.
A simple depth correlation algorithm was used for testing. The dual-camera system was used to capture image sequences, timed with the projected flash pattern. The data was processed with three techniques, using raw images from the $\lambda_A$ camera, using PDIR images and using LRBE images, then analyzed by the algorithm. The algorithm allows the NIII performance to be compared qualitatively in videos and quantitatively using the accuracy of the depth correlation.

**Known-Pattern Image Generation**

Known-pattern images (KPI\(^1\)) were created by manually segmenting a video of the paddle swinging in front of the background. All the manually segmented regions of the same depth were averaged together to create a single KPI for each depth. The average manually segmented image was then manually thresholded to get a logical pattern map at each depth. In these experiments the depth correlation was only between background and foreground (the paddle) depths. Figure 6.6, shows the known-pattern image at the background depth.

\(^1\)“Reference Image” would be a sensible alternative term to KPI, however it is not used to avoid confusion with the “reference image” referred to in Chapter 2
Correlation Image Generation

Each image obtained from the video was then processed to categorize the depth of each image region. Figure 6.11 shows examples of the processing steps of a raw image. NIII was used to obtain flash-free images using the PDIR and LRBE techniques presented in Chapter 2 and raw images were used for comparison. A raw image, taken with minimum ambient light, is shown in Figure 6.11a. First the images were thresholded using an automatic threshold to convert them to a logical (1-bit) image e.g. Figure 6.11b. The threshold was determined using Otsu’s method [71], similar to the segmentation experiments in Chapter 4. Otsu’s method was selected since it doesn’t require any coefficients, thus providing an unbiased threshold. The thresholded image was then compared to the KPIs generated earlier, to find the best depth match, as discussed below.
Depth Image Matching

The depth for each image region is found from the depth of the KPI which best matches the thresholded image. The image was separated into regions of 20-pixels x 20-pixels. Each image region was compared to the KPI region at each depth using an XNOR (not-exclusive-or) logical operation which resulted in a 1 at each pixel where the subject image matched the KPI, and a 0 where they differed. An example XNOR image, combining all the regions, is shown in Figure 6.11c. The average value of the XNOR image region was computed (equivalent to counting 1’s) for each KPI region for each depth. Figure 6.11d shows the regional averages for the XNOR image with the background depth KPI. The correlation was repeated for each depth. For each region, the KPI region which had the highest score (greatest number of 1’s) was determined to be the depth. If a region did not correlate
with any KPI, with greater than 60% of the pixels matching, the region depth was marked unknown (black). Repeating this process over all the image regions created the depth correlated regional map.

**Improved Algorithms**

The presented algorithm is simple in only working on two depths and with large, fixed regions, giving poor depth and spatial resolution. The algorithm can easily be extended to have greater depth resolution by increasing the number of KPIs the thresholded image is compared to. Increased depth resolution would not require changes in the acquisition, NIII technique or thresholding, only an increase in the number of KPIs the thresholded image is compared to. Increasing the depth resolution and thus the number of KPIs increases the computation requirements, making
the logical correlation used here more attractive than arithmetic correlation. The presented algorithm has poor spatial resolution due to fixed region boundaries that don’t necessarily align with object boundaries in the scene. Using a finer projected pattern would enable a reduced region size. Regions which did not match any KPI could be further subdivided and the sub-regions and compared to the KPIs for the depths of neighboring regions. Comparing and evaluating sub-regions could be performed efficiently using the logical images. The depth and spatial resolution of the processing algorithm for NIII depth correlation images can be improved.

6.3 Results

Two lighting cases are examined to compare the performance of depth correlation with NIII techniques under different conditions. The first case was used as a baseline. It was created with only background room lighting, so none of the lights (c) in Figure 6.2 were on. A sample image of the video is shown in Figure 6.12. The figures are explained in detail below. The entire videos are available as listed in Appendix A. Depth correlation was also tested with all four lights on, which created strong ambient shadows which confused the algorithm when used on raw images, shown in Figure 6.13. Table 6.1 numerically summarizes the performance, and is discussed in detail below. These cases demonstrated that NIII can significantly improve the performance of logical depth correlation.

The results are presented in three parts. First the layout and interpretation of the video frames is presented. Second, the numerical analysis is explained. Finally the results of applying the different techniques is discussed.

6.3.1 Videos

Sample frames from the videos for the weak ambient lighting and strong ambient lighting cases are shown in Figure 6.12 and Figure 6.13 respectively. The top row (a-c) are the images. The images are individually scaled to maximize the contrast. The bottom row (d-f) are the depth correlation results. The black areas were not sufficiently correlated to any KPI, which is expected in some regions where the projector pattern is occluded as shown in Figure 6.10. The dark gray areas represent
the background, and the light gray areas represent the foreground. The first column (a,d) is using a raw image, with no NIII processing. The second column (b,e) is using PDIR to create the flash-only image, and the third column (c,f) uses LRBE.

### 6.3.2 Categorization Accuracy

The videos are compared numerically (Table 6.1) using categorization accuracy. Categorization accuracy was found by counting how many regions depth measurement (excluding undetected regions) by the algorithm matched a manual reference segmentation. It is expressed as a percentage of the total regions in the video.

### 6.3.3 Technique Comparison

The lighting cases show the performance improvement due to NIII. In the weak ambient illumination (flash intensity was 43% of ambient intensity) case, used as a reference and shown in Figure 6.12, there is little variation between the raw, PDIR and LRBE images as expected since there is almost no background illumination to remove. In Table 6.1 the categorization accuracy is very similar between all three techniques with very weak ambient light. All three techniques show the flash clearly, and as a result the depth can be accurately categorized in all cases. In the strong ambient illumination case (flash intensity only 2% of ambient intensity), which NIII is intended for and is shown in Figure 6.13, the ambient illumination makes discerning the flash far more difficult. In the raw image (Figure 6.13a) the flash is too weak to separate effectively since the intensity difference between the shading, the foreground paddle and the background appears larger than the intensity difference due to the flash. As a result, the depth correlation obtained is poor, as shown in Figure 6.13d. Using PDIR (Figure 6.13b) there are large light and dark areas caused by the motion of the paddle which overwhelm the projected pattern, which results in poor depth correlation as seen in Figure 6.13e. Using LRBE (Figure 6.13c) the background illumination, and the effects of movement are substantially removed and the projected pattern can be clearly correlated as in Figure 6.13f. The numerical categorization accuracy presented in Table 6.1 when there is strong ambient light shows the performance with raw images is very poor, while PDIR is better, but LRBE performance is close to the performance with weak
ambient lighting.

<table>
<thead>
<tr>
<th>Case</th>
<th>Mean Light Intensities (Normalized Camera Measurement)</th>
<th>Categorization Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ambient</td>
<td>Flash</td>
</tr>
<tr>
<td>Weak-Ambient</td>
<td>0.0357</td>
<td>0.0155</td>
</tr>
<tr>
<td>Strong-Ambient</td>
<td>0.5893</td>
<td>0.0117</td>
</tr>
</tbody>
</table>

Table 6.1: Depth Categorization Accuracy. The results is the percentage of regions where the algorithms automatic segmentation matched the manual segmentation. The weak-ambient case is used as a reference, the strong-ambient case is what NIII was intended for.

6.4 Computational Feasibility of Higher Depth Resolution

A key drawback of depth correlation is the computation requirements, especially for power sensitive mobile applications. NIII produces a flash-only image that can easily be thresholded to a 1-bit logical image, substantially reducing the operations required for the correlation from arithmetic computation with multiplication operations to a simple logical XNOR operation. e.g. In order to resolve 256 depths (thus 256 KPIs) with a VGA image would require comparing $640 \times 480 \times 256 = 78,643,200$ pixels, and a similar number of additions to sum each region. Performing this at a typical video frame rate of 24 frames per second (fps) would require over 1.8 billion operations per second. While this is within the capability of modern computer CPUs and GPUs, it is too complex and power hungry for mobile applications. Using logical images, the XNOR could be performed in groups the size of the micro-controller bus, typically 32-bits. This would require only 59 million XNOR operations per second, which could be performed by a cell phone or other mobile processor. Logical image processing is well suited to hardware implementations, which would be very fast and energy efficient. Using NIII flash-only images and converting them to logical images can substantially reduce the processing requirements for depth correlation.
6.5 Summary

NIII can improve logical depth correlation performance in high-ambient light situations by substantially reducing the apparent intensity of the ambient illumination. Logical correlation is much less computationally intensive than arithmetic correlation. NIII could allow depth correlation to be implemented even on power sensitive mobile devices where the scenes contain significant ambient illumination.
Figure 6.11: Raw Image Processing Example

(a) Raw Image
(b) Thresholded Raw Image
(c) XNOR Image. White indicates the pixels in the thresholded image (b) and KPI for the background depth (shown in Figure 6.6) match.
(d) XNOR Regional Average Image. Lighter sections have a better correlation to the background KPI.
Figure 6.12: Weak Ambient-Light Scene Depth Detection Example. Raw, PDIR flash-only and LRBE flash-only images are shown on the top row, and are adjusted for maximum contrast. Depth matched images produced from each technique are shown on the bottom. Black indicates no depth found; Dark Grey indicates background regions; Light Grey indicates foreground regions. The full video is available in Appendix A.
Figure 6.13: Strong Ambient-Light Scene Depth Detection Example. Raw, PDIR flash-only and LRBE flash-only images are shown on the top row, and are adjusted for maximum contrast. Depth matched images produced from each technique are shown on the bottom. Black indicates no depth found; Dark Grey indicates background regions; Light Grey indicates foreground regions. The full video is available in Appendix A.
Chapter 7

Bounds on Ultimate Performance

The limitations of NIII techniques for specific applications are determined by several factors which are examined in this chapter. The wavelength selected for \( \lambda_A \) for the flash and camera could allow NIII to operate outside the visible range, or at a wavelength where ambient light is weaker. The selection of the wavelength also affects the safe operating distance, and the range at which the device will produce a usable image. Finally the computing operations required for different algorithms is examined.

7.1 Wavelength Selection

With suitable equipment, NIII could operate over a wide range of wavelengths, including UV and NIR. Visible light was selected for these experiments for safety and ease of use. Using invisible light is feasible with the appropriate (usually more expensive) equipment and proper safety precautions. The ambient light source, atmosphere, scene, filters, optics and camera need to be considered when selecting the wavelength. A practical consideration for experimentation is the ability to see the light, both to aid in setup and improve safety. In a working application, an invisible wavelength which doesn’t bother the operator is more attractive.

7.1.1 Ambient Light Source

NIII aims to remove the appearance of ambient light in an image. This is easier if the image is taken at a wavelength where the ambient source is weaker. Sunlight is the most common ambient light source, and a standardized sunlight spectrum is shown Figure 7.1 based on data from [3]. The actual intensity and spectrum of sunlight varies with the position of the sun in the sky and the atmospheric condi-
tions (e.g. clouds, smoke). Using wavelengths where sunlight is weaker, such as UV (<400nm), reduces the required flash power, but may be challenging for other components as discussed below.

Figure 7.1: ASTM-G178 Solar Irradiance

7.1.2 Atmosphere

Air is substantially transparent around the visible wavelength range. In specific IR wavelengths the oxygen and CO$_2$, among other compounds in the air, absorb light passing through it, creating a dip in the spectrum of a distant source. If there is substantially more atmosphere between the ambient source (e.g. sun) and the object as compared to between the flash and the object the absorption could be beneficial to reduce the ambient light in the scene without substantially weakening the flash. If the absorption is too high, the flash itself may not be visible.
7.1.3 Scene

Objects are visible by their reflections, of ambient or flash illumination. NIII will be ineffective on a totally black scene (e.g. coal or bitumen). Colored objects reflect more strongly in one wavelength than in others. Selecting a flash wavelength where objects of interest in the scene have high reflectance will improve performance.

7.1.4 Filters

Narrow-bandpass interference filters are available for a wide range of wavelengths. Typically they are manufactured to match common laser emission wavelengths.

7.1.5 Optics

Various optical materials are useful at different wavelengths. Most optical glass has good transmission from visible through near-infra red wavelengths. Ultraviolet is blocked by many types of glass, so special UV lenses made with fused silica [19] are used for UV applications. Most lenses intended for visible light use several types of glass to combat dispersion and chromatic aberration, but often have poor NIR performance. Dispersion occurs because the refractive index of materials changes with wavelength, causing chromatic aberration where the focal point is color dependent. Since the cameras in the dual-camera system are filtered to operate near a single wavelength, chromatic aberration in the lens is not an issue for NIII. Lenses are available for other wavelengths as in table 7.1, but are typically more expensive than readily available visible wavelength lenses.

Table 7.1: Selected Typical Optical Material Ranges, from data in [46]

<table>
<thead>
<tr>
<th>Material</th>
<th>Useful Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quartz</td>
<td>400nm – 4µm</td>
</tr>
<tr>
<td>Crown Glass</td>
<td>350nm – 2.5µm</td>
</tr>
<tr>
<td>Plastics</td>
<td>350nm – 2µm</td>
</tr>
<tr>
<td>CaF$_2$</td>
<td>100nm – 10µm</td>
</tr>
<tr>
<td>GaAs</td>
<td>850nm – 10µm</td>
</tr>
</tbody>
</table>
7.1.6 Camera

Camera sensitivity varies with wavelength depending on the bandgap of the semiconductor. Most cameras are silicon-based CMOS or CCD technology. Silicon based cameras are sensitive from 200nm-1000nm [46], which covers the visible range. Figure 7.2 shows typical silicon-based camera quantum efficiency. The vertical axis is quantum efficiency, a ratio of the signal received at each wavelength compared to the most sensitive wavelength. Other camera technologies are available with different sensitivity ranges, e.g. Table 7.2.

![Figure 7.2: Typical Si Camera Sensitivity based on Sony ICX 267 sensor used in the Grasshopper cameras in the Dual-Camera apparatus, based on [85].](image)

Table 7.2: Some Typical Optical Detector Sensitivity Ranges, from data in [46].

<table>
<thead>
<tr>
<th>Detector Material</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>200nm – 1μm</td>
</tr>
<tr>
<td>InGaAs</td>
<td>600nm – 2.5μm</td>
</tr>
<tr>
<td>CsTe</td>
<td>105nm – 375nm</td>
</tr>
<tr>
<td>Ge</td>
<td>800nm – 2μm</td>
</tr>
<tr>
<td>Ge:Cd</td>
<td>2μm – 30μm</td>
</tr>
</tbody>
</table>
7.1.7 Model

A Simulink model was created to simulate these factors. (Atmosphere and optics were ignored since they are both substantially transparent over the visible light range examined.) An excerpt from the model is shown in Figure 7.3. The full model considers four cases, $\lambda_A$ filtered image of only sunlight ($A', A''$), $\lambda_B$ filtered image of only sunlight ($B', B''$), $\lambda_A$ filtered image of sunlight and flash ($A''$) and $\lambda_B$ filtered image of only sunlight ($B''$).

The imaging model proceeds starting from the top left of Figure 7.3. The model inputs are an array of values by wavelength. First the light sources, sunlight and the LED, are added together. The intensity of sunlight doesn’t change substantially with distance on this scale since the sun is already some distance away. The flash intensity does change with distance, which is modeled using the field of view of the Enfis LED. The incident light is then multiplied by the reflectivity spectrum of the object (in this case a 20% reflectivity target which is constant over visible wavelengths) to get the spectrum of the reflected light. As the camera and object get further apart, the light intensity from a given surface area decreases by the inverse square law, but conversely the surface area imaged by each pixel increases by the square of the distance. These effects cancel out, hence objects do not get dimmer as they get further away from the camera, except due to atmospheric attenuation (discussed in Section 2.3). An atmosphere term could be added between the top and bottom rows for evaluating NIR or UV wavelength operation, but it isn’t significant in the visible spectrum. The top row models light to the point it enters the camera.

Next (bottom row of Figure 7.3), the light that has been reflected off the object and travels through the air to the camera assembly. The incident light is then multiplied by the filter spectrum based on the manufacturer specifications to get the filtered light spectrum. A camera optics spectrum could be included after the filter if it varied substantially over the evaluation range, which wasn’t the case here. Camera detectors are not uniformly responsive, as discussed in Section 7.1.6 and shown in Figure 7.2. The filtered light is multiplied by the camera sensitivity spectrum from the datasheet, and the result is summed together across wavelengths to get the quantity of electrons collected by the camera pixel per second. The number of electrons is multiplied by a conversion factor to convert it to a digital pixel value.
per second. The pixel value per second of exposure duration is multiplied by the exposure time to get the output pixel value. Physically the exposure control usually takes place on the camera sensor, but since it is a single factor for all wavelengths it is included after summation to reduce computation. The camera readout has a finite range, so the pixel value is limited by saturation to that range. Then finally the pixel value is output.

The camera noise is proportional to the light intensity (discussed in depth in sec. 4.3.2) and is calculated from the pixel value to get an expected noise level for the simulated configuration. The difference between the modeled camera values for the ambient-only and ambient+flash images gives the signal level. These can be combined to calculate the expected SNR. Using this model the operational range and SNR was calculated for different proposed configurations as in Section 7.3.

7.2 Optical Safety

A principal concern in any application is safety. The international standard IEC 60825-1 [39] specifies the Maximum Permissible Exposure (MPE) for laser radiation. Standard IEC 60825-9 [38] is for incoherent LED sources, but defers substantially to -1. The standard describes the three types of hazards from optical radiation and specifies limits for each of them. The lower of the three limits sets the maximum exposure.

7.2.1 Optical Radiation Hazards

There are three optical radiation hazards that need to be considered, skin exposure, retinal thermal hazard and photochemical hazard. Skin exposure permissible exposure limits, of about twice the irradiance of a sunny day, are far higher than retinal thermal or photochemical hazards in the visible range. Skin exposure would only need to be considered if glasses were mandatory or working in deep infrared, mitigating the retinal thermal and photochemical hazards. Retinal thermal hazard is from light focused in the eye being powerful enough to heat up, and effectively cook, the retinal tissue at the focal point. The cornea only focuses light between about 400nm to 1400nm wavelength, so light outside this range has a low risk of
Figure 7.3: Simulink Model of System by Wavelength
retinal damage. Photochemical hazard, sometimes referred to as blue light hazard, increases with shorter wavelengths, and correspondingly higher photon energies, starting in the visible range. Photochemical hazard comes from molecules being excited by photons, changing their behavior and properties, especially in the cornea. It can occur even without focusing and can be cumulative over a long period. Within the visible range the maximum expected exposure time for retinal thermal and photochemical hazards is low, since a strong visible stimulus triggers blinking or looking away. Someone could unwittingly stare at an invisible light source, requiring an invisible source to be evaluated for longer potential exposure. The retinal thermal and, for shorter wavelengths, the photochemical hazard are the key concerns.

7.2.2 Evaluation of Optical Hazards

NIII flash illumination is intended to operate over a large area and thus the light is spread out from the source, unlike laser light. The intensity of the light reduces with distance according to the inverse square law, thus there is always some distance where the light will fall below the MPE. The safety evaluation for NIII flashes focuses on how far from the source a user needs to be to ensure they are below the MPE. In practice, safety guards could be put in place for small distances. Applications like mining shovels could use more powerful flashes with larger unsafe distances by mounting the flash in areas that workers can’t access without special equipment and wouldn’t be accessible when the machine is in operation.

The distance to the MPE for a set of test conditions with a range of source intensities was evaluated using the IEC 60825 standard, shown in Figure 7.4. The source power is plotted on the X-axis. There are two distances (Y-axis) to MPE, shown with the green and blue lines. The green line assumes the source is always an infinitely small point, which would focus to a single spot (as small as possible within the limits of diffraction and lens aberrations.) on the retina. The blue line accounts for the size of the source listed. As the power increases and the viewer goes further away the apparent angular size of the source decreases until it reaches the theoretical minimum where the blue and green lines meet. The red line shows the distance to a target intensity of $1 \text{Wm}^{-2}$ in this example. The operational range
(red line) is independent of the MPE range. If an application requires a very intense flash, the flash may never be bright enough and eye safe. In this example case a 1W source, 5cm wide would be safe at any distance, while an infinitely small version of the same source would be safe at 0.6m and either would have sufficient intensity for RBE up to 2.3m.

![Distance to MPE by Source Intensity according to IEC 60825 for a 470nm wavelength, 2ms flash, pulsed at 10Hz with a 25° field-of-view.](image)

Figure 7.4: Distance to MPE by Source Intensity according to IEC 60825 for a 470nm wavelength, 2ms flash, pulsed at 10Hz with a 25° field-of-view.

### 7.3 Operational Range

The operational distance of NIII can be calculated from the target SNR, the camera noise model, the spectra of the light sources and filters; using the model discussed in Section 7.1.7. The object reflectance does not contribute significantly since it affects the measured intensity of both the flash and ambient illumination the same
way. Object reflectance will change the intensity of the measured signal, thus slightly increasing the noise; though the measured intensity can also be adjusted by changing the exposure. Figure 7.5 shows the SNR as compared to range for the configuration as follows:

- Enfis 405nm Violet LED module, with 25° Field-of-view
- $\lambda_A$ camera using 405nm interference filter with 3ms exposure
- $\lambda_B$ camera using dichroic red filter with 0.3ms exposure
- Sunlit scene, with 20% reflectance surface

For example, a 1% segmentation error rate would require an SNR of 5 according to Figure 4.8, which could be obtained at approximately 3m according to Figure 7.5.

### 7.4 Computational Speed

One of the principal trade offs in using these techniques is the computation speed of the algorithm. To test the algorithms speed they were implemented in C using Apple vector libraries and timed on an Intel Core 2 Duo at 2.4GHz. Figure 7.6 shows the time each technique took in comparison to the residual. Lower computational time and lower residual (bottom left corner) are more desirable, though the balance between the two will depend on the needs of the application. RBE consistently provides lower residual and thus better performance than DIR for the same weighting technique, though RBE techniques are slower than DIR techniques. All the techniques could be run in real time (<33ms, allowing greater than 30 frames per second) on the test system.

The algorithms are highly parallelizable so the performance is expected to be improved with more processors or on a graphics processing unit with many subprocessors. Initial experiments using the 32 stream processors on an Nvidia 8600M GT graphics card approximately doubled the processing speed, though this does not include the additional overhead of transferring the data to the graphics memory. If the image processing required after RBE is also highly parallelizable and can be efficiently implemented using a GPU, it may justify the required additional time to transfer the data to the GPU memory.
The performance improvement from using past-only (P) or averaging (A) to linear (L) or sigmoid (S) weighted is substantial. However, there is only a small performance improvement from linear (L) weighting to sigmoid (S) weighting and is sensitive to the selection of $\alpha$. Sigmoid weighting requires more computation and provides a slight benefit. Linear weighting is a better choice for most applications.

### 7.5 Image Processing Algorithms

NIII is intended to reduce background illumination and is most applicable to algorithms and scenes which are adversely affected by background illumination. Three
challenges caused by shadows were discussed in Chapter 1, and illustrated in Figure 1.2: false edges, discontinuities and lowered contrast. The prevalence of false edges and discontinuities will depend on the shadows cast on the scene. In a scene with complicated shadows (e.g. shadows from the sun shining through leaves) false edges and discontinuities are prevalent and NIII will offer substantial improvements. These two problems will affect algorithms which operate on small areas of an image (e.g. filtering with a 2D finite impulse response filter) in the vicinity of the shadow edges. The lowered contrast from a shaded scene feature compared to a brightly lit scene feature will reduce the intensities of differences between neighboring pixels, thus reducing the response of algorithms which rely...
on neighboring pixel differences (e.g. Sobel filter [29]). A pattern projected on
a scene using a NIII flash will not change contrast substantially between ambi-
tently lit and ambiently shaded areas, provided the camera doesn’t saturate. In the
case of a projected pattern, NIII image subtraction is principally useful to reduce
false edges and discontinuities, but won’t change the contrast substantially. Neigh-
borhood difference-based operations are often invariant to a constant background
offset, for example due to ambient light or shadow. NIII improves algorithm per-
formance on scene features throughout an image, though in the case of a projected
pattern, NIII is principally useful around the shadow edges.
Chapter 8

Conclusions and Future Work

Illumination is both indispensable and can be frustrating for machine vision. Without light, there are no photons to create an image. With most lights there are shadows, reflections and specular highlights confounding analysis. Using NIII there is something to see, which is consistent and potentially shadow free, irrespective of whether the ambient conditions are a bright sunny day or a dark night. NIII simplifies many applications by providing a simple, uniform image.

Natural Illumination Invariant Imaging can be used for shadow removal or segmentation for object recognition. It can be used for distance segmentation or depth-keying. It could provide an image which is easier to interpret by many machine vision algorithms.

Ten techniques for NIII were examined using a flash and a dual camera system. The accuracy of the estimated reference of a flash-free image (how close an estimate of a flash-free image was to a real flash-free image) was evaluated using video clips. The best method, SRBE, showed 63% less error than the baseline PDIR method. Segmentation using NIII methods demonstrated segmenting an object 55% more accurately compared to PDIR. Shadow removal was demonstrated using simple outdoor scenes, and was able to remove 65% of the shadow intensity compared to PDIR, or 70% compared to raw. Depth correlation using a simple logical algorithm was improved by over 60% in absolute terms using NIII images, or 45% in absolute terms compared to PDIR. NIII is a building block which can create images which improve the performance of many machine vision applications.

8.1 Principal Contributions

The principal contributions of this thesis are:
• Examining the impact of reference image subtraction on a variety of image processing applications. The impact of an existing NIII technique (PDIR) was quantitatively evaluated on three applications, segmentation, shadow removal and depth correlation.

• Mitigating the movement blurring artifacts in reference subtraction (PDIR) using two cameras, one able to see the flash (λA), the other unable to see the flash (λB) using NIII techniques. The second camera provides additional information allowing a more accurate reference image to be created. Previous techniques have either simply subtracted the past image (referred to as PDIR in this thesis) or required three cameras. The design considerations and trade offs of the cameras and flash were also explored.

• Proposing and comparing eight alternative NIII techniques (PRBE, ORBE, ARBE, LDIR, LRBE, SDIR, SRBE, TSEL) which take advantage of the additional information from the second camera. Two other methods (PDIR and ADIR) which only require a single camera were used as references. All the methods were compared numerically on a range of data. The relative computational demand of the techniques was considered. All the NIII techniques were demonstrated on 640x480 images in less than the 33ms required for 30frames per second.

• Demonstrating the most relevant techniques on several real applications and quantitatively assessing their performance. Shadow removal, segmentation and depth correlation applications were all used to test the best performing techniques. These applications showed substantial performance improvement. These NIII techniques could improve the performance of a wide range of machine vision applications. In open pit mining, they could potentially reduce the hazards from missing teeth and collisions by improving segmentation and reduce the explosives consumption by improving aggregate particle size measurement.

• Evaluating the bounds of NIII techniques and application performance. An imaging model was used to assess the noise intensity in the input images.
Statistical methods were then used to propagate the noise in the source images through to the output images to find the noise expected with each NIII technique. Finally, the impact of image noise on segmentation was examined, providing a complete analysis from scene and equipment parameters to application performance. The optical safety limits and the computing operations required for NIII techniques were examined.

8.2 Potential Future Improvements

There are several ways to improve the proposed NIII techniques using better hardware. With further development, budget and time custom components could be developed to improve NIII performance and potentially operate with a smaller, cheaper system. The feasibility of these improvements is reviewed below.

8.2.1 Alternate Wavelengths

Wavelength selection is a major factor in NIII performance. The current demonstration of NIII operated using violet (405nm) for the outdoor shadow removal and distance segmentation experiments and green (520nm) for the depth correlation testing. These wavelengths were selected in part due to available parts (e.g. cameras at desired frequency band, filters), and to improve safety by staying within the visible range so the operator would be aware of any exposure. As discussed in Section 7.1, there are dips in the solar radiation spectrum which would reduce the ambient light intensity, allowing a weaker light source to be used. Wavelengths outside the visible spectrum are attractive since they do not disturb users. Laser modules are available in the 1400nm wavelength range (frequently used for communications) where sunlight is highly attenuated, though they require a special camera. Another potential wavelength is 1064nm, since extremely powerful Nd-YAG industrial lasers [73] are readily available. The disadvantages of using 1064nm are that sunlight is not heavily attenuated and silicon based cameras are insensitive, so InGaAs or other cameras are needed. Near Infrared, 700nm-900nm has additional solar attenuation, reasonable silicon sensor sensitivity and readily available LED modules. High powered NIR lasers are available, but are more expensive and less
common than the 1064nm Nd-YAG lasers. UV wavelengths have very low sunlight, but present a greater optical hazard, and require highly specialized optics and cameras. NIII could operate at a range of wavelengths, the choice of wavelength will depend on the application requirements.

8.2.2 Higher Intensity Light Source

The higher intensity the light source, the greater distance NIII can be used at and the shorter the exposure time that can be used. There are two relevant aspects to the intensity - the bandwidth of the light, and the output power.

The narrower the bandwidth of the light source, the narrower the filter that can be used, and thus more ambient light is excluded and more of the flash light is passed by the filter. Few filters are available with bandwidths below 10nm, so this sets a lower bound on how narrow the light source bandwidth needs to be. Lasers usually have very narrow bandwidths (<1nm). LEDs usually have a bandwidth over 10nm, though shorter wavelength LEDs typically have narrower bandwidths. Using a laser for the light source is preferable for a narrower bandwidth.

A higher output power light allows a shorter exposure and a higher SNR or increased operating range. Most LEDs and their drive circuitry, including the ones used for this work, are designed for continuous use. NIII techniques only require brief flashes. Most LEDs can tolerate brief peak intensities, typically double their continuous operating power, depending on duration and manufacture (e.g. the OS-RAM module used [70], has a continuous rating of 1000mA forward current per chip, but will tolerate a 2000mA surge for < 10µS). A key limitation on LEDs and their drive circuitry is the heat dissipation; with a low duty cycle, short duration flash, the heat sink requirements are much lower than continuous use. More compact, intermittent use only, drivers and LEDs could be used. LEDs can be easily arrayed into larger modules, with a higher total power for shadow removal and distance segmentation applications. Large-area LED modules are less effective for depth correlation, since the light is emitted over a wider area and more complicated to focus into a pattern. Using existing LED modules with specially designed heat sinks and drive circuitry could create LEDs for NIII that are both more compact and more powerful than off the shelf products used in this work. Lasers are an-
other option, they are available with very high pulse power, but they are usually much more expensive and larger. DOEs (discussed below) require coherent light and only work with lasers. There are several options to increase the output power including pulsing LEDs and lasers.

8.2.3 Diffractive Optical Elements

Diffractive optical elements (DOE) [69, 44] use the wave properties of light to efficiently distribute a single coherent, monochromatic source, such as a laser, into a specific pattern. For shadow removal and distance segmentation applications they could be used to create a uniform light field matching the field of view of the camera. This would reduce the variation in lighting due to variations in the source. For structured light or depth correlation applications they can efficiently project a pattern. DOEs coupled with a laser source could yield an efficient and compact flash or pattern projector.

8.2.4 Shorter Time Between Images

The fundamental need for an estimation algorithm, as opposed to simply subtracting the previous image (PDIR), is due to the change in the scene between the images. The less time between images, the less motion in the scene needs to be compensated for using estimation. There are several limitations to infinitely close together images. Most limitations in typical camera technology can be addressed with special cameras. The limitations of dealing with low light levels are more fundamental. The need for estimation can be reduced, but not completely eliminated.

Standard cameras have a number of limitations however, increasing the time between exposures, or specialized cameras can address many of them. Typical cameras are designed to provide continuous 24 fps video, with each frame evenly spaced. The camera requires time between the frames to read the camera sensor and send the information to the computer. The camera sensor is usually read in serial, with each pixel sequentially read by a single ADC. High speed and large format cameras can speed this up by adding two or more ADCs, so only a portion of the pixels are read by each ADC, reducing the time required to read all the pixels. Sending information to the computer is limited by the speed of the link to
the computer. Some cameras have on-board processors, though since NIII requires
two cameras, this is of limited use. Many cameras have on board memory which
can save several images, so the transmission time to the computer doesn’t hold up
capturing the next image. Specially designed cameras are available to take images
in rapid succession, but the underlying physical limitation of exposure time still
limits the time separation.

The exposure time fundamentally limits the separation between exposures. The
narrow bandwidth of the filters reduces the incident light on the sensor, requiring
a longer exposure time. As the intensity falls, the noise does not decrease linearly
(see Section 4.3.2). The requirement of a minimum SNR with a narrow wave-
length filter sets a minimum exposure time. For example the outdoor experiments
presented in Section 5.3 were done with 5ms exposures on the $\lambda_A$ camera. The
scene will likely change within the exposure time causing blurring. Larger lenses
can improve the low light performance by increasing the amount of light collected
by the lens. Intensified cameras can be sensitive down to single photons, but are
still effected by shot noise, have additional noise sources and are expensive. With
higher intensity sources, less exposure time may be needed, allowing exposures to
be closer together in time. Fundamentally the subsequent exposures cannot occur
simultaneously, setting a minimum separation time between exposures, though that
time can be minimized using suitable hardware.

8.2.5 Smaller Separation Between Wavelengths

Estimation is required since we need a flash-free image, apparently taken while the
flash is on. If $\lambda_A$ and $\lambda_B$ were infinitely close together, there would be no need
for estimation as they would match except for the flash only being visible in one.
The bandwidth of the light source limits how close $\lambda_A$ and $\lambda_B$ can be to each other.
The $\lambda_A$ filter must include as much of the flash as possible while the $\lambda_B$ filter must
exclude the flash. As discussed above, laser sources have a very narrow bandwidth
allowing the filters to be closer together. Laser wavelength can change slightly with
temperature, so additional wavelength separation may be needed. The limitation
is then the bandwidth and availability of filters, specifically ensuring the $\lambda_B$ filter
does not pass any of the flash illumination. Off-the-shelf filters are available with
10nm FWHM bandwidth, but only at specific wavelengths. Custom filters could be built at specific wavelengths for NIII so the separation between the wavelengths is as low as 15nm. The differences between the $\lambda_A$ and $\lambda_B$ images will depend on whether the spectrum of the ambient light source or reflectivity of the objects change substantially between the adjacent wavelengths. If the differences between them are small enough, the $\lambda_B$ image could simply be subtracted from the $\lambda_A$ image taken with flash to leave a flash only image. In practice the absolute transmission of the filters will likely be different, so a single linear correction coefficient may be needed before subtraction. Careful selection of filters could improve the estimation accuracy, even reduce it to a single linear coefficient.

### 8.2.6 Miniaturization

The entire apparatus could be miniaturized considerably, creating a single compact module. The flash or projector could be miniaturized using a DOE as discussed in Section 8.2.3. The dual-camera apparatus could easily be reduced simply using smaller cameras. A key miniaturization would be splitting the image after a single lens, as opposed to in front of two lenses as in the current dual-camera system as in Figure 8.1. Off-the-shelf lenses are designed for a specific back-focal length (the distance between the lens and the camera sensor). The short back-focal length of most lenses doesn’t allow the relatively thick interference filters and splitter to be inserted into the optical path behind the lens. A custom designed lens with a longer back focal length would allow the splitter and filters to be inserted between the lens and the sensors, requiring only one lens and reducing the size of the apparatus. Designing a lens with a longer back focal length is relatively straightforward, though it is a limited market so they are uncommon. Some high end video cameras (e.g. Sony HDR-FX1 [84]) already use a splitter behind the lens to separate the image to independent red, green and blue sensors. Bandpass interference filters reflect a substantial portion of the out-of-band light, allowing the $\lambda_A$ filter to potentially also serve as a mirror, as in Figure 8.1c. The transmission wavelength of the filter is dependent on the incident angle of the light, so the filter wavelength needs to be selected accounting for the 45° mounting angle. A second filter would likely be required over the $\lambda_B$ sensor to ensure all the flash illumination was excluded and
to reduce the incident light on the $\lambda_B$ sensor to allow a similar exposure time to the narrow bandpass filtered $\lambda_A$ sensor. Many of the components could be easily miniaturized with engineering, but miniaturization is fundamentally limited by the lower light gathering of smaller lenses and higher noise of smaller sensors.

8.2.7 Single Multi-Wavelength Camera

Further miniaturization would be possible if a single sensor could be used to detect both wavelengths. Standard color image sensors do this with tiny filters in front of individual pixels and alternating the red, green and blue color filters, as shown in Figure 8.2a. A camera for NIII could be implemented using alternating filters over each pixel as shown in Figure 8.2b. The key challenge for NIII is manufacturing the very narrow bandwidth filters required right on each pixel. The filters used in color cameras are very broad bandwidth and don’t use interference, so they have a simpler manufacturing process than the required narrow wavelength interference filters. Implementing NIII on a single sensor would require considerable development.

8.2.8 Color Natural Illumination Invariant Imaging

While monochrome images, such as discussed in this thesis, are suitable for many applications, in some cases color images are useful for both human and machine understanding. Two techniques could be used to extend NIII to create color images. A hardware approach is to add additional cameras at different wavelengths. The flash would need to be active at any wavelength where images are produced. One solution would be to use an RGB camera, a white flash with a filter blocking NIR and a fourth camera (or pixels on a multi-wavelength camera as in Section 8.2.7) to capture an NIR reference image. The drawbacks of this approach are it requires a very powerful flash and complex camera apparatus. A flash designed to operate only in three narrow wavelengths corresponding to the RGB camera colors would reduce the flash power required, but would produce images that don’t appear as natural. Another approach is using a standard RGB color camera in addition to the proposed NIII apparatus. The NIII apparatus would be used to find the intensity of the scene without shadows while the RGB camera would be used to find the
Figure 8.1: Dual-Camera Miniaturization
hue and saturation of the scene. The RGB hue and saturation information would be combined with the NIII intensity information to create a whole scene. This approach and other relighting methods such as proposed in [81] and others but could result in unnatural looking images. These two techniques could allow NIII to be extended to color images if required for the application.
References


31, 2006. Available from: http://dx.doi.org/10.1002/scj.20423.3.2.3


[41] Bernd Jähne and Horst Haußecker. *computer vision and applications: a guide for students and practitioners*. academic press, 525 B Street, Suite 1900, San Diego, CA, 92101, USA, 2000. 1.5.3


[46] Laurin Publishing, Box 4949, Pittsfield, MA, 01202-4949, USA. *The Photonics Spectrum Reference Wall Chart*, 2009. 7.1, 7.1.6, 7.2


1.5.3

133


[70] OSRAM Opto Semiconductors GmbH, Leibnizstrasse 4, D-93055 Regensburg. OSTAR - Projection LE T H3A, LE B H3A Preliminary Data, July 2008. 4.1.1, 8.2.2


---

1 n.b. the section number(s) where each reference appears in the text are listed at the end of each citation.
Appendix A

Video Attachments

This thesis is about taking images of moving scenes, and videos are an indispensable part of illustrating the performance of the techniques. All the videos listed are available from the UBC library at:

http://hdl.handle.net/2429/33535

These videos are compressed, as most video player software can’t load uncompressed video files. The compression will introduce some artifacts, typically smoothing the edges and some transitory blockiness. In some cases video has been cropped to allow compression. The layout of the videos is similar to the figures, but there are aesthetic differences due to the difference between the typesetting and video software.