

Spectral Image Utility for Target Detection Applications

by

Marcus S. Stefanou

B. S. United States Naval Academy, 1990
M. S. United States Naval Postgraduate School, 1997

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy,
in the Chester F. Carlson Center for Imaging Science,

College of Science

Rochester Institute of Technology

2008

Signature of Author _____

Accepted by _____
Coordinator, Ph. D. Degree Program

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Marcus S. Stefanou
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation requirement for the
Ph.D. degree in Imaging Science

Dr. John Kerekes, Dissertation Advisor

Dr. David Messinger

Dr. Carl Salvaggio

Dr. Joseph DeLorenzo

Dr. William Basener

Date

Spectral Image Utility
for Target Detection Applications

by

Marcus S. Stefanou

Chester F. Carlson Center for Imaging Science,
Rochester Institute of Technology

Abstract

In a wide range of applications, images convey useful information about scenes. The “utility” of an image is defined with reference to the specific task that an observer seeks to accomplish, and differs from the “fidelity” of the image, which seeks to capture the ability of the image to represent the true nature of the scene. In remote sensing of the earth, various means of characterizing the utility of satellite and airborne imagery have evolved over the years. Recent advances in the imaging modality of spectral imaging have enabled synoptic views of the earth at many finely sampled wavelengths over a broad spectral band. These advances challenge the ability of traditional earth observation image utility metrics to describe the rich information content of spectral images. Traditional approaches to image utility that are based on overhead panchromatic image interpretability by a human observer are not applicable to spectral imagery, which requires automated processing. This research establishes the context for spectral image utility by reviewing traditional approaches and current methods for describing spectral image utility. It proposes a new approach to assessing and predicting spectral image utility for the specific application of target detection.

We develop a novel approach to *assessing* the utility of any spectral image using the target-implant method. This method is not limited by the requirements of traditional target detection performance assessment, which need ground truth and an adequate number of target pixels in the scene. The flexibility of this approach is demonstrated by assessing the utility of a wide range of real and simulated spectral imagery over a variety

of target detection scenarios. The assessed image utility may be summarized to any desired level of specificity based on the image analysis requirements.

We also present an approach to *predicting* spectral image utility that derives statistical parameters directly from an image and uses them to model target detection algorithm output. The image-derived predicted utility is directly comparable to the assessed utility and the accuracy of prediction is shown to improve with statistical models that capture the non-Gaussian behavior of real spectral image target detection algorithm outputs.

The sensitivity of the proposed spectral image utility metric to various image chain parameters is examined in detail, revealing characteristics, requirements, and limitations that provide insight into the relative importance of parameters in the image utility. The results of these investigations lead to a better understanding of spectral image information vis-à-vis target detection performance that will hopefully prove useful to the spectral imagery analysis community and represent a step towards quantifying the ability of a spectral image to satisfy information exploitation requirements.

Acknowledgements

“Whatever you do, work at it with all your heart, as working for the Lord...”

Colossians 3:23

I would like to thank the members of my committee for their time, support, and patience. I also greatly appreciate the DIRS faculty and staff, who were always willing to entertain any of my questions or requests.

I am indebted to my advisor, John Kerekes, for providing the inspiration to investigate this topic, the foresight to tailor it to “play to my strengths,” and the discipline to hold me accountable along the way. His outstanding mentorship through personal example and work ethic are life lessons that I will always appreciate.

I am grateful to the Air Force for giving me the opportunity to attend this program. I would also like to thank all of the great men and women at the Air Force Research Laboratory Sensors Directorate at Hanscom Air Force Base, who encouraged me to pursue this path and taught me what good basic research entails.

Finally, I would like to thank my wife, Carrie. Her understanding, love, and positive attitude about life and family have made this endeavor possible.

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

Contents

List of Figures	x
List of Tables	xiv
1. Introduction	1
1.1 Semantics: Image Fidelity and Image Utility	1
1.2 Spectral Image Utility for Target Detection Applications	4
1.3 Assessment and Prediction of Image Utility	5
1.4 Motivation for Spectral Image Utility Prediction	7
1.5 Goal and Organization	8
2. Objectives	11
2.1 Objectives	11
2.2 Scope	11
2.3 Key Contributions	13
3. Background: Image Quality Review	15
3.1 Image Fidelity Measures	16
3.1.1 Sensor-derived	16
3.1.1.1 Scene-dependent	17
3.1.1.2 Scene-indepent	26
3.1.2 Image-derived	29
3.1.2.1 Single Image	29
3.1.2.2 Multiple Image	30
3.2 Image Utility Measures	35
3.2.1 Task Performance Measures	35
3.2.1.1 Direct Performance Measures	36
3.2.1.2 Performance Estimate Measures	38
3.2.2 Information Theory Measures	39
3.3 Image Utility Prediction Models	43
3.3.1 Parameter-based Models	44
3.3.1.1 Empirically-derived Predictive Models	45
3.3.1.2 Analytical Predictive Models	49
3.3.1.3 Image-derived	50
3.3.2 Simulation	51
3.4 Summary	52

4. Background: Spectral Imaging	55
4.1 Spectral Imaging Fundamentals	55
4.1.1 Physical Basis	56
4.1.2 Image Formation	56
4.1.3 Motivation for Spectral Imaging	58
4.1.4 Comparison with Panchromatic Imaging	58
4.1.5 Views of the Data	59
4.1.6 Spectral Imagery Characteristics	61
4.1.6.1 Spectral Contrast	61
4.1.6.2 Spectral Resolution	62
4.1.6.3 Spatial and Spectral Resolution Tradeoff	62
4.1.6.4 Spectral Mixing	63
4.1.6.5 Noise and Variability Characteristics	63
4.2 Spectral Imagery Application Areas	64
4.2.1 Target and Anomaly Detection	64
4.2.2 Unmixing	65
4.2.3 Change Detection	65
4.2.4 Classification	66
4.3 Spectral Image Target Detection	66
4.3.1 Reflectance Spectra	67
4.3.2 Theory and Design of Detectors	68
4.3.3 Detector Taxonomy	72
4.3.3.1 Full pixel	72
4.3.3.2 Subpixel	73
4.4 Target Detection Performance Assessment	74
4.5 Target Detection Performance Prediction	75
4.6 Summary	76
5. Prior Work: Spectral Image Quality	79
5.1 Overview	80
5.2 Spectral Image Fidelity Measures	81
5.2.1 Sensor-derived	81
5.2.1.1 Scene-dependent	81
5.2.1.2 Scene-independent	82
5.2.2 Image-derived	82
5.3 Spectral Image Utility Measures	84
5.3.1 Direct Performance Measures	84
5.3.2 Performance Estimates	84
5.4 Predicting Spectral Image Utility	86
5.4.1 Regression Approaches	86
5.4.1.1 General Image Quality Equation (GIQE)-like	87
5.4.1.2 Spectral Quality Equation (SQE)	87
5.4.1.3 Spectral Quality Rating Scale (SQRS)	88
5.4.2 Spectral Vectors	90
5.4.3 Spectral-spatial	91
5.4.4 Spectral Statistical Analytical Model	92

5.5 Summary	93
6. Approach	95
6.1 A Framework for Image Quality	95
6.2 Defining a Spectral Image Utility Metric	98
6.3 Assessing Spectral Image Utility	100
6.3.1 Target Implant Method	101
6.3.2 Subpixel Target Detection Algorithms	105
6.3.3 Summary Utility Metric	106
6.4 Predicting Spectral Image Utility	109
6.4.1 Image-derived Prediction Method	110
6.4.2 Statistical Models	111
6.5 Experiments to Demonstrate the Viability of the Utility Metric	113
6.6 Experiments to Characterize the Accuracy of Predicted Spectral Image Utility	115
6.6.1 Metrics for Prediction Accuracy and Performance	115
6.6.2 Experiments to Explore Sensitivity of Prediction Accuracy to Model Parameters	118
6.6.3 Experiments to Compare Image-derived Utility with Other Spectral Image Quality Methods	120
6.6.3.1 Target Type Specific Methods	121
6.6.3.2 Target Type/Size Independent Methods	122
6.6.3.3 Target Type/Size Specific Methods	122
6.7 Experiments to Investigate Utility Sensitivity	123
6.7.1 Conceptual Approach	124
6.7.2 Baseline Images	126
6.7.3 Scene Parameters	126
6.7.4 Image Acquisition Parameters	127
6.7.5 Preprocessing Parameters	128
6.7.6 Target Detection Scenario Parameters	129
6.8 Image Descriptions	129
7. Results	137
7.1 Demonstrate Viability of Utility Metric	137
7.1.1 Utility Metric Formation	138
7.1.2 Simple Observations of Utility Behavior	142
7.1.2.1 Detector Type	142
7.1.2.2 Target Type	145
7.1.2.3 Target Size	149
7.1.3 Utility Trends	152
7.1.3.1 Scene Composition	152
7.1.3.2 GRD	154
7.1.3.3 Spectral Resolution	156
7.1.3.4 Signal to Noise Ratio	159
7.1.4 Versatility of the Utility Metric	163
7.1.4.1 Specified PFA	163
7.1.4.2 Target Types	165
7.1.4.3 Target Sizes	166
7.1.4.4 Target Detection Algorithms	168

Contents

7.1.5 Comparing Images Using the Utility Metric	169
7.1.5.1 Different Sensors	169
7.1.5.2 Comparison of Different Scenes Using HYDICE Images	172
7.1.5.3 Comparison of Same Scene Using DIRSIG and HYDICE Images	174
7.1.6 Summary of Utility Metric Results	175
7.2 Image-derived Utility Prediction Results	176
7.2.1 Behavior of Predicted Utility for Three Images	177
7.2.1.1 Predicted Filter Output Distributions	177
7.2.1.2 Predicted ROC Curves	179
7.2.1.3 Predicted Utility	180
7.2.2 Exceedance Metric	182
7.2.2.1 PFA Goodness of Fit	182
7.2.2.2 PD Goodness of Fit	185
7.2.2.3 Selection of Optimal T-DOF Using PFA Goodness-of-Fit	188
7.2.3 Characterizing the Accuracy of Predictions	189
7.2.3.1 Absolute Error Between Predicted and Assessed Utility	189
7.2.3.2 Metric for Quantifying Utility Error over 4 Targets	191
7.2.4 Sensitivity of Predicted Utility to Prediction Parameters	194
7.2.4.1 T-Distribution Degree of Freedom Parameter	194
7.2.4.2 Number of Classes Used in Sum of Gaussian Model	196
7.2.4.3 Sample Support for Model Parameter Estimates	197
7.2.5 Prediction Accuracy and Time Tradeoff	198
7.2.6 Comparison with Other Spectral Image Quality Methods	201
7.2.6.1 Target Type Specific Methods	201
7.2.6.2 Target Type/Size Independent Methods	203
7.2.6.3 Target Type/Size Specific Methods	205
7.2.6.4 Observations	206
7.2.7 Summary of Utility Prediction Results	207
7.3 Sensitivity Analysis	208
7.3.1 Scene Parameters	209
7.3.1.1 Target Presence	209
7.3.1.2 Scene Composition	211
7.3.2 Image Acquisition Parameters	213
7.3.2.1 GRD	214
7.3.2.2 SNR	214
7.3.2.3 Spectral Resolution	215
7.3.2.4 Misregistration	216
7.3.2.5 Spectral Shift	219
7.3.3 Preprocessing Parameters	221
7.3.3.1 Compensation Method	221
7.3.3.2 ELM Parameters	224
7.3.3.3 Spectral Range	226
7.3.3.4 Bad Band Definition	228
7.3.3.5 Bad Reflectance Values	230
7.3.4 Target Detection Scenario Parameters	231
7.3.4.1 Target Variance	231
7.3.4.2 Target Type	234
7.3.4.3 Target Size	236
7.3.4.4 Specified PFA	237
7.3.5 Ranking of Parameters	239
7.3.6 Summary of Sensitivity Analysis	241
8. Summary	243
8.1 Summary	243

Contents

8.2 Original Contributions	245
8.3 Lessons Learned from Research Objectives	247
8.3.1 Image Utility Framework	247
8.3.2 Target Implant Image Utility Assessment	247
8.3.3 Image-derived Image Utility Prediction	249
8.3.4 Image Utility Sensitivity to Image Chain Parameters	250
8.4 Future Work	253
8.4.1 Within Current Research Scope	253
8.4.2 Beyond Current Scope	254
References	255

List of Figures

Figure 1.1	Simple image chain depicting concepts of image fidelity and utility	2
Figure 1.2	Some factors influencing image utility for target detection	3
Figure 1.3	Assessing and predicting image utility for target detection.....	6
Figure 1.4	Motivation for a spectral image utility metric	8
Figure 3.1	Organization of image quality review	15
Figure 3.2	Rayleigh Criterion of resolution (from Biberman, 1973).....	17
Figure 3.3	Ground sampled distance (from Fiete, 1999)	18
Figure 3.4	Edge spread function (from Schott, 2007).....	19
Figure 3.5	Relative edge response (from Leachtenauer and Driggers, 2001).....	20
Figure 3.6	Contrast transfer function (from Schott, 2007).....	21
Figure 3.7	MTFA (from Barten, 1990).....	24
Figure 3.8	Image scale (from Schott, 2007)	27
Figure 3.9	Ground spot size (from Fiete, 1999).....	28
Figure 3.10	λ FN/p detector sampling in the spatial domain (from Fiete, 1999).....	28
Figure 3.11	Four image degradations(from Wang and Bovik, 2002)	33
Figure 3.12	Probability distributions of observations in a binary decision task (from Leachtenauer and Driggers, 2001).....	37
Figure 3.13	Receiver operating characteristic curve (from Leachtenauer and Driggers, 2001)	37
Figure 3.14	Information theory view of an imaging system (from O'Sullivan, et. al., 1998).....	40
Figure 3.15	Visual communications channel (from Huck, et. al., 1992).....	41
Figure 3.16	Notional image utility prediction model (from Hailstone, 2005)	43
Figure 3.17	GIQE conceptual model (from Leachtenauer and Driggers, 2001).....	46
Figure 3.18	Objects and their corresponding bar targets (from Driggers, et. al., 1998)	47
Figure 3.19	Target acquisition system model(from Driggers, et. al., 1998)	48
Figure 3.20	Block diagram of FASSP model (from Kerekes, 2004).....	49
Figure 3.21	Synthetic image generation components (from Schott, 2007).....	52
Figure 4.1	Three spectral image formation methods(Descour and Dereniak, 1995)	57
Figure 4.2	Spectral variability models (from Manolakis, et. al., 2003)	60
Figure 4.3	Spectral variability (from Shaw and Burke, 2003).....	64
Figure 4.4	The target detection problem (from Manolakis and Shaw, 2002)	69
Figure 5.1	Organization of spectral image quality review.....	80
Figure 5.2	Surface of constant utility (from Kerekes and Hsu, 2004a)	89
Figure 6.1	Task-based framework for objective image quality assessment (from Barrett and Myers, 2004).....	96
Figure 6.2	Framework for image quality with spectral image utility shown as an example.....	97
Figure 6.3	Probability of detection at specified probability of false alarm as a utility metric	99
Figure 6.4	Definition of utility using the normalized area under the ROC curve.....	100
Figure 6.5	Target implant method of assessing spectral image utility.....	102
Figure 6.6	Flow diagram for image-derived spectral image utility prediction	111
Figure 6.7	Exceedance metric.....	116
Figure 6.8	Distance metric for prediction accuracy.....	118
Figure 6.9	Method for calculating prediction accuracy sensitivity to prediction parameter variation.....	119
Figure 6.10	Some of the image chain parameters that impact assessed utility	124
Figure 6.11	Conceptual approach to defining utility sensitivity to one parameter	125
Figure 6.12	HYDICE Forest Radiance I Run 05, 07, and 09 images	130
Figure 6.13	HYDICE image from Desert Radiance II Run 03	131
Figure 6.14	HYDICE Terrain (left) and Urban (right) images	131
Figure 6.15	MISI image of RIT Campus	132
Figure 6.16	AVIRIS Lunar Lake (left) and Moffett Field (right) images.....	132
Figure 6.17	HyMap image subset of Cooke City	132
Figure 6.18	Quickbird image and spectral response functions	133
Figure 6.19	ALI Cuprite image and spectral response functions.....	134
Figure 6.20	DIRSIG Megascene image analyst images	135

List of Figures

Figure 7.1	Complete and detailed views of the spectral matched filter output target absent (TA) and target present (TP) distributions for Forest Radiance I Run 05 0.7 m c6 target	138
Figure 7.2	Receiver operating characteristic curve and utility metric formation for FR I Run 05 with 0.7 m c6 target	140
Figure 7.3	Area under ROC curves for two target detection scenarios.....	141
Figure 7.4	Distributions (left) and ROC curves (right) for the 0.7 m c6 target in the FRI Run 05 image for SMF (top), SAM (middle), and ACE (bottom) detectors	143
Figure 7.5	Distributions (left) and Spectra of First False Alarms (right) for 1 m c6 target in the DIRSIG 4 m 100 SNR image for SMF (top), SAM (middle), and ACE (bottom) detectors	145
Figure 7.6	Distributions and ROC Curves for the 0.7 m c6 (left) f8 (right) targets in the FRI Run 05 Image for SMF (top), SAM (middle), and ACE (bottom) detectors	146
Figure 7.7	Comparison of image and target spectra.....	147
Figure 7.8	Target c6 and f8 demeaned target vectors (left), FR I Run 05 inverse image covariance matrix (center), and resulting spectral matched filter vectors (right)	148
Figure 7.9	Targets f2, f4, f8, c6, vf124	149
Figure 7.10	Target absent and present distributions for four targets and ten implant fractions in FR I Run 05 image	150
Figure 7.11	ROC curves for four targets and ten implant fractions in FR I Run 05 image.....	150
Figure 7.12	Utility plots for four targets and ten implant fractions in FR I Run 05 image	151
Figure 7.13	Utility versus target size curves for four targets and ten target sizes in FR I and DR II images	153
Figure 7.14	Utility versus target size curves for four targets and ten target sizes in 2 m and 4 m DIRSIG Megascene images.....	154
Figure 7.15	Target absent and target present distributions for DIRSIG 2 m and 4 m images with first false alarm spectra.....	155
Figure 7.16	Utility versus target size for four targets in DIRSIG 4 m 100 SNR 37 channel and 145 channel images.....	156
Figure 7.17	Mean spectra (top) and covariance matrices (bottom) of DIRSIG 4 m 100 SNR 145 channel (left) and 37 channel (right) images	157
Figure 7.18	Filter vector, resulting filter output distributions, and spectra of first 20 false alarms for 1 m c6 target in DIRSIG 4 m 100 SNR 37 channel and 145 channel images.....	158
Figure 7.19	Spatial distribution of first twenty false alarms for 1 m c6 target in DIRSIG 4 m 100 SNR 37 channel and 145 channel images.....	158
Figure 7.20	Utility versus target size for four targets in DIRSIG 4 m 100 and 10 SNR image	159
Figure 7.21	Mean spectra (top left), inverse variances (top right), and filter vectors (bottom) for 1 m c6 target in DIRSIG 4 m 100 and 10 SNR images	160
Figure 7.22	Filter output distributions (left) and first twenty false alarm spectra (right) for 1 m c6 target in DIRSIG 4 m 100 (bottom) and 10 (top) SNR images.....	161
Figure 7.23	Spatial distribution of first twenty false alarms for 1m c6 target in DIRSIG 4m 10 and 100 SNR images.....	162
Figure 7.24	Window functions for utility calculation.....	164
Figure 7.25	Effect of utility window function on summary utility	165
Figure 7.26	Two summary utilities for target type	166
Figure 7.27	Detector type effect on utility	168
Figure 7.28	Utility of six spectral images plotted against target size.....	171
Figure 7.29	Four HYDICE image utilities plotted versus target implant fraction	173
Figure 7.30	Utility versus target size for two AVIRIS images	173
Figure 7.31	Utility of HYDICE FR I Runs 05, 07, and 09.....	174
Figure 7.32	Utility plots for nine images from the DIRSIG Megascene.....	175
Figure 7.33	Assessed and predicted filter output distributions for target f8 in FR I Run 05.....	178
Figure 7.34	Assessed and predicted ROC curves for target f8 in FR I Run 05.....	179
Figure 7.35	Assessed and predicted utility versus target size in FR I Run 05 for four targets at a specified PFA of 5×10^{-4}	180
Figure 7.36	Assessed and predicted utilities versus target implant size in HyMap image for four targets at a specified PFA of 5×10^{-4}	181
Figure 7.37	Assessed and predicted utilities versus target implant size in Quickbird image for four targets at a specified PFA of 5×10^{-4}	181

List of Figures

Figure 7.38 Assessed and predicted PFA vectors in FR I run 05 image for four targets	182
Figure 7.39 Assessed and predicted PFA vectors in HyMap image for four targets.....	183
Figure 7.40 Assessed and predicted PFA vectors in Quickbird image for four targets.....	184
Figure 7.41 Assessed and predicted PD vectors in FR I Run 05 image for one target.....	185
Figure 7.42 Predicted exceedance errors as a function of target size in FR I Run 05 image for four targets	186
Figure 7.43 Predicted exceedance errors as a function of target size in HyMap image for four targets	187
Figure 7.44 Predicted exceedance errors as a function of target size in Quickbird image for four targets	187
Figure 7.45 Optimal DOF selection for T-distribution PFA vector in FR I run 05 Image for four targets	188
Figure 7.46 Optimal DOF selection for T-distribution PFA vector in HyMap and Quickbird images for four targets.....	189
Figure 7.47 Utility error in FR I Run 05 image for four targets.....	190
Figure 7.48 Utility error in HyMap image for four targets	190
Figure 7.49 Utility error in Quickbird image for four targets	191
Figure 7.50 Utility scatter plot for FR I run 05 image for four targets and four prediction	192
Figure 7.51 Utility scatter plot for HyMap image for four targets and four prediction.....	192
Figure 7.52 Utility scatter plot for Quickbird image for four targets and four prediction.....	193
Figure 7.53 Utility error for T-distribution plotted against DOF parameter for FR I Run 05 and HyMap (left) and Quickbird (right) images.....	195
Figure 7.54 Utility error plotted against number of classes for FR I, HyMap, and Quickbird images	196
Figure 7.55 Utility error plotted as a function of sample support for normal model parameter estimates in three images	198
Figure 7.56 Tradeoff between prediction accuracy and time	200
Figure 7.57 Comparison of target-specific spectral image quality methods for ten images at specified PFA 1×10^{-3}	202
Figure 7.58 Scatter plot of assessed utility versus predicted utility for target-specific spectral quality methods in ten images.....	203
Figure 7.59 Comparison of target type/size independent methods with assessed and predicted utility for ten images.....	204
Figure 7.60 Comparison of GSUM with assessed and predicted utility methods for ten images	205
Figure 7.61 Utility versus number of c6 pixels in the FR I Run 05 image.....	210
Figure 7.62 Utility for DIRSIG 4 m 100 SNR image plotted against number of c6 target pixels in the image	210
Figure 7.63 Utility for 1 m c6 target versus number of spectral classes in the DIRSIG 4 m 100 SNR image	211
Figure 7.64 Utility plotted against target-background class statistical distance in DIRSIG image.....	212
Figure 7.65 Utility plotted against target-background class statistical distance in FR I image.....	213
Figure 7.66 Utility sensitivity to GRD for DIRSIG images	214
Figure 7.67 Utility plotted against SNR for DIRSIG images.....	215
Figure 7.68 Utility plotted against spectral resolution for 4 m 100 SNR DIRSIG image	216
Figure 7.69 Image mean and inverse covariance diagonal illustrating misregistration.....	216
Figure 7.70 Utility plotted against pixel misregistration.....	217
Figure 7.71 Spatial distribution of false alarms in original and misregistered DIRSIG images	218
Figure 7.72 FR I Run 05 subset mean vector with different amounts of spectral shift	219
Figure 7.73 Effect of spectral shift on inverse covariance matrix diagonal (left) and filter vector (right).....	220
Figure 7.74 Utility plotted against spectral shift (left) and filter output distributions at three spectral shifts (0 – top right, 0.3 – middle right, 1 – bottom right)	220
Figure 7.75 Comparison of retrieved reflectance spectra for 4% calibration Panel in FR I Run 05 image.....	222
Figure 7.76 Utility plotted against retrieved reflectance error for FR I Run 05 image	223
Figure 7.77 Utility plotted against reflectance errors for various ELM gain and offset scalings.....	224
Figure 7.78 Utility plotted against scale factor for elm gain and offset parameters in the DRISIG image	225
Figure 7.79 Effect of scaling ELM gain and offset on image mean spectrum	226
Figure 7.80 Image and three target spectra and Mahalanobis distances in FR I Run 05 image	227
Figure 7.81 Utility plotted against spectral range in FR I run 05 and DIRSIG images.....	227
Figure 7.82 False alarm spectra and filter vector for 0.7 m c6 target in FR I Run 05	229
Figure 7.83 Utility plotted against percent of bad bands included.....	229
Figure 7.84 Utility as a function of target covariance scaling.....	232
Figure 7.85 Distributions and ROC curves for 1 m c6 utility in DIRSIG image for three different target variability scalings	232
Figure 7.86 Utility plotted against target covariance scaling for two targets in DIRSIG	233

List of Figures

Figure 7.87 Filter output distributions for 2 m vf124 target in DIRSIG image.....	234
Figure 7.88 Filter vectors for three targets in the FR I Run 05 image.....	235
Figure 7.89 Distributions and ROC curves for three targets in FR I Run 05	235
Figure 7.90 Utility of 19 targets plotted against their Mahalanobis distance for the FR I run 05 image (left) and the DIRSIG image (right)	236
Figure 7.91 Utility versus target size for FR I Run 05 and DIRSIG 4m 100 SNR images	237
Figure 7.92 Filter output distributions and ROC curves for three specified PFAs.....	238
Figure 7.93 Utility plotted against Specified PFA in FR I and DIRSIG images.....	238
Figure 7.94 Ranking of parameters based on utility sensitivity for the DIRSIG image.....	241

List of Tables

Table 5.1	Multispectral Image Interpretability Rating Scale (from Irvine, 1997)	85
Table 7.1	SCRs for Five Targets in FR I Run 05 Image.....	149
Table 7.2	SCRs for Four Targets in FR I and DR II Images	153
Table 7.3	Summary of Utility Using Target Size Ranges	167
Table 7.4	Average Utility Summarized over Different Target Size Ranges for Six Images	171
Table 7.5	Exceedance Metric for Predicted PFA Vectors in FR I Run 05 Image for Four Targets	183
Table 7.6	Exceedance Metric for Predicted PFA Vectors in HyMap Image for Four Targets	184
Table 7.7	Exceedance Metric for Predicted PFA Vectors in Quickbird Image for Four Targets	185
Table 7.8	Average Distance of Predicted from Assessed Utility for Four Prediction Methods	193
Table 7.9	Utility Errors and Computation Times with Decreased Sample Support for Three Images and 20% f4 target	199
Table 7.10	Parameters Used in Calculation of SQRS and SQE Spectral Image Quality Metrics	204
Table 7.11	Rank Ordering of Image Utilities by Spectral Quality Methods	209
Table 7.12	Effect on Utility of Data Cleansing for FR I Run 05 ELM, ATREM, and FLAASH Images	230
Table 7.13	Sensitivity Scores for FR I Run 05 Image Parameters	240
Table 7.14	Sensitivity Scores for DIRSIG Image Parameters	240

Chapter 1

Introduction

Webster's Dictionary defines *quality* as the degree of excellence of a thing or process. This vague statement needs further focus in order to be useful for our purposes. As a start, it is helpful to anchor the general concept of quality with a mental picture of a specific process. In the case of images, this mental picture is the image chain, which describes the process of image formation and subsequent processing for information extraction in the most general terms. Fundamentally, the objective of imaging is to gain information about a scene. This process begins when energy from some source interacts with the material constituents of a scene, propagates through a medium, and is transformed into a measurable quantity by the sensor. The observations at the sensor are inherently noisy due to the variability and randomness of the scene, the propagating medium, and the sensor itself. These noisy observations constitute the image from which we seek to gain information.

We are interested in describing the quality of images. As will become clear shortly, this is a very open-ended proposition. If we are to successfully navigate this topic, and make a useful contribution, then we need to be deliberate in our approach. Therefore, we begin with our working definitions of terms, so that there is no ambiguity as we proceed.

1.1 Semantics: Image Fidelity and Image Utility

We propose that the *fidelity* of an image is concerned with how closely the image represents the scene. It may be defined along spatial, radiometric, and geometric dimensions (Schott, 2007). Spatial fidelity describes the extent to which the image preserves scene's relative size, shape, and detail; radiometric fidelity describes how well the image preserves the scene's energy distribution; and geometric fidelity quantifies how well the scene's relative or absolute positions are preserved (Leachtenauer and Driggers, 2001). Thus, an image with high fidelity would be one that closely resembles the object being imaged (assuming that we can quantitatively compare the image and object) and is free of artifacts and other distortions.

When we consider the concept of image quality, it is likely that we are thinking about the fidelity of the image since the human observer perceives the accurate representation of spatial detail as the primary attribute of a high quality image. During sensor design

evaluation, the fidelity of the image is readily measured when an object of known spatial extent is present in the scene to be imaged. The ability of the sensor to faithfully represent such a test pattern (e. g. the standard bar target) is quantified by the image fidelity, and has become a well established standard for specifying the sensor performance.

It is our view that image fidelity is one part of a more encompassing definition of image quality, in which image quality is comprised of an *image fidelity* component and an *image utility* component. Figure 1.1 depicts the image chain along with the activities of image acquisition and information exploitation. The concepts of image fidelity and image utility are also shown, and are defined with reference to their location in the image chain and

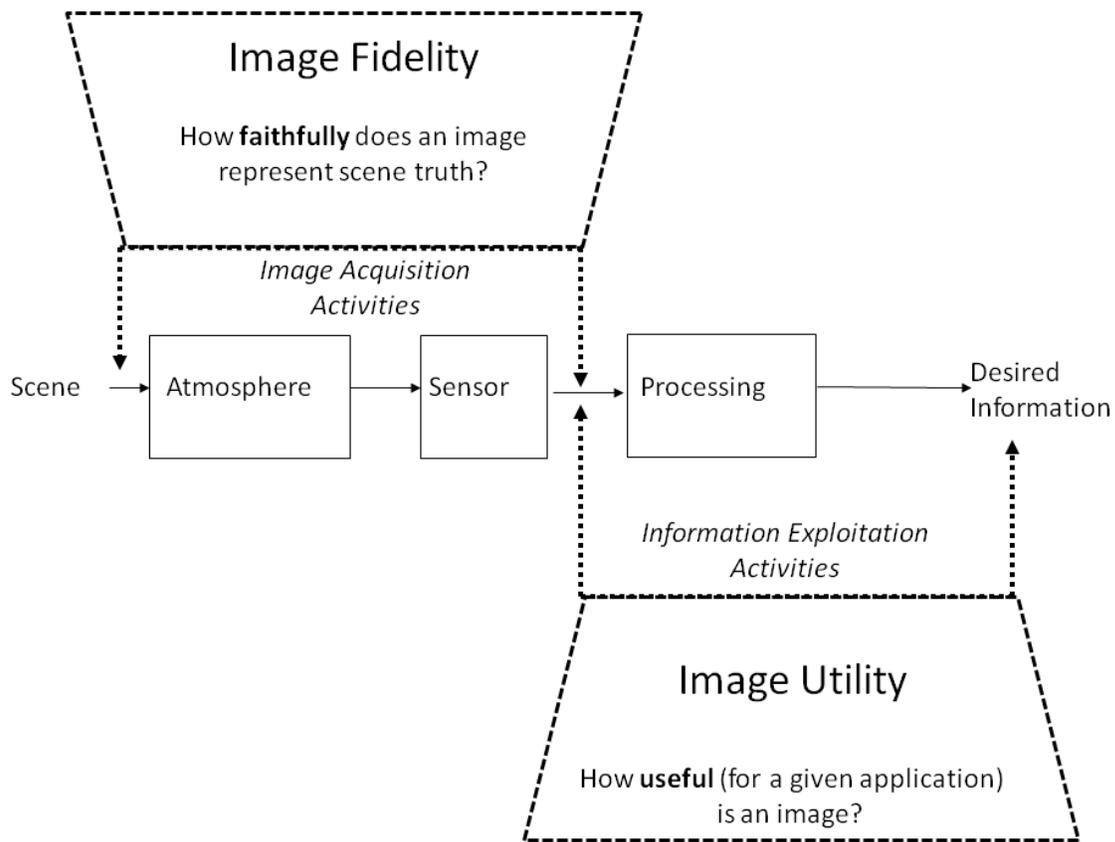


Figure 1.1: Simple image chain depicting concepts of image fidelity and utility

the particular question that each seeks to answer. Image fidelity compares the image to the scene and asks how faithfully the image has represented the scene, whereas image utility compares the information product to the image and asks how useful the image has been in yielding the desired information. Historically, image fidelity has been virtually synonymous with image quality.

The *utility* of the image is the image’s ability to deliver information about the scene. If an image is unsuited for delivering a certain type of information product, then it is judged as having low utility. We specify the image utility relative to the specific information desired from an image as defined by the ultimate task or application. The specific processing employed to exploit information greatly influences the selection of an appropriate image utility metric. The utility is also influenced by factors over which the observer (image analyst) has no control, such as the scene and the conditions under which the image acquisition occurred.

To illustrate the challenging nature of specifying image utility, consider Figure 1.2, which depicts some of the many factors that influence the ultimate utility of an image (in this particular case for a target detection application). Despite the apparently complicated nature of image utility, it is not an unfamiliar concept to us. Image utility is simply another term for

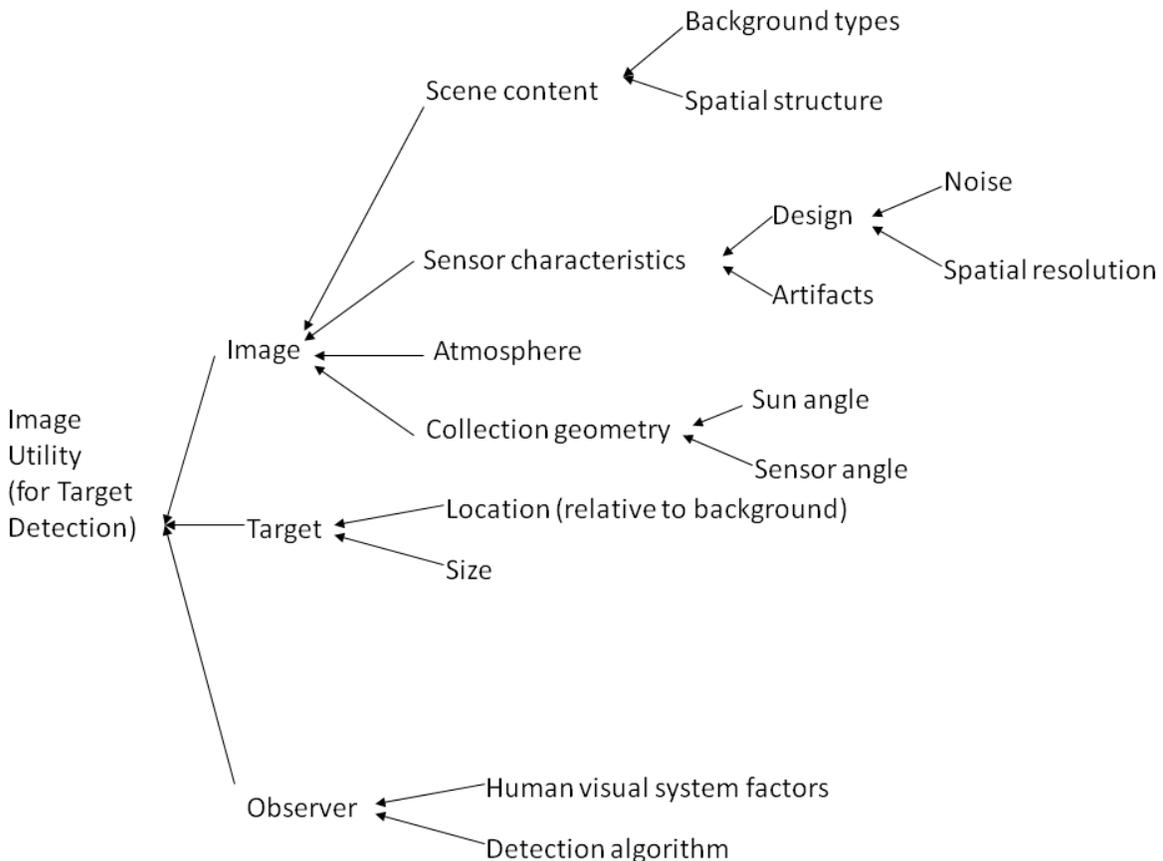


Figure 1.2: Some factors influencing image utility for target detection

the measured performance of the information exploitation operations in the image chain of Figure 1.1. Metrics such as interpretability or probability of detection are used to describe the

image utility for image analysis tasks. The subtle difference between an image utility metric and an application performance measure is based on one's perspective: image utility focuses on labeling a specific image's usefulness in a particular application whereas a performance measure focuses on describing the performance of the application across many images. While image utility has been studied in the context of panchromatic imagery and predicting image utility, there has been very little systematic exploration of spectral image utility.

We stress that despite the apparent separate domains of image fidelity and image utility in Figure 1.1, image fidelity and image utility are not mutually exclusive terms. The utility of an image is impacted by the fidelity with which it was gathered. This is because the scene information collected by the sensor during image acquisition is “embedded” in the image that serves as the starting point for the information exploitation. Likewise, sensor design activities specify the desired image fidelity based on the intended purpose of the image in many cases. E. H. Linfoot, in a 1958 paper, discusses the interdependent nature of image fidelity and image utility (which he calls ‘information’), “...it is worth considering whether a departure from maximum image-fidelity may not some times allow a useful gain in the amount of information recorded. If the arithmetical recoding of optical images were a standard practice today, instead of a prospect for the future opened up by the advent of the fast computing machines, we would go on to add that informationally optimized designs were always to be preferred” (Biberman, 1973).

1.2 Spectral Image Utility for Target Detection Applications

The breadth of the topic of image quality is enormous. The model presented in Figure 1.1 allows us to consider the various modalities of image formation and many possible applications that use images as the source of information. Our interest lies in considering the utility of aerial and satellite spectral imagery of terrestrial scenes in the context of subpixel target detection. Spectral imaging is the measurement, analysis, and interpretation of radiation arriving at an electro-optical sensor at many wavelengths over a broad spectral band and a large spatial area (Shaw and Burke, 2003). Subpixel target detection is a spectral imagery application in which the presence of a material of interest is detected at finer resolution than the minimum spatial sampling extent of the sensor by virtue of the spectral characteristics of the target and background materials. In order to gain traction on the image

quality topic, we narrow the focus of this research to image utility for a specific imaging modality and application area.

The choice to narrow the research is not just a pragmatic decision – it represents an opportunity to explore uncharted territory. As a means of providing perspective, consider that spectral imagery based on imaging spectrometers has only been in existence for about 40 years, compared to some 200 years for photography. Thus, there is much about this imaging modality that still needs to be understood if we are to maximize its potential. The utility of spectral imagery for target detection provides a rich area to investigate for several reasons. First, image fidelity and utility are well characterized in the literature of sensor design tradeoffs for panchromatic aerial and satellite imaging systems, but not for spectral imaging systems. Second, we take the perspective of the consumers of spectral imagery (the image analyst community) rather than sensor designers in this research. Thus, the relevant topic is not how faithfully the spectral image reproduces the scene, but rather what information we can glean from various types of spectral images. The image analyst is not concerned with sensor design parameters but with figures of merit that pertain to the ability to accomplish his or her job – extracting information from images. Third, whereas the information exploitation of panchromatic images relies almost exclusively on the human observer, spectral imagery is exploited by computer algorithms that operate on the multi-dimensional spectral image space. The performance of these algorithms across a wide range of spectral images is not well understood. Thus, we hope that understanding the utility of spectral images in a specific application will give insights into target detection performance that will be extensible to other application areas, thereby increasing the appreciation for the useful information contained in spectral imagery.

1.3 Assessment and Prediction of Image Utility

Two further definitions are central to our research of spectral image utility for subpixel target detection. These definitions pertain to the manner in which we estimate the utility of an image. *Assessing* image utility is an activity that frequently occurs in dealing with images. This is because utility is measured or assessed using the metrics created for measuring application performance. Thus, the activity of assessing utility is synonymous with measuring performance. In the case of spectral imagery, this is done most simply by applying

an information exploitation algorithm to the image and noting the results. *Predicting* the utility of an image necessarily involves a prognostication of performance in lieu of actual application of an information exploitation algorithm to the image. The prediction of utility (performance) requires some sort of model of the information exploitation processing. This model may take different forms, some of which include: a simulation of the image formation process, in which the simulated image is subjected to information exploitation; a parametric representation of the image, in which the parameters are processed as surrogates for the image; or a functional expression that relates sensor parameters to utility. Figure 1.3 depicts the generic activities of assessing and predicting the utility of an image for a subpixel target detection task, with the utility metric represented as a receiver operating characteristic (ROC) curve that expresses the probability of detecting the target as a function of the probability of false alarm. We see that the activities of assessing and predicting utility are two different means of arriving at the same utility metric. We also see that the assessment of utility entails operating on the real or simulated image, whereas predicting utility involves using image or sensor parameters as inputs to a model.

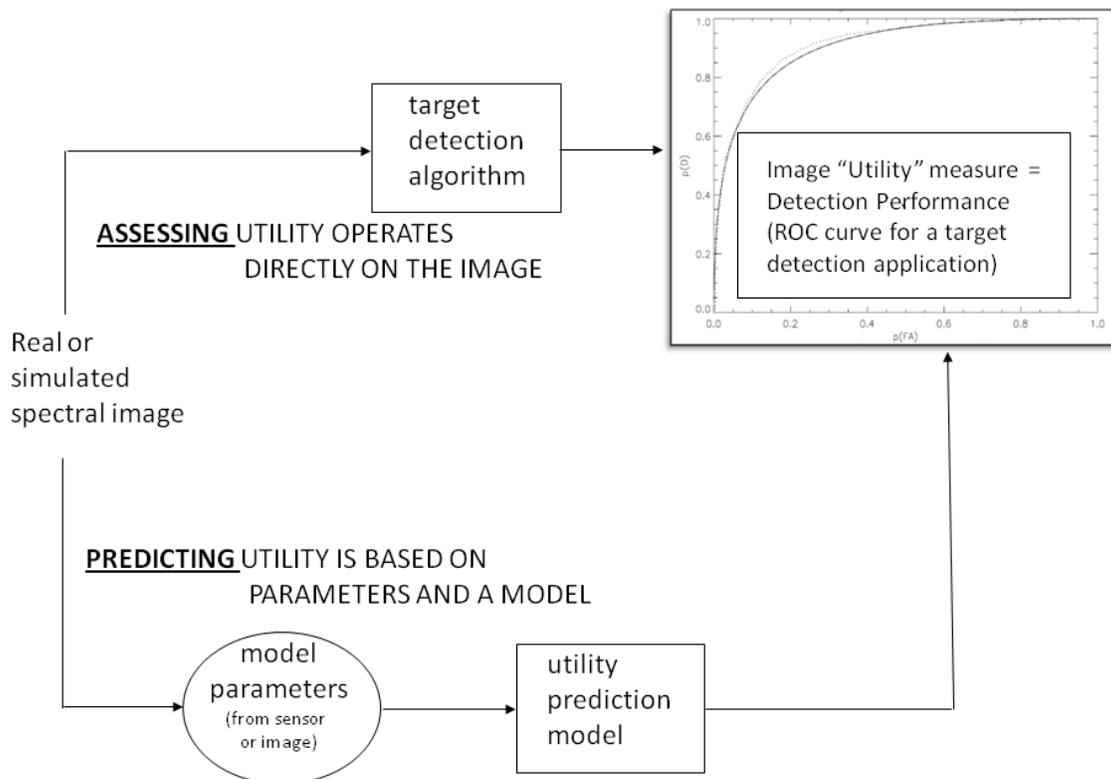


Figure 1.3: Assessing and predicting image utility for target detection

1.4 Motivation for Spectral Image Utility Prediction

There is a compelling need for a spectral image utility metric and the capability to assess images using such a metric. Such a metric could be applied to many images to build a catalog of utility-labeled images covering many image acquisition and application-specific scenarios. The robust ability to assess the image utility for a given application would create the foundation for an image archive indexing scheme. An archive of spectral images, each with a utility label, would then facilitate sensor design trade studies and provide a basis for guiding future image collection activities. These important capabilities, which use current images to optimize future sensor designs and imagery collection activities to acquire the most useful images, are only possible with a reliable utility metric and a consistent method to assess the utility of many images.

There is an equally important need to be able to accurately predict the assessed utility of any spectral image more rapidly than assessing it. The benefit of such a predictive capability is that it does not rely on an extensive scene simulation or a notional parametric description of the scene composition, because the parameters used in the utility prediction are estimated directly from the image. The ability to predict an easily calculated, quantifiable, application-specific measure of utility on a per-image basis would be beneficial for several reasons. It allows imagery requirements to be documented, facilitating the prioritization and tasking of imagery collection activities based on the types of images most likely to fulfill desired informational requirements. It enables communication of image usefulness relative to specific exploitation tasks, providing a means of evaluating the credibility of information extracted from a particular image. If the predicted image utility is associated with specific sensor parameters, trade studies of image utility sensitivity relative to design parameters would facilitate the development of future imaging systems. Figure 1.4 depicts the above reasons for a spectral utility metric by posing analysts' questions and analysis tasks which a utility metric and the ability to assess and predict it for any spectral image could help answer.

Most fundamentally, we believe that the exploration of spectral image utility will lead to a deeper understanding of spectral image information content, and is a topic worth considering now. The ability to consistently assess and predict the utility of spectral images is a capability needed before the next generation of imaging spectrometers produces a flood of imagery.

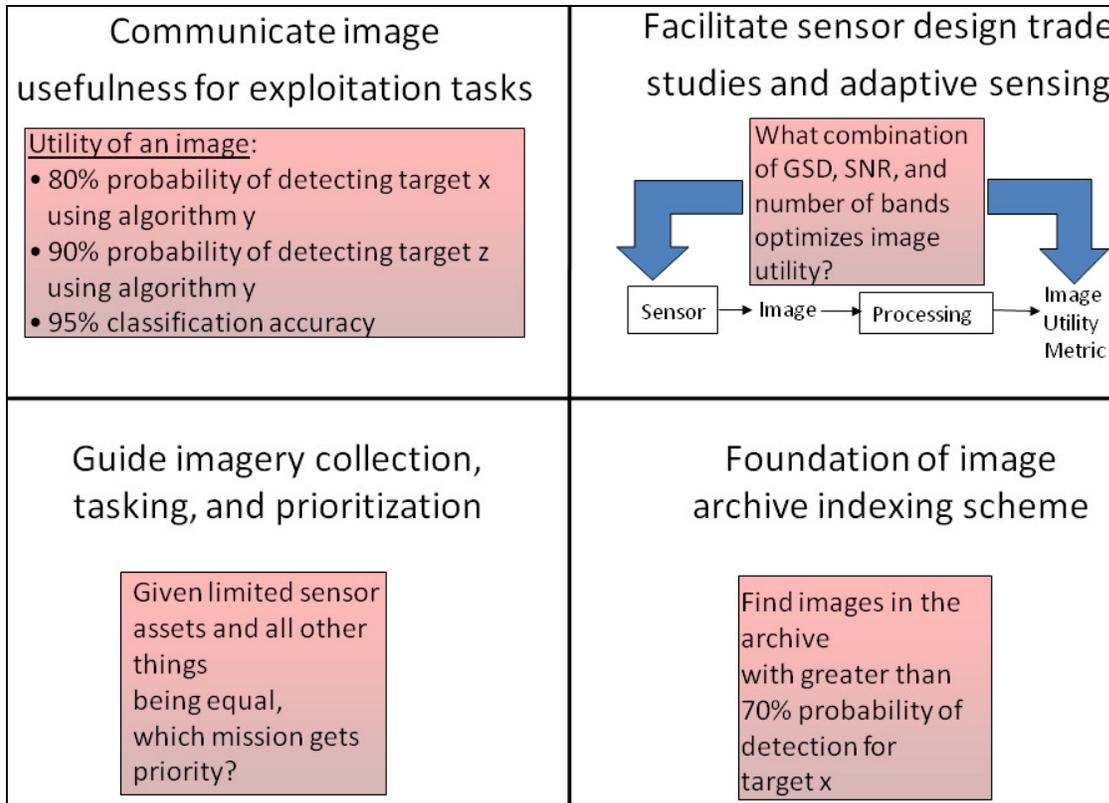


Figure 1.4: Motivation for a spectral image utility metric

1.5 Goal and Organization

The goal of this research is to create a spectral image utility metric and the capability to assess and predict it for any spectral image. The first step towards achieving this goal is to review existing image quality approaches and create an overarching framework through which to view image quality and specifically spectral image utility. The second step is to create a simple, reliable method to assess spectral image utility along with methods to predict it and to understand the characteristic behavior of the assessed and predicted utility for a variety of images and image chain parameter settings.

The organization of this proposal proceeds as follows. Chapter 2 describes the objectives and scope of the proposed research. Chapters 3, 4 and 5 represent the background necessary to establish the foundations and context that enable the meaningful discussion of spectral image utility prediction. Chapter 3 reviews traditional image quality metrics, approaches, and methods primarily drawn from optical remote sensing. It examines how image fidelity and utility are assessed and how image utility is predicted. Having established the context of general image quality, Chapter 4 then looks at spectral imagery in detail. It

Chapter 1. Introduction

develops the concept and characteristics of spectral imaging, examines how spectral image analysis differs from traditional panchromatic digital image analysis paradigms, and looks at the specific application of subpixel target detection in spectral imagery along with how target detection performance is measured. Chapter 5 investigates the current literature for assessing and predicting the fidelity and utility of spectral imagery. Chapter 6 presents the approach to accomplish the research objectives. Chapter 7 discusses results observed toward accomplishment of the research objectives. Chapter 8 provides a summary, describes the original contributions of this work, and proposes future research areas.

Chapter 2

Objectives

2.1 Objectives

This research is motivated by the belief that a thorough understanding of existing approaches to image quality is a prerequisite to developing a general framework robust enough to address the various applications of spectral imaging. Likewise, an appreciation for the unique nature of spectral imagery and the associated information exploitation techniques is necessary in order to create a spectral image utility assessment and prediction capability. With these thoughts in mind, there are four objectives that this research seeks to accomplish:

- a. Establish a unifying framework that enables image quality to be consistently discussed across applications.
- b. Develop a spectral image utility metric and assess image utility using the metric for a variety of spectral images and detection scenarios in the subpixel target detection application.
- c. Develop a means to predict spectral image utility and quantify the accuracy of the predictions relative to assessed utility.
- d. Understand the sensitivity of utility assessments and predictions to perturbations of prediction model and image chain parameters.

The first objective pertains to the activity of creating a larger framework and placing spectral image target detection within its context. The last three objectives develop, characterize, and evaluate specific methods of assessing and predicting spectral image utility for subpixel target detection applications.

2.2 Scope

In addressing a topic as broad as image quality, specific choices need to be made in order to make a meaningful contribution. After the broad discussion of image quality, we make the first choice of focusing on image utility rather than image fidelity. It is our belief that image utility is the more relevant aspect that is accessible to a wider audience than image fidelity.

Although we have tried to be as comprehensive as possible in the review of image quality, there are undoubtedly specific image quality measures that are overlooked due to the far ranging nature of the topic, while other areas are intentionally not reviewed in great depth. While we touch on the area of human visual perception and cognition for the sake of completeness, we do not dwell on them because our ultimate interest is not in the human observer, but in computer processing to exploit information. Similarly, we do not touch the topic of color perception by the human visual system.

We narrow our consideration further to the specific modality of spectral imaging and the specific application of subpixel target detection. Spectral imaging represents an area full of potential for increased understanding, and target detection is an activity which has clearly defined performance measures and for which utility is a relatively well understood concept.

We next make decisions that make the development of image utility assessment and prediction methods tractable. First, we use remotely sensed images collected by spectral imagers operating in the visible to shortwave infrared portion of the spectrum (0.4 – 2.5 μm). This is because there are many available images in this domain. Second, we assume that the images for which we will be assessing and predicting utility have been atmospherically compensated so that we operate in the reflectance domain. While somewhat tedious to ensure that all images are in the reflectance domain, we use this approach so that we may employ our reference target spectra. However, we note that our approach could be applied in a forward manner, operating on images in radiance or sensor counts, provided that we can propagate the target spectra forward into this domain. Third, we search for subpixel targets which have been fractionally mixed with background spectra in a linear fashion. This simple linear mixing model allows us to control the fractional pixel area that we hypothetically assign as target. This mixing model does not take the nonlinear effects of shadow, non-Lambertian surfaces, or sloped earth surfaces into account. Fourth, while we consider three target detection algorithms in our utility assessment method, we only use one of them, the spectral matched filter, in utility prediction. This is because the filter's action on image pixel vectors may be represented as the linear transformation of multivariate to scalar random variables. Fifth, we assume that the output of the detection algorithm is the end of the image chain. In other words, no further image processing or human assessment is required in order to assess the utility of the image. The detection algorithm output is the information that we

seek, not the conclusion of a human observer. This makes the problem simpler and avoids the difficulty of describing how an image analyst makes an assessment based on various sources of information or how the human visual system processes spatial information.

2.3 Key Contributions

This research offers five key contributions to the state-of-the-art of image quality and spectral imaging. First, the definitions and framework for considering image quality provide a robust means by which to organize the many existing techniques both for traditional and spectral imagery. Such an organizational scheme allows us to see common themes and differences and helps point out areas where further opportunities for new methods exist. Second, a simple, widely applicable utility metric that describes the usefulness of a spectral image for a particular target detection task or range of tasks will provide great flexibility to image analysts in accomplishing common image analysis tasks. Third, the capability to predict the utility of a spectral image using image-derived parameters allows image-specific utility predictions that will be instrumental in planning collection activities, conducting sensor design trade studies, and labeling images in a manner relevant to image analysts. Fourth, the spectral image utility assessment capability will build a foundation for a fuller understanding of the behavior of target detection (and eventually other applications) in spectral imagery. This will hopefully help realize the full potential of spectral imagery by capitalizing upon the unique information captured by imaging spectrometers. Fifth, the understanding of the relative role of image chain parameters on the utility metric will streamline the development of a robust utility metric by focusing efforts on those parameters having the greatest impact on utility. An appreciation of utility sensitivity to image chain parameters will also help reveal conditions under which the utility metric may fail to accurately characterize the situation, thus adding value to the analyst's confidence in interpreting the utility metric.

Chapter 3

Background: Image Quality Review

The large body of literature on image quality might be categorized in many ways. We could review the works historically, by discipline, or by process. Barrett and Myers (2004) discuss several means by which imaging systems might be classified – by type of radiation or field used to form the image, by property of the object to be displayed in the image, by imaging mechanism, by whether the image is directly captured or reconstructed, or by whether the imaged radiation is part of the imaging system. Given the hundreds of imaging modalities and applications, the task of categorizing image quality is challenging. We simplify matters by using our constructs of image fidelity and image utility to show how traditional measures of image quality fit into one of these two categories. We simplify further by confining our survey to passive optical remote sensing.

Realizing that we must pick an organizational scheme, we organize our image quality review into three major parts: image fidelity measures, image utility measures, and image utility prediction. Figure 3.1 provides a roadmap for this review scheme and illustrates the interconnection of the three major parts. Section 3.1 begins with a discussion of image

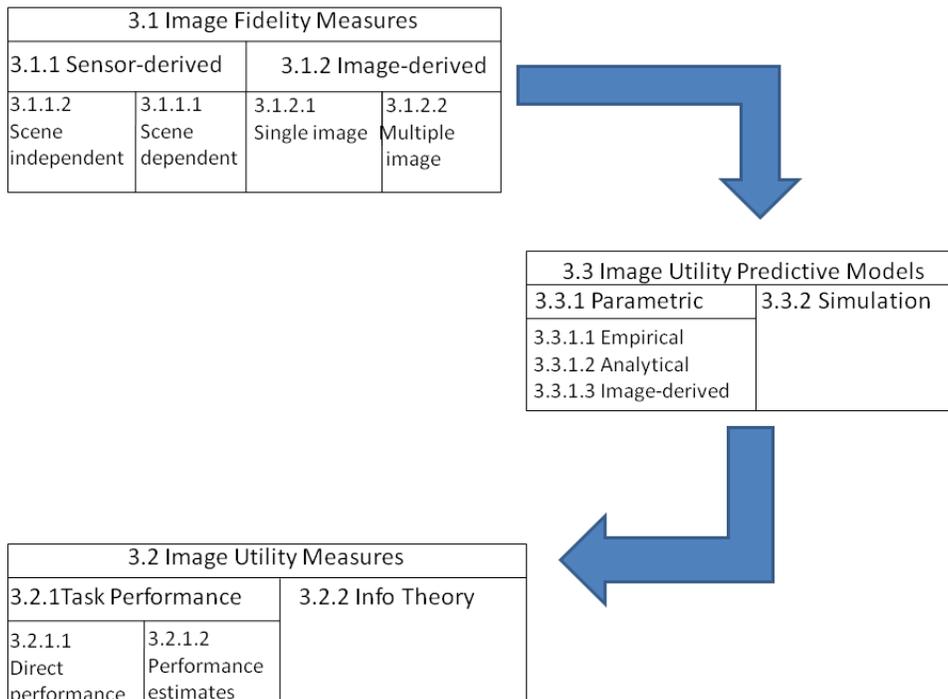


Figure 3.1: Organization of image quality review

fidelity measures. Image fidelity is assessed (measured) using either sensor- or image-derived parameters. Sensor-derived parameters (Section 3.1.1) are divided according to whether or not their calculation depends on the scene being imaged. Image-derived (Section 3.1.2) parameters are categorized based on the number of images used to calculate the image fidelity parameters. Section 3.2 presents image utility measures. These are categorized by methods based on task performance (Section 3.2.1) or information theoretic (Section 3.2.2) approaches. Task performance approaches are further classified by whether the metric is measured based on statistical decision theory (direct performance measures) or estimated via human observers. Section 3.3 presents image utility prediction models. These form the connection between the measures of image fidelity and image utility when one seeks to make a prediction about the utility of an image. The models may either be parametric (Section 3.3.1) or based on simulations (Section 3.3.2). The parametric models are further divided according to the method of deriving the parameters: empirical methods, analytical, or from the image. We will employ this organizational scheme as we explore traditional image quality approaches.

3.1 Image Fidelity Measures

Image fidelity measures may be broadly categorized as either sensor-derived or image-derived based on the primary user group of the measures. The sensor-derived approaches are more relevant to sensor designers since they can be readily collected from operation of the sensor. These are associated with physical image fidelity measures and primarily relate to the ability of an observer to detect objects in the image. The image-derived approaches are oriented towards those who work with images, and want to know how the sensor or a noisy transmission channel will impact the fidelity of the image. Historically, the sensor-derived approaches came first in characterizing camera systems, and image-derived approaches developed later with the advent of digital image processing.

3.1.1 Sensor-derived

These approaches to image fidelity measurement are predicated on physical parameters of the sensor. Some of these parameters may be obtained independently of the scene being imaged,

whereas others depend on scene truth to be derived. This distinction of how the parameters are obtained defines the dichotomy of our sensor-derived approaches.

3.1.1.1 Scene-dependent

These measures of image fidelity are derived by having some reference objects or ground truth in the scene of which the image is made.

Resolution

Since the 1940s, spatial resolution continues to be one of the primary means by which we express end-to-end camera system performance. This measure is attributable to Lord Rayleigh, who introduced the term resolving power in the 1880s to describe the quality of optical devices and components (Biberman, 1973). In order to quantify resolution, he defined two equally luminous point sources as “just resolved” when the center of the Airy pattern of the first fell into the first dark ring of the diffraction pattern of the second. Figure 3.2 depicts the images of two point sources, A and B, being resolved according to the Rayleigh

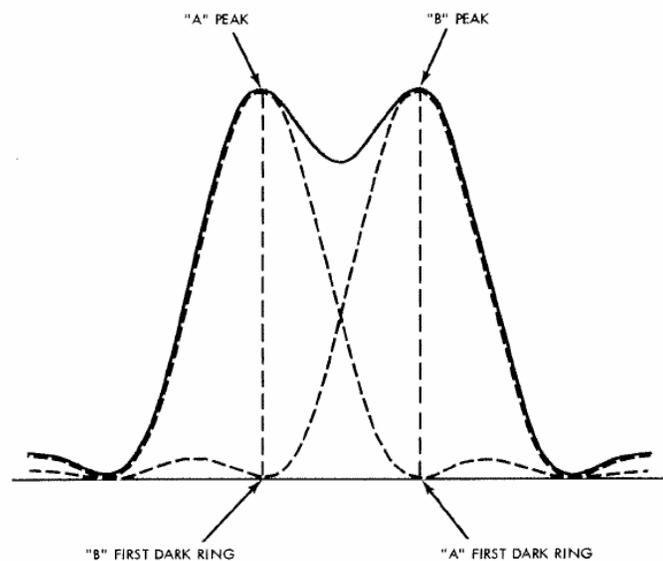


Figure 3.2: Rayleigh Criterion of resolution (from Biberman, 1973)

Criterion. Resolving power can also be measured in the laboratory by acquiring images of standard test targets and expressed as the number of line pairs per unit distance.

Resolution encompasses all imaging chain components, and has an intuitively clear meaning as a measure of the size of the smallest objects or finest detail that can be seen in an image. It is a pervasive standard, and Brock in 1967 described resolution as, “...the universal

standard of image quality without which photographic engineers cannot talk to each other, to their customers, or to the users of other imaging devices.” (Riehl and Maver, 1996) However, as an image fidelity measure, resolution has some limitations. First, it varies with target contrast so that the greater the contrast, the better the measured resolution. Second, it is not as appropriate for electro-optical systems as for photographic systems, which are described in terms of ground sampled distance (GSD). The GSD is the pixel pitch dimension projected to the object plane and indicates the energy footprint that will be sampled. It is a rough indicator of the size of objects that can be resolved by the imaging system. For a nadir-viewing sensor, the GSD in one dimension is defined as:

$$\text{GSD} = \frac{p}{f} H \quad (3.1)$$

where H is the distance from sensor to scene, p is the detector sampling pitch dimension, and f is the focal length. Figure 3.3 shows the concept of GSD for a linear detector array as well

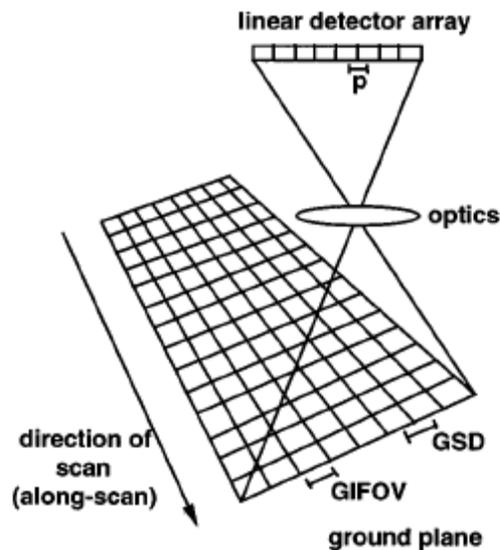


Figure 3.3: Ground sampled distance (from Fiete, 1999)

as the ground instantaneous field of view (GIFOV). The GIFOV is the linear extent of each detector projected onto the ground, and is equal to GSD if the fill factor for each detector element is 100%.

Sharpness

Acutance is a measure of the sharpness of an edge and is expressed as the mean squared gradient of the luminous flux or density with distance from the edge for photographic systems (Biberman, 1973). Applying linear systems theory, the point spread function represents the response of an imaging system to a point source and completely characterizes the system. Assuming that the imaging system is linear and shift invariant, the output signal, $g(x,y)$, where x and y are spatial coordinates, is described by the convolution in the spatial domain of the input with the point spread function $h(x,y)$:

$$g(x, y) = f(x, y) ** h(x, y) = \int \int_{-\infty}^{+\infty} f(\alpha, \beta) h(x - \alpha, y - \beta) d\alpha d\beta \quad (3.2)$$

The point spread function (psf) is difficult to measure experimentally, so we think of it in terms of a line spread function (lsf), which is a slice through the origin of the radially symmetric 2-D point spread function. The line spread function, $l(x)$, is the point spread function that has been integrated over the y direction:

$$l(x) = \int_{-\infty}^{+\infty} h(x, y) dy \quad (3.3)$$

Sharpness may be measured using a well-defined edge such as a knife edge to form an image of a step function. The sharp transition associated with a step function in the x direction will be softened so that it transitions over a finite x from the high to the low exposure value. Figure 3.4 shows the input knife edge and the resulting image to illustrate this point.

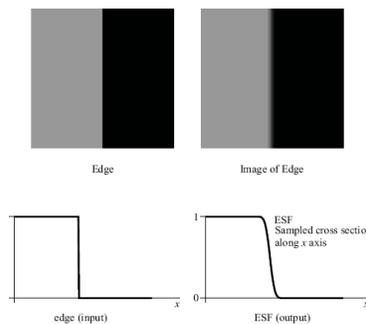


Figure 3.4: Edge spread function (from Schott, 2007)

In other words, the edge appears smeared in the image and the image appears in an area that was not exposed. It is the edge spread function that we can measure most readily, since it is the sampled cross section of the knife edge image, so starting with this quantity; we take the

first derivative to attain the line spread function. The Fourier transform of the line spread function is taken to obtain the optical transfer function, the magnitude of which is the modulation transfer function (MTF). The MTF is used to calculate the normalized edge response (ER):

$$ER_x(d) = 0.5 + \frac{1}{\pi} \int_0^{u_c} \frac{MTF_x(u)}{u} \sin(2\pi ud) du \quad (3.4)$$

where d is the response position from the center of a horizontal pixel, MTF_x is the system MTF in the x direction, u_c is the normalized optics cutoff spatial frequency, and u is the spatial frequency variable associated with the x direction. For electro-optical systems, the analog to acutance is the relative edge response (RER) and is a measure of the slope of a normalized edge response within 0.5 pixels of the edge for a noise-free signal as shown in Figure 3.5 (Leachtenauer and Driggers, 2001).

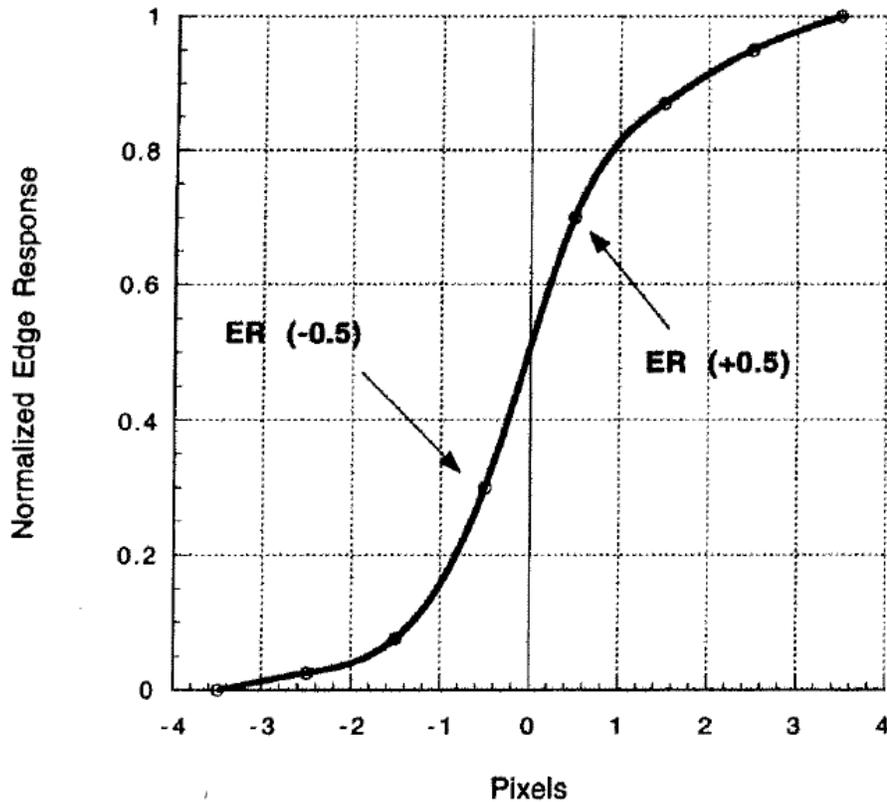


Figure 3.5: Relative edge response (from Leachtenauer and Driggers, 2001)

Contrast

Contrast is a measure of an average energy difference. The ability of a system to preserve the contrast in the scene is a measure of fidelity. For photographic processes, contrast is a function of the photographic transfer process that describes the relationship between energy incident on the camera lens and the film transmittance. The density, D , is inversely related to the transmittance and is related to the exposure, H , by the well-known H and D curve. The slope of the curve is called gamma and is a measure of relative contrast and is defined as:

$$\gamma = \frac{\Delta D}{\Delta \log H} \tag{3.5}$$

For a digital system, the analogous measure is the dynamic range, or the energy range over which the system response monotonically changes. When the energy level exceeds the capability of the detector to provide an increased response, saturation is said to have occurred, resulting in an inability to separate different energy levels even when they exceed the system signal-to-noise ratio (SNR) threshold. Dynamic range is expressed as the ratio of maximum signal to system noise (Leachtenauer and Driggers, 2001).

The contrast transfer of an imaging system is the ratio of the contrast of the input and output square-wave target images. The contrast, C , is the square wave modulation between the maximum (max or white) and minimum (min or black) brightness values of the image:

$$C = \frac{\max - \min}{\max + \min} \tag{3.6}$$

Figure 3.6 depicts how the contrast transfer of the system varies as a function of the spatial

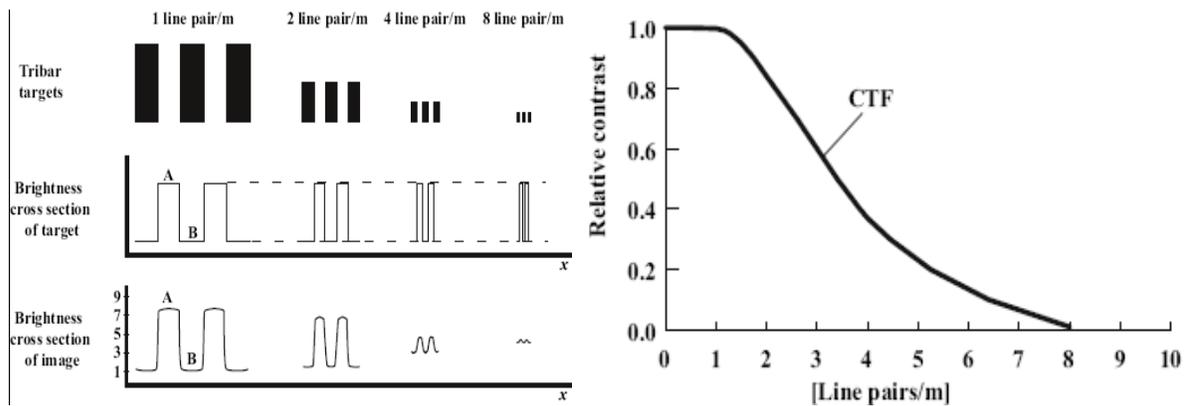


Figure 3.6: Contrast transfer function (from Schott, 2007)

frequency of the tri-bar target. Quantities A and B correspond to the maximum and minimum brightness values, respectively. The resulting contrast transfer function (CTF) characterizes image fidelity over all spatial frequencies in the image (Schott, 2007).

In electro-optical systems, a primary performance parameter is the minimum resolvable contrast (MRC) and is used to describe the sensor's sensitivity as a function of bar pattern resolution (spatial frequency). The MRC describes imager performance and is dependent on the sensitivity and resolution characteristics of the system, which in turn are dependent on a large number of parameters such as focal length, entrance pupil diameter, detector size, etc. It is measured by varying the contrast of a four bar target until the bars are just resolvable by a person using the sensor. The differential contrast is plotted for each frequency. Because the noise and background currents depend on the amount of light falling on the detector, the MRC is a family of curves that depends on light level (Driggers, et. al., 1998).

Modulation Transfer Function (MTF)

The MTF is the same as the CTF except it uses sinusoidal rather than square wave input functions to quantitatively describe the spatial frequency transfer characteristic of an imaging system. It describes how the imaging system changes the spatial frequency content of the object as represented in the image or the efficiency of image detail acquisition from object to recorded image. It is mathematically stated as the ratio at any given spatial frequency, u , of the fractional decrease of amplitude of the constituent sine waves of a function as it passes through the imaging system:

$$MTF(u) = \frac{m_{out}}{m_{in}} = \frac{\frac{g_{max} - g_{min}}{g_{max} + g_{min}}}{\frac{f_{max} - f_{min}}{f_{max} + f_{min}}} \quad (3.7)$$

In the equation, m represents the modulation of the input and output signals, assuming that the input is a nonnegative sinusoid with additive bias b of the form:

$$f(x, y) = b + a \cos(2\pi u(x, y) + \theta) \quad (3.8)$$

where u is the spatial frequency, x and y are the spatial variables, and θ is the initial phase shift of the sinusoid. The f_{max} is the sum of amplitude and bias, $a+b$ and f_{min} is their difference $b-a$, and g_{max} and g_{min} are the corresponding output signal amplitude and bias sums and differences. The modulation, m , of a nonnegative sinusoid is the ratio of the amplitude a

to the bias b , and therefore provides a measure of the relative brightness of the maxima and minima of the function.

It is a characteristic of a linear system that it is completely specified by its point spread function $h(x,y)$ in the spatial domain or its transfer function $H(u,v)$ in the spatial frequency domain (using the spatial frequency variables, u and v , in units of [cycles/unit spatial length]). The magnitude of the transfer function is the scale factor that must be applied to each sinusoidal component of $f(x,y)$ and is the MTF of the system. Using Fourier analysis, the 2-D convolution that describes the input-output relationship of the imaging system is:

$$G(u, v) = F(u, v)H(u, v) = \int \int_{-\infty}^{+\infty} f(x, y) * h(x, y) e^{-i2\pi(ux+vy)} dx dy \quad (3.9)$$

MTF is a useful fidelity metric because it relates contrast and resolution and forms the basis for defining the relative edge response. The MTF indicates how well a sinusoidally varying brightness of a given frequency is reproduced by the imaging system and is indicative of the contrast throughput of the system. The system MTF is the cascaded value of all component MTFs which typically include the optics, environmental wavefront error, atmospheric dispersion and turbulence, detector, aperture, clocking, carrier diffusion, charge transfer efficiency, and modulation transfer function compensation.

MTF is an important quality measure of displayed images for two reasons: higher modulation transfer values lead to higher luminancy contrast of the displayed pattern and higher cutoff frequency allows for higher spatial frequency patterns to be displayed. This implies higher contrast and sharper detail in displayed images, so one way to characterize the fidelity of a display system is the area under its MTF curve (Janssen, 2001). The area between the MTF curve and the detection threshold function, M_t , is called MTFA and is a summary measure of the quality of a photograph. MTFA is the shaded area in Figure 3.7.

The detection threshold function is defined by a combination of the modulation threshold of the human visual system (HVS) and the film grain. Other measures are related to MTFA and incorporate the contrast sensitivity function (CSF) of the HVS. The CSF expresses the contrast sensitivity of the eye as a function of angular frequency. A measure of the nonlinear HVS response to a single frequency is called the contrast threshold function and is measured over the visible radial spatial frequencies from 0 to 60 cycles/degree. The

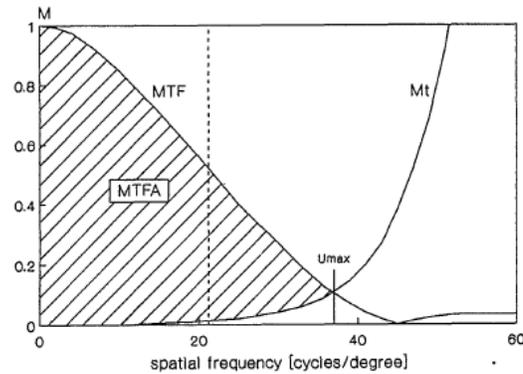


Figure 3.7: MTF A (from Barten, 1990)

contrast threshold function is the minimum amplitude necessary to just detect a sine wave of a given angular spatial frequency. Inverting the contrast threshold function gives a frequency response, the CSF, which is a linear spatially invariant approximation to the HVS (Barten, 1990). Granger and Cupery (1972) combined the visual system MTF and the image system MTF to generate a combined image quality metric for visual assessment called the subjective quality factor (SQF). Barten (1990) introduced the square-root integral (SQRI) to describe the effect of the display device on image quality. SQRI uses the eye's contrast sensitivity function as a weighting function to the MTF.

Noise Measures

Image noise manifests itself as the random fluctuations in recorded luminance that interfere with the ability to detect detail. The noise is the variance of the observations about a mean value relative to a defined signal. In photographic systems, film grain is the dominant noise source, with the random distribution of silver grains in the photographic emulsion producing random fluctuations in film density. This noise is called granularity and it correlates with the subjective appearance of graininess of the image (Leachtenauer and Driggers, 2001). In digital imaging systems, the noise is usually characterized by the variation in the instantaneous signal level, which is typically specified for a scene of a known signal flux level and imaging conditions. The noise, N , is described by the root mean square deviation in the instantaneous signal, S_i , at a fixed input:

$$N = \left(\frac{\sum_{i=1}^n (S_i - S_{avg})^2}{n} \right)^{\frac{1}{2}} \quad (3.10)$$

S_{avg} is the mean signal level and n is the number of samples (Schott, 2007). Noise can also be expressed in radiometric units rather than signal output units. The noise equivalent power (NEP) is the amount of incremental flux at a particular wavelength required to change the signal level by an amount equal to the noise or equivalently, produce a SNR of one:

$$NEP(\lambda) = \frac{N}{R(\lambda)} \quad (3.11)$$

where $R(\lambda)$ is the spectral responsivity of the sensor. Flux variations must be above the NEP in order to be detected, meaning that a radiometrically sensitive sensor (or one which would have good radiometric fidelity) would have a small NEP. The noise equivalent change in reflectance, $NE\Delta\rho$, represents the reflectance difference between two pixels that is equivalent to the noise standard deviation.

The signal-to-noise ratio (SNR) is commonly used to express the sensitivity of the sensor. The signal is defined as the measured output of a detector and the noise is the random signal variation due to detector, photon, temperature, and amplifier noise terms. For images with large signal, the primary noise contributor is photon noise, which arises from the random fluctuations in the arrival rate of photons, and is Poisson distributed. Most SNR metrics compare a known reflectance mean scene signal, S_{avg} , with the standard deviation of the noise, N .

$$SNR = \frac{S_{avg}}{N} \quad (3.12)$$

Fiete and Tantalo (2001) point out that there are many ways to define the SNR, including describing the signal in terms of the reflectance difference of a target and its background, calculating the SNR as a function of spatial frequencies, and using signal and noise calculations made from the image data.

Another noise-related measure is detective quantum efficiency (DQE), first proposed by Rose in 1946 in the context of radiation detectors. The motivation for this metric was to compare the performance of different detector types. It is based on the principle of referring the output fluctuations through the system input/output operating characteristics and

expressing them in terms of equivalent input fluctuations. The DQE is expressed in terms of the ability of the detector to discern a signal against a background of ambient radiation. The DQE compares the quantum fluctuation of the ambient radiation manifested in the randomness of the detector response with the randomness of the detector itself (Hailstone, 1995). The DQE can be viewed as the ratio of squared input and output SNRs in order to describe the transfer of SNR in the image formation process.

$$DQE = \frac{SNR_{out}^2}{SNR_{in}^2} \quad (3.13)$$

SNR_{in} is the ratio of signal mean and standard deviation at the aperture of the imaging system, and SNR_{out} is the ratio of signal mean to standard deviation measured at the output. The DQE will be less than or equal to one, since the output SNR will always be degraded with respect to input SNR due to sensor noise, with equality occurring in the ideal case of no additive noise (only photon noise-limited performance). In another sense, DQE compares the noise level of a real radiation detector to that of an ideal one working at the same exposure level, so that the number of quanta through a real detector required to produce an equivalent output in an ideal detector can also be used to express system performance. This is known as the noise equivalent quanta (NEQ) (Dainty and Shaw, 1974).

3.1.1.2 Scene-independent

These measures of image fidelity do not require reference objects or ground truth in the scene being imaged, as they are derived directly from sensor parameters.

Scale

Perhaps the most intuitive image fidelity metric is scale. Scale is defined as $s = f/H$, where f is the camera focal length and H is the camera distance from the object. Scale defines the size relationship between ground and image distance and is a measure of what can be seen in the image as illustrated in Figure 3.8. In many applications, image quality is a function of object-image scale. In aerial surveying, a scene imaged at a scale of 1000:1 would be visually assessed as having higher fidelity for interpretation of ground objects than the same scene imaged at 10,000:1 scale assuming constant imaging system performance (Nill and Bouzas, 1992).

But scale is misleading as a figure of merit. When aerial cameras and films were generally of the same quality, image scale served as an adequate predictor of image

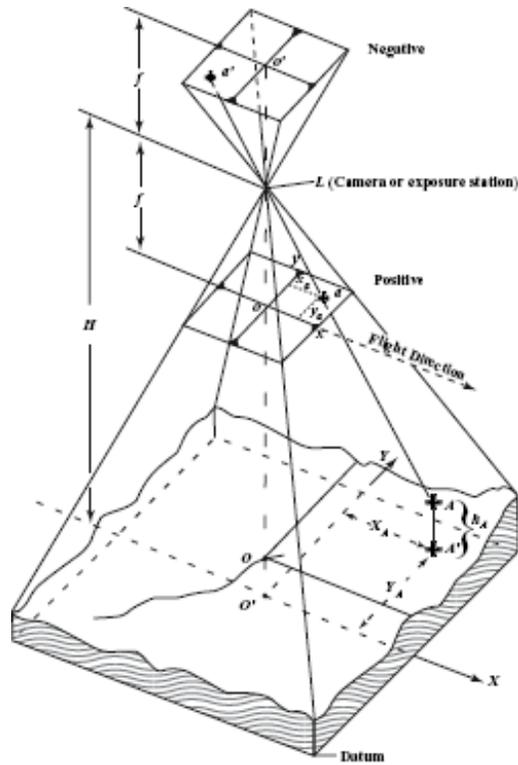


Figure 3.8: Image scale (from Schott, 2007)

interpretability or information content. All images of the same scale were of roughly the same quality and interpretability. As cameras and films diversified and improved in terms of quality, scale ceased to be a good predictor (Leachtenauer, et. al., 1997).

$$\lambda FN/p$$

$\lambda FN/p$ is the ratio of the digital sensor spatial sampling frequency to the optical bandpass limit of an incoherent diffraction-limited optical system, and is a fundamental system design parameter (Fiete, 1999). FN is the f/number of the system, λ is the mean wavelength, and p is the detector sampling pitch. The interaction between the detector sampling and performance of the optics plays an important role in determining the final image quality, as detector sampling pitch limits the highest spatial frequency that can be sampled without aliasing. The diffraction resolution of a diffraction limited incoherent imaging system is determined by the optics and is described by the point spread function (PSF). The optics PSF width projected through the imaging system onto the ground is taken to be λFN and called the ground spot size (GSS) as illustrated in Figure 3.9. In terms of spatial frequency, the metric $\lambda FN/p$ is the

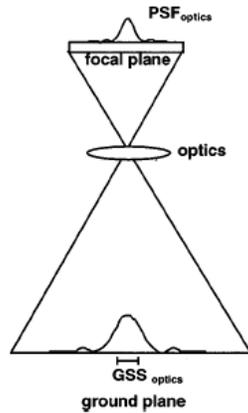


Figure 3.9: Ground spot size (from Fiete, 1999)

ratio of the sampling frequency to the optical bandpass limit of the optical system:

$$\frac{\text{detector sampling frequency}}{\text{optical bandpass limit}} = \frac{\frac{1}{p}}{\frac{1}{\lambda(f/D)}} = \frac{\frac{1}{p}}{\frac{1}{\lambda FN}} = \frac{\lambda FN}{p} \quad (3.14)$$

In the spatial domain, $\lambda FN/p$ can be interpreted as a measure of how finely the detector samples the diffraction limited optics PSF or how finely the GRD samples the ground scene with respect to GSS. Figure 3.10 depicts this concept of sampling.

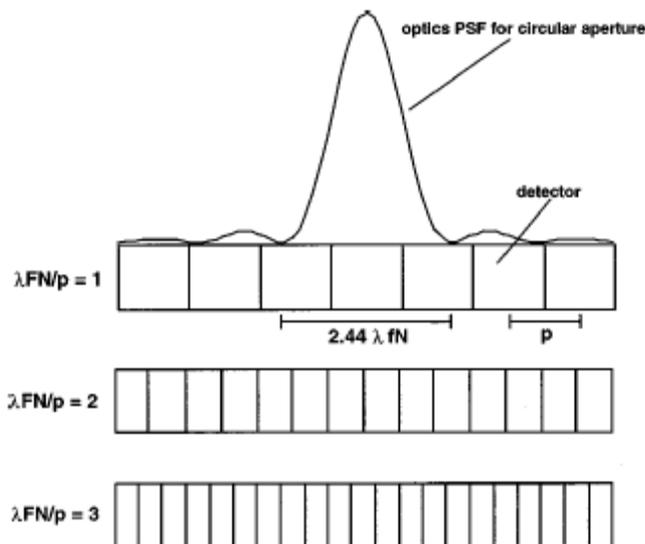


Figure 3.10: $\lambda FN/p$ detector sampling in the spatial domain (from Fiete, 1999)

As a design parameter, this measure is very useful. When $\lambda FN/p = 2$, the diffraction resolution and detector resolution are equal because the optics MTF falls to zero at the Nyquist frequency. $\lambda FN/p < 2$ implies that the spatial resolution is limited by the detector, and $\lambda FN/p > 2$ implies that spatial resolution is limited by the optics diffraction (Fiete, 1999). Thus, a $\lambda FN/p=2$ design will optimize the system resolution by maximizing both the detector and diffraction resolution of the system.

3.1.2 Image-derived

These approaches to image fidelity measurement only rely on the image and not on any sensor parameters. They are divided into two categories: those derived by using a single image and those that use more than one version of the image. We will call these two distinctions the single image and the multiple image subcategories in the discussion that follows.

3.1.2.1 Single image

Power Spectral Density

Nil and Bouzas (1992) demonstrate this approach for visible panchromatic digital aerial images for the application area of man-made object detection, recognition, and identification. The approach is built on the assumption that the imaging system input scene power spectrum is invariant from scene to scene. The 2-D power spectrum of the image, $H(u,v)$ is obtained and normalized to compensate for the effect of image-to-image brightness variations on power spectra magnitudes. The normalized image power spectrum is:

$$P(u,v) = \frac{|H(u,v)|^2}{\mu^2 M^2} \quad (3.15)$$

where μ is the average gray level of the image and M^2 is the total number of image pixels. The modulation transfer function of the HVS, directional differences (scale), and a filter to account for imaging system noise can be incorporated to obtain an objective image quality measure. This measure has been shown to have good correlation with the visual quality of aerial images assessed for their interpretability by image analysts (Nil and Bouzas, 1992).

Impairment

Janssen (2001) describes an approach based on defining image distortions that can be directly related to one or more visually distinguishable effects on image appearance. This approach does not need to compare an impaired image to its original version. Impairment strength is calculated directly from the image by estimating the relevant physical parameters such as blur kernel spread or noise variance, and relating these parameters to perceptual attributes such as unsharpness and noisiness using psychophysically established relations. The psychophysical relations between physical parameters regarding the image distortion and the impairment strength are the outcome of the impairment approach. In using this approach, there is ambiguity as to whether a hypothetical unimpaired version of an image is really the version with the highest image quality.

3.1.2.2 Multiple image

These approaches operate by comparing distorted and original versions of an image. The comparison could be between an original uncorrupted version of the image and the received or reconstructed version following operations such as compression or transmission. It could also be a comparison between the original image and one subjected to intentional degradation such as additive noise, blurring, or spatially localized brightness modifications. We will examine statistical, HVS-based, and subjective measures of image-derived image fidelity.

Statistical

These measures originate from signal processing applications where original (called the ‘object’) and distorted (called the ‘image’) versions of a signal need to be compared. Image comparisons are made on a pixel-by-pixel basis and are attractive due to their ease of calculation and independence from intended use or viewer of the image. This group of measures is based on the assumption that the goal in imaging is to reproduce a likeness of the object so that the best imaging system is the one that gives the smallest discrepancy between object and image. A problem with statistical measures is their sensitivity to different image properties. Small image rotation or magnification changes, distortions, and gray-scale errors produce large discrepancies between image and object. By the same token, these fidelity measures may be completely insensitive to small details that we seek to capture in the image (Barrett and Myers, 2004).

Avcibas, Sankur, and Sayood (2002) evaluate image quality metrics from the digital image processing literature and categorize them into four categories according to how the distortion is calculated. The first category of metrics is called the pixel-difference measures. These calculate the distortion between two images on the basis of their pixel-wise differences or certain moments of the difference image. The L_γ norm of the error between two images can be calculated by taking the Minkowski average of pixel differences (between the original image, C_k , and its degraded version, \hat{C}_k) spatially (using i and j as the spatial coordinate indices over an image with N^2 pixels) and then over the spectral dimension (indexed by k in an image assumed to consist of K spectral bands):

$$L_\gamma = \frac{1}{K} \sum_{k=1}^K \left\{ \frac{1}{N^2} \sum_{i,j=0}^{N-1} |C_k(i,j) - \hat{C}_k(i,j)|^\gamma \right\}^{\frac{1}{\gamma}} \quad (3.16)$$

The most popular pixel-difference metric is the mean squared error (MSE), which is a special case of the L_γ norm where $\gamma = 2$. The use of MSE is common in digital imaging to specify how well an imaging system is functioning. Since MSE is symmetric, it cannot be used to predict which of the two image versions is better. Related to MSE is the peak signal to noise ratio (PSNR), which is a normalized version of MSE obtained by dividing the square of the image dynamic range by the MSE. PSNR is expressed in decibels:

$$PSNR = 10 \log_{10} \frac{(2^n - 1)^2}{MSE} \quad (3.17)$$

where n is the number of bits describing the count dynamic range of the image. PSNR is the ratio of peak signal power to average noise power. Raw error measures such as MSE work best when the distortion is due to additive noise contamination, but they do not correspond to all aspects of the observers' visual perception of errors nor do they correctly reflect structural coding artifacts (Avcibas, Sankur, and Sayood, 2002). Other metrics in the pixel-difference approach include modifications to the Minkowski infinity metric, a difference over a neighborhood to penalize spatial displacements in a graduated way in addition to gray level differences, and a multiresolution distance measure to enable the comparison of images at other than full image resolution.

The second category of statistical measures is termed the correlation-based measures. These quantify the similarity between two images based on structural content, normalized

cross-correlation, Czenakowski distances, angles between image pixel vectors, or the combined angle-magnitude difference(Avcibas, Sankur, and Sayood, 2002).

The third category is based on the idea that edges will contain the most information in the image, and task performance in various applications depends on the quality of edges and other two dimensional features such as corners. Edge maps are obtained from the original uncompressed image and the compressed image using an edge-detector algorithm (Canny detector). The edge-based measures consider the accuracy of the edge location, missing or spurious edge elements, and the consistency (stability) of the edge across different scales (Avcibas, Sankur, and Sayood, 2002).

The fourth category operates on the complex Fourier transform of the original and compressed images. The spectral distortion measures examine the MSE of the image phases, magnitudes, and weighted spectral distortions. An image may also be divided into blocks and blockwise spectral distortions can be calculated. Minkowski averaging of block spectral distortions may help deal with the localized nature of the distortion and rank order averages such as median block distortion, maximum block distortion, and average block distortion may be considered (Avcibas, Sankur, and Sayood 2002).

Wang and Bovik (2002) propose a universal objective image quality index that is designed by modeling any image distortion as a combination of correlation loss, luminance distortion, and contrast distortion. This statistical measure of image fidelity is defined as:

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2 + (\bar{y})^2]} \quad (3.18)$$

in which \bar{x} is the original image mean, \bar{y} is the test image mean, σ_x^2 is the variance of the original image, σ_y^2 is the test image variance, and σ_{xy} is the cross-correlation between the original and test images. Statistical measures that use image differencing such as any Minkowski-based metrics are not an appropriate mathematical form for image quality evaluation since image differencing does not capture an estimate of the correlation between the two images. This is because information loss occurring during the image degradation process is best measured using correlation rather than differences, which are sensitive to the energy of the errors. As a demonstration of the efficacy of the Q metric as opposed to MSE in responding to image degradations, consider the following four images in Figure 3.11.



Figure 3.11: Four image degradations (from Wang and Bovik, 2002)

Degradation (a) corresponds to a mean shift, (b) is contrast stretched, (c) is blurred, and (d) has been JPEG compressed. The MSE between the original and degraded images for (a), (b), and (c) is computed as 225 and (d) is 215, whereas the Q value for (a) is 0.9894, (b) is 0.9372, (c) is 0.3461, and (d) is 0.2876 (Wang and Bovik, 2002).

Human Visual System (HVS)

The HVS category of image-derived image fidelity measures is based on the assumption that the goal of image processing is to create an image that is perceptually equivalent to the original. Another assumption is that inclusion of human visual processing characteristics will help assess the perceived difference between a reproduced image and the original (Janssen, 2001). Image degradation is measured in units of just-noticeable differences (JND) between the original image and degraded version. One JND unit corresponds to a fixed probability that an observer will detect the difference between two images or image regions. The JND approach is based on the threshold theory of vision in which signal detection occurs when a signal's perceptibility exceeds an observer's threshold (Barrett and Myers, 2004).

Quality measures based on linear HVS models assess image quality by computing an error image as the difference between the original and distorted image and then weighting the error image by the frequency response of the HVS, which is represented with a lowpass

contrast sensitivity function (CSF) or some other component of a HVS model. Components of HVS models are drawn from physiologically and psychophysically established mechanisms of human visual processing that include: optical blurring due to imperfect optics of the eye, luminance adaptation to account for variation in visual sensitivity as a function of light level, contrast sensitivity function to account for variation in visual sensitivity as a function of angular frequency, decomposition into multiple frequency bands to account for spatial frequency selectivity of the HVS, orientation-selective filters, and luminance and contrast masking to account for visual sensitivity as a function of background structure (Damera-Venkata, et. al., 2000). Regardless of the complexity of the model employed, the resulting error image characterizes the regions in the test image that are visually different from the original image, quantified on a per pixel basis.

The JND approach is objective and correlates with subjective assessments of image quality. Like the statistical approaches, the HVS-based approaches quantify some form of image discrepancy, and whereas the statistical measures equally weight image differences, the HVS approaches weight image differences according to their predicted manifestation at the output of the visual system. Barrett and Myers (2004) point out that in order to calculate perceptual image differences, the HVS approach requires an image pair in which each image has the same noise realization—a situation that is only achievable with simulated images.

Subjective

Nil and Bouzas (1992) claim that the most meaningful image quality measures are based on visual assessments of images, because the human observer is the ultimate viewer of images. However, obtaining a large enough sample size of visual image quality assessments to overcome inherent observer variability is very time consuming and resource intensive. The picture quality scale (PQS) attempts to overcome this challenge. PQS is based on the perception properties of human vision and a set of partial distortion measures are defined as the function of error calculated between the original and decoded pictures (Miyahara, et. al., 1998). It pools the effects of noise, luminance coding mistakes, end of block disturbances, correlated errors, and problems near high contrast changes. It models the effects of these errors using five factors as coefficients in the equation obtained by multiple linear regressions against human mean opinion score (Dosselman and Yang, 2005). PQS was

originally proposed for image quality estimation of monochromatic image coding, but Sun and Fairchild (2003) extend the idea to multispectral images.

3.2 Image Utility Measures

In the broadest definition, image utility conveys the ability of the image to provide the information required by the image observer. A number of factors other than image fidelity and the image acquisition processes affect information extraction performance. The target and its environment, individual observer differences, and the information extraction algorithm (if the image is being machine-processed) will all affect the image utility, as Figure 1.2 illustrated. We can describe image utility in terms of tasks that can be accomplished or information that can be extracted. If we take a task performance approach, the theory of signal detection is used to describe task performance in terms of mathematically tractable probabilities. If we take an information extraction approach, information theory provides a viewpoint that has a vocabulary designed to quantify the information passing through the system. This section will investigate the task performance and information theory approaches to image utility measures.

3.2.1 Task Performance Measures

From a task performance perspective, image utility is defined as the ability to correctly perform a well-defined task. To a radiologist, an x-ray image is considered to be of good utility when a well-defined task such as the detection of a certain lesion can be performed correctly. In this medical application example, the radiologist is the observer and the signal to be detected is an abnormality in a medical image. Task performance measures are usually derived during experiments since often no adequate models exist that can reliably predict performance. Tasks are defined along a hierarchy ranging from detection, to recognition (classification), to identification. Task performance is not easy to measure, since an unbiased measurement of performance can only be obtained once from a given image of a scene. Once an observer has viewed a scene, learning has taken place, and performance on successive scenes is based partly on image quality and partly on the learning. Thus, images of different scenes are needed to accurately measure performance. It is for these reasons that assessing

task performance is challenging and has led to different methodologies for characterizing performance depending on the particular application area. Leachtenauer and Driggers (2001) establish a structure for considering task-based approaches based on whether the assessed image utility is measured or estimated. These two groupings are called direct performance measures and performance estimates, and we employ these subdivisions in our discussion of the task performance view of image utility.

3.2.1.1 Direct Performance Measures

Barrett and Myers (2004) contend that for an image-assessment method to be acceptable, it must objectively quantify the usefulness of images for performing a given task. The task performance of an imaging system is defined by how well inferences about an underlying scene can be made using an image. Because task performance must be assessed on the basis of average performance of some inference task by an observer or decision-maker and many of the tasks involve detecting objects in the scene, concepts from statistical decision theory and signal detection theory find good application. Statistical inference tasks are divided generically into classification and parameter estimation tasks.

In classification tasks, there are a finite number of possible outcomes. The tasks are referred to in various disciplines as pattern recognition, signal detection, discrimination, discriminant analysis, differential diagnosis, segmentation, and hypothesis testing (Barrett and Myers, 2004). An observer makes a decision about the class membership of the observed data by computing some test statistic, $\lambda(\mathbf{g})$, and comparing it to a decision threshold. The observer usually has some prior information about the possible objects being imaged, the distortions caused by image acquisition, and the sources of randomness in the data. The conditional density function, $p(\mathbf{g}|\mathbf{f})$, called the likelihood function, defines how the data are distributed given a particular underlying object or hypothesis. A classification problem is categorized according to the number of hypotheses to be distinguished, the nature of the hypotheses, the structure of the data, and the statistics of the signal and noise. Classification tasks with two underlying hypotheses from which the data might be drawn are known as binary decision problems. Class 1 is the signal (plus noise) present case and class 2 is the signal absent (noise only) case. Figure 3.12 depicts the conditional density functions of the

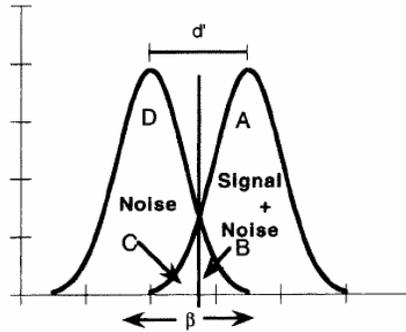


Figure 3.12: Probability distributions of observations in a binary decision task (from Leachtenauer and Driggers, 2001)

observations of the two hypotheses. The observer's decision is based on the data, which constitute only incomplete clues as to the underlying object since they were obtained through an imaging system contaminated by noise. The observer must make a decision based on whether the observations fall to the right or left of his decision threshold, β . The amount of overlap between the distributions relative to the threshold will determine the amount and type of errors that occur. The area to the right of the threshold under the signal and noise distribution constitutes the *probability of detection* (labeled as A in Figure 3.12), while the area to the right of the threshold under the noise only distribution is the *probability of false alarm* (labeled as B in Figure 3.12).

A more useful tool for capturing the detection task performance is the receiver operating characteristic (ROC) curve. It summarizes the difficulty of the task and the utility of the data for enabling the observer to perform the specified task. The ROC curve shows all possible probability of detection (PD) and false alarm (PFA) outcomes for a given signal and noise distribution as the decision criterion is varied. Figure 3.13 shows two ROC curves

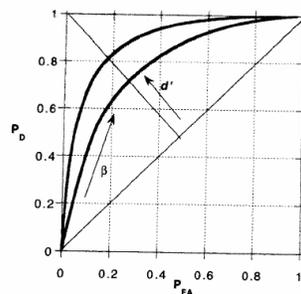


Figure 3.13: Receiver operating characteristic curve (from Leachtenauer and Driggers, 2001)

along with d' , which is called the detectability statistic and was also annotated on Figure 3.12 as a measure of separability between the two normally-distributed conditional density functions. It may be calculated using:

$$d' = \frac{\bar{i}_2 - \bar{i}_1}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}} \quad (3.19)$$

in which \bar{i}_2 and σ_2^2 are the noise distribution mean and variance, and \bar{i}_1 and σ_1^2 are the signal plus noise distribution mean and variance. The area under the ROC curve is also a measure of task performance, and ranges from 0.5 for a worthless (no better than guessing) test to 1.0 for a perfect one. The area under the ROC curve is a statement about the average PD over all PFAs. For classification tasks performed by a human observer, psychophysical studies and ROC curve analyses provide a reproducible, quantitative measure of image utility. However, such studies are time consuming since the threshold must be systematically varied over the range of the ROC curve and large numbers of images are required (Barrett, et. al., 1993).

A related classification task performance utility measure is the probability of discrimination. This performance metric is used with target acquisition sensor tasks, which involve visual detection, recognition, and identification of targets in broadband optical or infrared imagery. Overall performance is described using Johnson's criterion to obtain a probability of discrimination using the sensor's minimum resolvable temperature (MRT) or minimum resolvable contrast (MRC) parameter for infrared and electro-optical sensors, respectively (Driggers, et. al., 1998). Johnson's criterion will be discussed further in the image utility prediction models section (Section 3.3).

For completeness, we mention that parameter estimation tasks represent the limit of hypothesis testing when the number of hypotheses becomes infinite. Estimation begins with the assumption that we have data \mathbf{g} from some known probability law $p(\mathbf{g}|\theta)$, called the likelihood function, and the task is to estimate the scalar parameter θ or the parameter vector $\boldsymbol{\theta}$. These parameters can take on any value in a specified range.

3.2.1.2 Performance Estimate Measures

These measures represent observer estimates of information extraction performance. They are less resource intensive than directly measuring performance and less prone to bias effects from observer learning than direct performance measures involving human observers.

Although performance estimates may appear to be more subjective, there is no evidence to support the relative validity of direct versus estimate measures (Leachtenauer and Driggers, 2001).

National Imagery Interpretability Rating Scale (NIIRS)

NIIRS came about as a response to the inability of simple measures of physical image quality such as scale and resolution to successfully communicate image interpretability. NIIRS was developed in the early 1970s by a government and contractor team working for the Imagery Resolution and Reporting Standards Committee. The goal was to develop a scale that would communicate information that could be extracted from a specific image. The basis for the NIIRS concept was that imagery analysts should be able to perform more demanding interpretation tasks with higher quality imagery (Irvine, 1997). The scale was defined in terms of variations in resolution with each level of the scale representing a doubling of resolution. A set of tasks was selected at each of ten levels, where the tasks defined specific objects and levels of interpretation (detection, recognition, and identification). Each task consists of three parts: a recognition level (to detect, recognize, or identify), an object, and a qualifier. As a means of facilitating communication between image analysts, NIIRS provides a shorthand description of image interpretability (and thus utility). In the initial development process, a large sample of interpretation tasks was rated in terms of relative difficulty by a group of image analysts (Leachtenauer and Driggers, 2001). The most recent version of the visible panchromatic NIIRS was developed in 1994 and there is a civil NIIRS scale that is based on natural, agricultural, and urban/industrial interpretability tasks.

Task Satisfaction Confidence Ratings

These are observers' estimates of their ability to perform a specific information extraction task on an image. Unlike the very specific NIIRS criteria, the task may be defined as broadly as desired and any number of tasks may be addressed. The ratings attempt to capture the percent confidence that the image analyst could extract the feature of interest in the image (Leachtenauer and Driggers, 2001).

3.2.2 Information Theory Measures

The connection of Shannon's information theory to imaging systems was first proposed by E. H. Linfoot in the 1950s. Linfoot writes: "*The complex amplitude distribution over the entry*

pupil of an optical system can be regarded as an intercepted part of a message sent out by the object and the formation of an image in the focal plane of the system as a decoding of this intercepted message which presents information about the object in a more convenient form. An optical system can properly be said to be of high quality only if the amount of information contained in its images approaches the maximum possible with the prescribed aperture and receiving surface” (Biberman, 1973). Linfoot uses a simplified model of an optical system in which the source is divided into small discrete elements, each capable of a finite number of discrete brightness levels. The information content of the values of the elements is defined in terms of their unexpectedness. The information, I , carried by an object described by vector \mathbf{f} is the entropy of the set of values that an object assumes:

$$I(\mathbf{f}) = -\sum_{n=1}^N P(f_n) \log \left[\frac{1}{P(f_n)} \right] \quad (3.20)$$

The quantity f_n is the brightness of the n^{th} object element, and $P(\cdot)$ denotes probability. Using the object’s information measure as a starting point, Linfoot computes and maximizes the imaging system’s ability to transfer information. (Barrett and Myers, 2004).

O’Sullivan, et. al. (1998) depict the image formation process as an analog to the classic communications theory model, both of which are shown in Figure 3.14. The prior

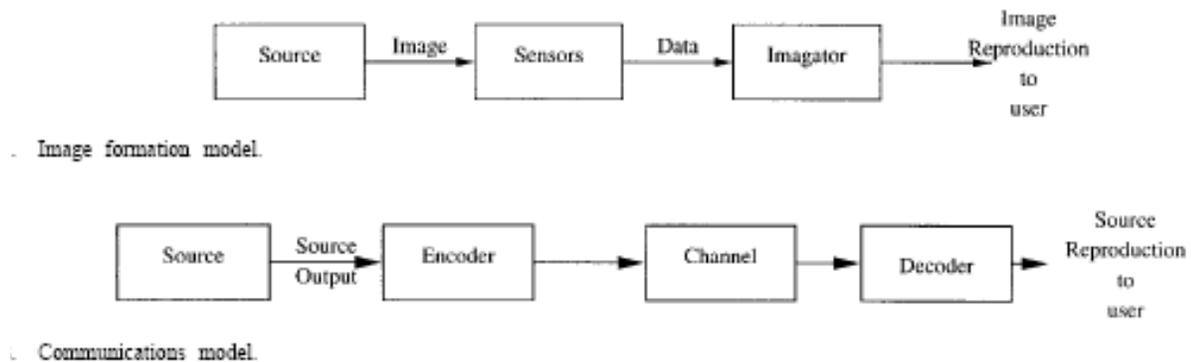


Figure 3.14: Information theory view of an imaging system (from O’Sullivan, et. al., 1998)

probability distribution on the image space (the set of all possible images from which a source selects an image) quantifies the information that is available prior to data acquisition. The sensor output data will be used to make an inference about the scene. Measurements provided by the sensors add information to this prior information, and the value of the new information depends on the imaging system’s goal. The goal of the imaging system also

determines the appropriate performance measure (such as the probability of detection, probability of false alarm, and probability of correct classification). From an information theory viewpoint, each of these measures seeks to quantify the information provided by a measurement for the specific class of problem. All of the measures, however, depend on the likelihood of the data, and they reduce the likelihood function to a form that is more appropriate to a particular problem (O’Sullivan, et. al., 1998).

Huck, et. al. (1997) present a model of the visual communications channel that accounts for three transformations: the continuous-to-discrete transformation of the continuous radiance field, $L(x,y)$, into the discrete signal, $s(x,y)$, with analog magnitudes, the analog-to-digital transformation of $s(x,y)$ into the encoded digital signal $s_e(x,y)$ quantized to k quantization levels, and the digital-to-continuous transformation of the restored digital image into the continuous observed image, $R(x,y)$. This model is shown in Figure 3.15:

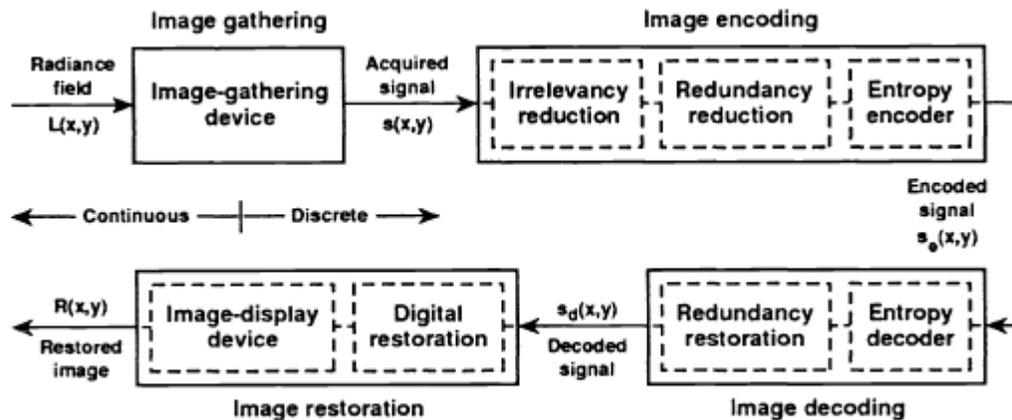


Figure 3.15: Visual communications channel (from Huck, et. al., 1992)

Because the resolution of the image-gathering device is coarser than the finest detail in the scene, the spatial frequency spectrum of the radiance field extends beyond the sampling passband of the image-gathering device. A high quality visual communication channel is one where the information rate from the scene to the observer approaches the maximum possible and the required data rate approaches the minimum possible.

The work of Huck, et. al. (1997) is based on Shannon’s concepts of rate of transmission of information in noisy channels and Wiener’s concept of minimum MSE restoration of signals corrupted by noise. Assuming a Gaussian signal, this describes the relationships that exist among information rate, theoretical minimum data rate, and

maximum-realizable fidelity. They relate information rate to the restorability of images, where image restoration is intended to produce a representation of the input rather than image reconstruction, which produces a continuous representation of the discrete output of the image gathering device. Their model discusses the quantitative assessment of visual communication in six figures of merit:

1) The information rate (H) of the image gathering system for the radiance field in its field of view. This is also the mutual information between the captured radiance field, $L(x,y)$, and the signal, $s(x,y)$, and is a measure of the amount of information in the acquired signal (represented by the entropy of the acquired signal, $E[s(x,y)]$) minus the part due to noise (represented by the conditional entropy of the signal given the captured radiance field, $E[s(x,y)|L(x,y)]$):

$$H = E[s(x,y)] - E[s(x,y)|L(x,y)] \quad (3.21)$$

H connects the performance of the visual communication channel to the quality with which images must be restored and it depends on the design of the image-gathering device.

2) The theoretical minimum data rate, E_{min} , is the entropy of completely decorrelated data associated with H . This is the mutual information between discrete signal, $s(x,y)$, and digital signal, $s_e(x,y)$ and it is the lower bound on the data rate that is associated with H .

3) The information efficiency, H/E_{min} , of the completely decorrelated data, ties the performance of the visual communication channel to the quantity of data that must be transmitted and stored.

4) The maximum realizable fidelity, F , of the digital image that can be restored from the received information, unconstrained by the display medium. It is a measure of the similarity between radiance field $L(x,y)$ and digital image. It is an MSE measure and depends on the mean spatial detail of the scene.

5) The information rate, H_0 , of the observed image that the image restoration system produces from the received information on an image display medium.

6) The maximum realizable fidelity, F_0 , of the observed image with information rate, H_0 , that can be restored in continuous form on an image display medium.

Huck, et. al. (1999) apply these metrics to images subjected to various degradations, and draw the following conclusions. The information rate of the acquired signal varies with mean spatial detail of the radiance field and reaches a broad peak when this detail is near the

sampling intervals. The peak occurs when the image power spectral density closely matches the sampling passband of the image gathering device. The information rate that a sampled imaging system conveys is closely correlated with the fidelity, sharpness, and clarity with which images can be restored. The associated theoretical minimum data rate is closely correlated with the lowest data rate with which the acquired signal can be encoded for efficient transmission.

3.3 Image Utility Prediction Models

Figure 3.16 offers a simplified view of a notional image utility prediction model. In this model, the three independent axes of the coordinate space represent the three parameters (or image fidelity measures) that determine the value of a point on the surface of constant image

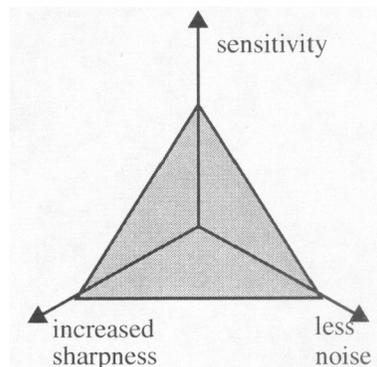


Figure 3.16: Notional image utility prediction model (from Hailstone, 2005)

utility. Ideally, the output of the model would be accurate and consistent predictions of image utility based on the input image fidelity measures discussed in Section 3.1.

This section describes how image utility is predicted, or more specifically, how the components of image quality assessment measures are applied to project forward to indicate image utility before actually assessing it. The literature does not describe these models as predicting image *utility*; rather it describes them as predicting sensor *performance*. This is just a matter of terminology, for as we pointed out in Chapter 1, system performance is synonymous with the utility of specific images, and thus we are predicting image utility.

Leachtenauer and Driggers (2001) state that performance prediction is based on some type of modeling process. At its simplest, this process may be a regression model relating

some image fidelity measure to a measure of information extraction performance (utility). At the complex end of the spectrum, the model may include computer simulations of the imaging process or theoretical models of the HVS. In order to model or predict performance, one or more predictor variables are needed. These variables may relate to the spatial and tonal characteristics of imagery (resolution, contrast, etc.), may be based on image measurements (physical image quality), or be based on image differences (pixel amplitude differences). The variables could also relate to the informational content of the imagery.

Holst (2006) asserts that models must be able to relate design parameters, laboratory measurements, and operational performance. He specifies three levels of models that satisfy these requirements: 1) component/phenomenology models that relate quantifiable design and environmental parameters to higher level summary parameters such as MTF and SNR, 2) system performance models that characterize total system performance for controlled tasks such as detection of periodic bar pattern in order to predict standard performance measures such as MRC, and 3) operational models that combine system models to characterize overall operational tasks such as probability of discrimination. Our framework for the role of utility prediction concerns Holst's third level, because we seek to predict image utility, not image fidelity measures such as SNR, MTF, and MRC.

We see two approaches to models that predict image utility. These approaches are reflected in the structure of this section. The first uses parameters of image chain components to predict a performance measure – a parametric approach. The second approach simulates the scene and predicts performance by assessing the utility of the synthetic image.

3.3.1 Parameter-based Models

There are three methods by which we can formulate a parametric relationship between the image fidelity measures and utility: 1) empirically deriving the relationship based on many observations of the behavior of the variables, 2) analytically deriving the relationship and applying it to sensor-derived parameters, or 3) deriving the parameters from the image and applying them to an analytical relationship. The next three subsections address each of these.

3.3.1.1 Empirically-derived Predictive Models

In traditional panchromatic image analysis, the ability to extract information from an image is based on the ability of the observer to detect objects of various types in a scene. This is based on the capacity to detect edges or tonal changes in the image. Consequently, image fidelity measures that relate to the ability to perceive small detail and edges, as well as measures relating to contrast and noise, are used to predict observer performance. Typical measures include the sensor-derived image fidelity measures such as scale, resolution, contrast, noise-related measures, and edge sharpness measures discussed in Section 3.1.1. The empirical model takes the form of a regression equation that predicts performance as a function of one or more independent variables and requires a set of empirical data to develop an initial model. Once the model is validated, performance predictions may be made on the basis of system design information (Leachtenauer and Driggers, 2001). There are two models in this category, the general image quality equation (GIQE), which predicts image interpretability in terms of NIIRS, and the probability of discrimination model, which predicts object detection probabilities.

General Image Quality Equation

The GIQE was developed to relate the NIIRS ratings more closely to quantities measurable in an image or based on system design parameters. It is based on regression analysis of measured parameters against analyst-defined NIIRS values for a large number of images over the range of NIIRS values (Schott, 2007). The GIQE predicts NIIRS as a function of perceptual-quality attributes of image scale, sharpness, resolution, and the SNR (contrast and noise). Fig 3.17 shows the conceptual model of GIQE and includes the image chain components that the model takes into account. The equation for visible imagery is given as:

$$NIIRS = 10.251 + a \log_{10} GSD_{GM} + b \log_{10} RER_{GM} + 0.656H_{GM} - 0.344 \frac{G}{SNR} \quad (3.22)$$

$$\begin{cases} a = 3.32 \text{ if } RER \geq 0.9 \\ a = 3.16 \text{ if } RER < 0.9 \end{cases} \quad \begin{cases} b = 1.559 \text{ if } RER \geq 0.9 \\ b = 2.817 \text{ if } RER < 0.9 \end{cases}$$

The terms in the GIQE derive from earlier research relating physical image quality to interpretability. GSD_{GM} is the geometric mean of the ground-sampled distance in inches and addresses sensor (scale and resolution) and target effects (aspect, size, contrast), RER_{GM} is the geometric mean of the normalized relative edge response and addresses sensor effects

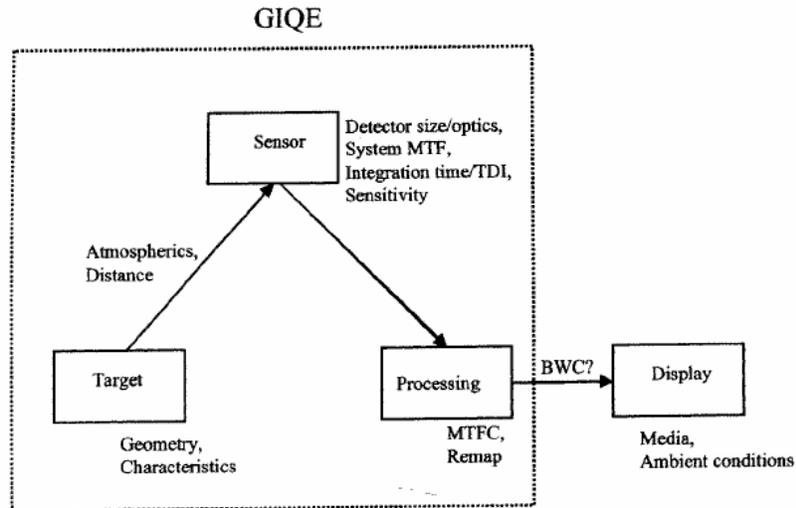


Figure 3.17: GIQE conceptual model (from Leachtenauer and Driggers, 2001)

(perceived sharpness), H_{GM} is the geometric mean height of overshoot across an edge due to MTF compensation (commonly used in image processing to enhance analyst evaluation), G is the noise gain due to MTF compensation, and coefficients a and b are defined according to the RER value. The GIQE was validated using a visible data set split into two equal subsets – one for model development and one for validation (Leachtenauer, et. al., 1997).

Probability of Discrimination

The probability of discrimination model is extensively used in the development of battlefield target acquisition systems and is dependent on parameters that capture the target dimensions, target-to-background contrast, and atmospheric conditions. This family of models has its origins in the work of Shade, who derived performance measures for photographic, motion picture, and television systems as a function of light level in the 1950s and 60s. These models are based on the ability of an observer to resolve a standard bar target in the presence of noise.

Producing a static probability of discrimination curve as a function of range requires four parameters: estimated target-to-background temperature (for infrared systems) or reflected light luminance (for electro-optic systems) differential, estimated height and width of the target, atmospheric transmission estimates within the spectral band of interest for a number of pertinent ranges, and the sensor MRT or MRC. The atmospheric degradation of the contrast is responsible for an apparent differential temperature or apparent contrast at the imager entrance pupil. The number of cycles, N , across the characteristic target dimension

that can be resolved by the sensor at a particular range is determined by the following relationship:

$$N = \rho \frac{d_c}{R} \quad (3.23)$$

where ρ is the maximum resolvable spatial frequency in [cycles/milliradian], d_c is the characteristic target dimension in meters ($d_c = \sqrt{\text{Width} \cdot \text{Height}}$), and R is the range from the sensor to the target in kilometers (Leachtenauer and Driggers, 2001).

The probability of discrimination for a given number of cycles, N , across a target is determined using the target transfer probability function for the desired level of discrimination:

$$P(N) = \frac{\left(\frac{N}{N_{50}}\right)^{2.7+0.7\left(\frac{N}{N_{50}}\right)}}{1 + \left(\frac{N}{N_{50}}\right)^{2.7+0.7\left(\frac{N}{N_{50}}\right)}} \quad (3.24)$$

In this equation, the term N_{50} corresponds to the number of cycles to achieve a 50% probability of discrimination. This is derived from Johnson's criteria which describes the number of bars in a bar pattern required to represent an object for a particular level of discrimination. Three objects and their corresponding bar targets are shown in Figure 3.18. Objects are represented as bar targets for purpose of sensor evaluation, analysis, and design. Johnson determined through experiments the number of bar pairs (cycles) needed to subtend an object in order for an observer to perform a particular discrimination (detection, recognition, or identification) task. MRT and MRC are then applied to the apparent target-to-

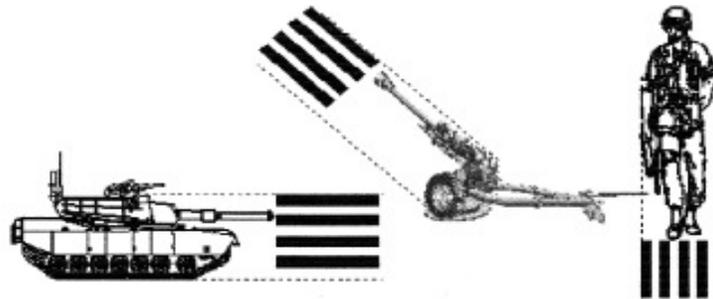


Figure 3.18: Objects and their corresponding bar targets (from Driggers, et. al., 1998)

background characteristics to give a frequency response that is compared to Johnson's criterion to obtain a probability of discrimination (Driggers, et. al., 1998). The entire target acquisition system model is shown in Figure 3.19 for an infrared sensor. Starting at the top

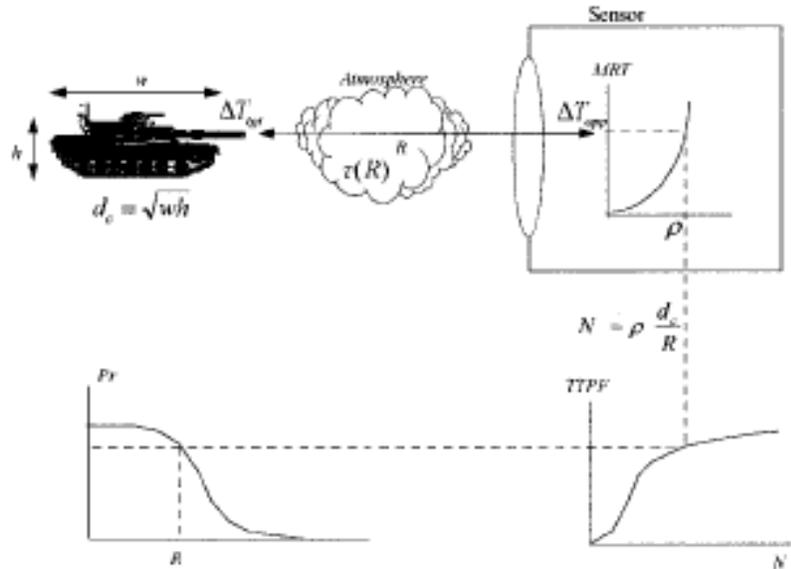


Figure 3.19: Target acquisition system model (from Driggers, et. al., 1998)

left with the object to be imaged, the characteristic dimension and differential temperature are propagated through an atmosphere, where the MRT is converted to a maximum resolvable spatial frequency at the top right. This spatial frequency is converted to a number of cycles using equation 3.23 and then converted to a probability of discrimination using the target transfer probability function at the bottom right. The bottom left converts the probability of discrimination into an equivalent maximum range corresponding to the probability of discrimination (Driggers, et. al., 1998).

3.3.1.2 Analytical Predictive Models

The analytical approaches describe the relationship between input parameters and output utility measures from physical process equations rather than empirically-derived expressions. The input parameters may either be image fidelity measures or they may be statistical descriptions of the image content.

Physique

This model was developed by the Eastman Kodak Company and predicts information content. Information is then related to performance measures such as NIIRS or probability of

recognition. Inputs are entered in 11 groups that include target type, orientation, height, density, and reflectance; solar altitude, sensor-target-sun angle, target zenith distance; percent of scene illuminated by five illumination types: f-number, focal length, filtration, solar altitude, vehicle altitude, atmospheric transmittance, obliquity, and aspect angle of sun and look vectors; target reflectance mean and variance; atmospheric model; exposure mean and variance; system MTFs due to optics, smear, cross talk, focus; photon and system noise; quantization noise, bit errors, and data transmission encoding and decoding. Physique is an encompassing model at the complex end of the spectrum of image utility models (Leachtenauer and Driggers, 2001).

Statistically-based parametric analytical models

The idea is to propagate a statistical description of the image through an image chain in order to predict utility for the modeled scenario in a specific application task. These parametric models rely on mathematical descriptions of various components of the image chain in order to make their predictions of utility. The underlying premise is that surface classes of interest may be represented by first- and second-order spectral statistics and that the effects of various processes in the end-to-end spectral imaging system can be modeled as transformations and functions of those statistics. These approaches do not rely on a real image, but rather use notional image statistics derived from real images. These statistical analytical models do not produce an image, but rather compute expected performance using analytical equations.

Kerekes and Baum (2002) introduce this approach to a spectral image subpixel object detection scenario. The system diagram of the model is shown in Figure 3.20. Their

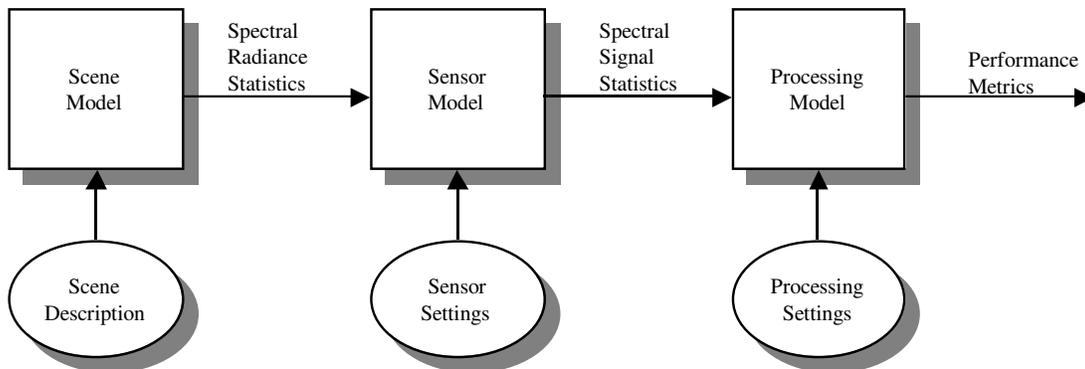


Figure 3.20: Block diagram of FASSP model (from Kerekes, 2004)

end-to-end spectral imaging system model includes the significant effects of the remote sensing process. The model is driven by a user-specified input set of system parameter settings that describe the scenario, scene classes, atmospheric state, sensor characteristics, and processing algorithms. Output is in terms of probability of detection for a given subpixel target scenario (ROC curves). Schwartz, et. al. (1995) developed a similar statistical parameter-based approach for a multispectral sensor in an anomaly detection scenario. The output is in terms of ROC curves. Although these models were designed with spectral imagery in mind, they can be applied in the limit of a single spectral channel to panchromatic image utility prediction, which is why they are discussed here.

3.3.1.3 Image-derived

These parametric models draw their parameters directly from the image being evaluated.

Image Quality Model

This model measures the power spectrum of a digital image and computes the information content of the image, which in turn is related to NIIRS. The input parameter is the power spectral density of the image. The image quality model is derived from the normalized 2-D power spectrum weighted by the square of the HVS MTF, the directional scale of the input image, and the modified Wiener noise filter (Nill and Bouzas, 1992). The image quality model output shows good correlation with image analyst assessed NIIRS ratings on digitized aerial photos.

Sarnoff JND model

This model computes differences between two images in terms of the HVS response expressed as JNDs. The number of JNDs is then related to some measure of relative quality or interpretability (NIIRS). Input to the model is a pair of digital images differing in some process (compression, smear, etc). Four other inputs are image sample spacing, observer-to-image distance, visual fixation depth, and eccentricity of the image in the observer's field of view. Successive stages model the various functions in the HVS. To validate the model, image analysts rated delta NIIRS on a set of images and then the JNDs were computed. The JND and delta NIIRS showed good correlation (Leachtenauer and Driggers, 2001).

3.3.2 Simulation

In simulation approaches, an image of a physical model (small scale mock up) or the simulated image of a scene is generated. The resulting image itself or the measurements of the image are the performance predictors. A specific scene scenario is generated and then information extraction is performed and the performance is noted. Unlike the parametric approaches, simulation does not take the parameters (fidelity measures) and put them into an equation to get a utility measure; it actually simulates the image formation process. Scene elements, radiation propagation, and sensor effects are simulated using computer models. Simulations attempt to model the physical processes taking place in the image chain with a first principles approach.

Kerekes and Landgrebe (1989) describe the elements of simulating an optical remotely sensed image. A high resolution scene file contains the spatial and spectral discrete version of the spectral radiance function present at the input aperture of the sensor. The user-defined sensor then operates on the scene file to create the remotely sensed image and specified processing algorithms are applied to the image to obtain the performance metric. A high spatial resolution implies scene pixel sizes several times smaller than those of the sensor. High spectral resolution means several spectral samples per sensor spectral channel. The scene model consists of the surface reflectance array of reflectance vectors arranged spatially by class and derived from field measurements and the solar illumination and atmospheric effects process (Kerekes and Landgrebe, 1989).

Schott (2007) states that simulation tools give sensor designers the ability to evaluate tradeoffs between image fidelity parameters. Synthetic images can be produced over a range of spatial, spectral, and radiometric performance specifications and the resulting images evaluated in terms of application-specific performance metrics to determine the utility of the sensor in a given application. While simulation allows for many variations in the scene and interaction process, the computational complexity in terms of scientific issues, coding, and run time are disadvantages (Schott, 2007). Components include object data, material database, scene database, meteorological, atmospheric databases and are shown in Figure 3.21. The block diagram also shows the flow and interactions needed to generate a synthetic image model.

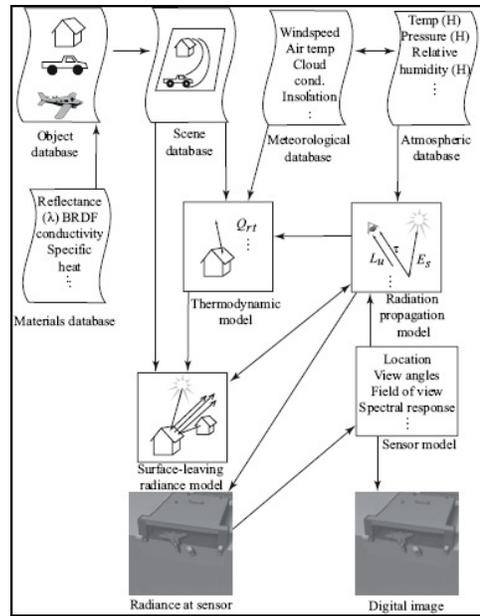


Figure 3.21: Synthetic image generation components (from Schott, 2007)

3.4 Summary

This chapter has presented a review of image quality approaches drawn primarily from the field of aerial and satellite remote sensing. Image fidelity measures were organized according to whether they were derived from sensor parameters or from the image itself. The sensor-derived approaches were either generated by imaging a reference scene or totally independently of a reference scene. Image-derived fidelity measures were either derived from a single image or by comparing multiple images. Image utility measures were organized based on the definition of utility: the ability to complete a task or the ability to convey information. The task performance approaches were categorized as either being direct performance measures or as performance estimates. The information theory utility measures were discussed and several metrics presented. The final section of this chapter focused on the models used to predict image utility. Image utility predictive models are built on image fidelity parameters with the purpose of predicting image utility. These models were categorized as either parametric or simulations. The parametric models consisted of those which were empirically derived, those that were analytically-based, and those that operated on image parameters derived from the image. Simulation approaches obtain image utility by

Chapter 3. Background: Image Quality Review

generating specific scene scenarios and then applying information extraction algorithms and noting the resulting performance.

Chapter 4

Background: Spectral Imaging

The image quality methods that we have discussed thus far have primarily been predicated on the visual interpretive skills of the human observer. The spectral image, however, does not readily lend itself to visual interpretation. Adams and Gillespie (2006) write, “*Our day-to-day visual observations depend heavily on an ability to derive meaning from the shapes, sizes and textures of the objects and patterns in the world around us. We make relatively little use of the spectral information...our visual experience is dominated by spatial, not spectral, information.*” They contend that because we are so engrained in the spatial aspect of image interpretation, many approaches to exploiting information from spectral images continue to rely on the more familiar spatial characteristics, resulting in an underutilization of the spectral information.

In order to establish the context for assessing and predicting spectral image quality, we need to explore what spectral imagery is and what makes it different than panchromatic imagery. We also review the major applications of spectral imagery that seek to exploit different types of information from spectral images. In particular, we examine the theory and mechanics of the target detection task as well as target detection performance assessment and prediction.

4.1 Spectral Imaging Fundamentals

Spectral imaging may be defined as the process of collecting spatially co-registered images in multiple spectral bands. Multispectral imaging (tens of spectral channels) has been in existence since the late 1960s. Advances in detector technologies in the 1970s and 1980s enabled sensor systems with increased spectral and spatial image resolution, so that by the late 1980s it became technically feasible to acquire simultaneous images in hundreds of narrow (tens of nanometers wide) spectral channels. This section examines several relevant aspects of spectral imaging in detail.

4.1.1 Physical Basis

Imaging spectroscopy combines the measurement, analysis, and interpretation of material spectra (the discipline of spectroscopy) with methods to acquire spectral information over large spatial areas (the process of imaging). Spectral imaging exploits the fact that materials comprising the scene will reflect, absorb, and emit electromagnetic radiation in ways characteristic of their molecular composition and shape. Spectral patterns are produced by fundamental differences in the capacity of surface materials to absorb, reflect, and emit radiant energy. The wavelengths where light has been absorbed are the most informative, as the photon energy is absorbed by activating certain electronic or vibrational processes. Processes of particular interest for spectroscopic identification of materials are electronic transitions within certain ions, transfer of electrons between ions, and assorted molecular vibrations and rotations (Adams and Gillespie, 2006).

4.1.2 Image Formation

A scene may be represented as a continuous function of space, wavelength, and time variables. The process of spectral imaging entails the spatial, spectral, radiometric, and temporal sampling of this function space, resulting in a finite-resolution image. The sensor's detector array spatially samples the image, and possibly the spectral spread of the dispersed (by prism, grating, or interferometer depending on the design of the spectrometer) radiation. The spatial resolution of the sensor characterizes the size of the smallest object that can be seen in the image as a distinct object separate from its surroundings. Spectral resolution is determined by the width of the spectral channels used to measure the radiance at the different wavelengths. Radiometric resolution is determined by the number of bits used to quantize the radiance value measured by the sensor at each spectral channel. Temporal resolution is related to how often the sensor revisits the scene to obtain a new set of data (Manolakis, Marden, and Shaw, 2003).

The process of collecting two-dimensional spatial images over many narrow spectral channels with a linear detector array or a two dimensional focal plane imaging sensor requires some form of time-sequenced imaging. This collection may be accomplished by either a time sequence of one-dimensional spectral pixel vectors at each spatial pixel (line

scanner design), a time sequence of two-dimensional spatial images at each waveband of interest (filter wheel design) or a time sequence of spatial-spectral images with multiple one-dimensional spatial images collected over time to obtain the second spatial dimension (pushbroom design). Figure 4.1 shows how each of these three sensor configurations

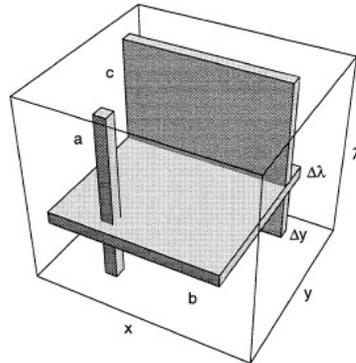


Figure 4.1: Three spectral image formation methods (from Descour and Dereniak, 1995)

construct a spectral image having spatial dimensions x and y and spectral dimension λ . In Figure 4.1, the letters a, b, and c represent the portion of the spectral image obtained in a single temporal sample (frame) of the sensor detector array. Letter a represents the pixel vector collected at one spatial location by the line scanner design, b represents a single spectral channel of an image collected by a filter wheel design of spectral resolution $\Delta\lambda$, and c represents a line of spatial pixels in all spectral channels formed in the pushbroom configuration as the sensor moves in the y direction. The feature common to all of these instruments is the need to scan in space or along the spectral dimension, sampling at each temporal sample to measure the entire cube (Descour and Dereniak, 1995).

Approaches to spectral image formation not shown in Figure 4.1 include Fourier transform spectrometer designs, in which a detector integrates spectral amplitude and spectral-interferogram phase for multiple spectral bands. Spectral imaging can also be accomplished without scanning by using computed tomography techniques to integrate the weighted sum of signals originating in different spectral bands and at different positions in the scene. This involves undoing the spatial and spectral multiplexing inherent in the images captured on the two-dimensional focal plane array that occurred because of projecting through the three-dimensional spectral image “volume.”

4.1.3 Motivation for Spectral Imaging

Landgrebe (2002) describes the motivation for the first spaceborne multispectral imaging system, *“The fundamental basis for space-based remote sensing is that information is potentially available from the electromagnetic energy field arising from the Earth’s surface and, in particular, from the spatial, spectral, and temporal variations in that field. Rather than focusing on the spatial variations, which imagery perhaps best conveys, why not move on to look at how the spectral variations might be used?”* The pragmatic impetus for this question is that spatial resolution is one of the most expensive parameters to achieve in a spaceborne imaging system. The search for a more economical approach that did not require such high spatial resolution was based on the belief that the characteristic spectral responses of materials in the scene would allow ground cover types to be classified based on spectral rather than spatial information. The fundamental assumption was that different classes of surface cover have unique spectral responses within a data set. In this situation, the spatial relationship between pixels was not needed to facilitate identification, since the spectral response of the pixel was assumed to be unique. Thus, the primary motivation for the spectral imaging concept was cost effectiveness to facilitate remote classification of large earth surface areas. Because data volume grows with the square of spatial resolution but only linearly with the number of spectral bands, reducing the number of pixels needed to survey a given spatial area was the goal of the first spectral imaging systems.

4.1.4 Comparison with Panchromatic Imaging

The characteristics of spectral and panchromatic imagery lead to different applications for each type of imagery. Spectral imaging is best suited for applications in which spectral information is more reliable or measurable than spatial or shape information. These include land cover classification, material identification, and labeling on a per pixel basis. The discrimination and detection of different materials using spectral information is sometimes termed nonliteral exploitation in reference to the fact that this process does not rely on literal interpretation of an image by means of morphological analysis and inference (Manolakis, Marden, and Shaw, 2003).

Given that the applications are different, the processing to gain the desired information is necessarily different. In spatial processing, information is embedded in the spatial arrangement of pixels using geometrical shape information, whereas in spectral processing the spectrum associated with each pixel individually is used to uniquely identify the material in the corresponding ground resolution cell. High spatial resolution is required to identify objects by shape, but only a single pixel is needed to identify a material spectrally (Manolakis, Marden, and Shaw, 2003). Whereas panchromatic images are analyzed with photointerpretation techniques in which image enhancement methods improve the visual interpretability, spectral images are processed in general with machine analysis employing numerical algorithms to label individual pixels (Richards and Jia, 1999).

4.1.5 Views of the Data

Spectral data may be viewed in three fundamental spaces. The first is the image space representation in which all spatial information is shown, but only one spectral channel at a time. This is useful in locating a pixel spatially, but its effectiveness does not extend beyond simultaneous display of two to three spectral bands. The second is the spectral space representation where the brightness values in each channel of a single spatial pixel are plotted. This view is useful for relating a pixel's spectral response to phenomenology, and leads one to believe that each ground cover material may be represented by a single spectral signature. In reality, there is substantial observational and scene variability, and it is difficult to display all of the diagnostic variation in this spectral space. The third representation is in the feature space, where spectral pixel vectors are defined as points in K -dimensional space (assuming K spectral channels). Feature space allows the diagnostic variation of all image pixel vectors to be represented mathematically and computationally, but not visually for more than three spectral channels (Landgrebe, 2003). The ability to describe the entire image in the K -dimensional spectral space offers the potential of finding a function that offers good discrimination between spectral materials, which is key to extracting information from spectral images.

Because the diagnostic variation of the spectral vectors in feature space is of great interest, we seek mathematical models to characterize this spectral variability. Manolakis, Marden, and Shaw (2003) identify three models to account for this variability. The first

model is that of a probability density model which represents the data as a mixture of conditional probability distributions specified by parameter vectors. If M is the number of distinct spectral classes, π is the prior probability for a specific class, and θ is a parameter vector that specifies the probability distribution, then the probability distribution of the data is $f(\mathbf{x})$ and is represented as the weighted summation of individual spectral class probability distributions, $f(\mathbf{x};\theta_k)$:

$$f(\mathbf{x}) = \sum_{k=1}^M \pi_k f(\mathbf{x};\theta_k) \quad (4.1)$$

The second model is based on subspaces, and it restricts the pixel vector to vary in the M -dimensional subspace of spectral channel space (K) (where $M < K$). The $K \times 1$ pixel vector \mathbf{x} is described as a weighted summation of linearly independent $K \times 1$ column vectors \mathbf{s}_k of the $K \times M$ matrix \mathbf{S} with the abundances of each described by scalars a_k in the $M \times 1$ vector of abundances, \mathbf{a} .

$$\mathbf{x} = \sum_{k=1}^M a_k \mathbf{s}_k + \mathbf{w} = \mathbf{S}\mathbf{a} + \mathbf{w} \quad (4.2)$$

The $K \times 1$ additive error vector \mathbf{w} is assumed to be random. The linearly independent column vectors define the variability subspace and can be obtained with statistical techniques such as eigendecomposition of the data correlation matrix. The third model is the linear mixing model that assumes that the observed reflectance spectrum is generated by a linear combination of endmembers (constituent material deterministic spectral signatures). Endmembers may be obtained from spectral libraries, in-scene spectra, or geometrical techniques. Figure 4.2 shows each of the three spectral variability models in $K=3$ feature space.

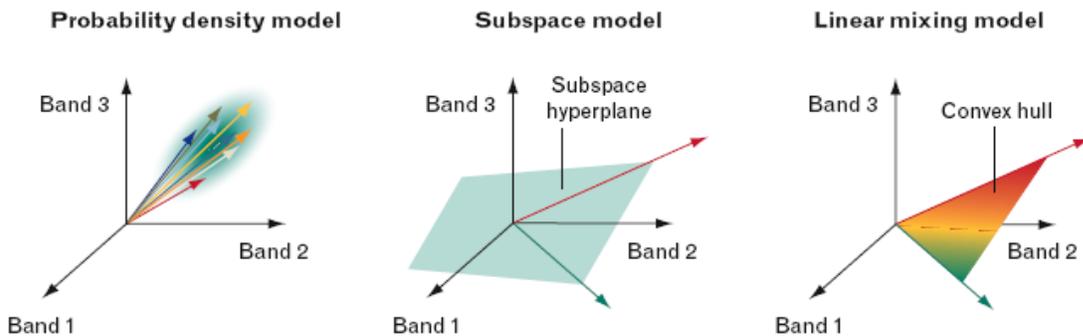


Figure 4.2: Spectral variability models (from Manolakis, et. al., 2003)

4.1.6 Spectral Imagery Characteristics

The image fidelity measures discussed in Section 3.1 are standard ways of characterizing panchromatic images. In this section, we explore several aspects of spectral imagery which help us more fully understand the nature of these types of images.

4.1.6.1 Spectral Contrast

Adams and Gillespie (2006) offer two ways in which to describe the spectral *contrast* of spectral imagery. The first is wavelength-to-wavelength contrast. A reflectance spectrum contains information about the relative absorbing power from one wavelength to another. The wavelength-to-wavelength contrast of a pure material becomes smaller as the spatial pixel size increases due to the diluting effect of other material spectra and geometrically diverse surfaces. The loss of wavelength-to-wavelength contrast affects the amount of information that we can extract from spectral images about the composition of a sample. The second description of contrast is termed pixel-to-pixel contrast and is defined based on the difference in spectral vector length and angle in K-dimensional spectral space. Spectral length is defined as the Euclidean distance between vectors and measures the overall spectrum lightness but is ambiguous in the presence of topographic shading and shadowing. Spectral angle is relatively insensitive to a variety of gains, including those imposed by topographic shading but is also insensitive to the overall reflectance. The spectral angle in radians between two pixel vectors \mathbf{x} and \mathbf{y} is defined by Yuhas, Goetz, and Boardman (1992) as:

$$\cos^{-1}\left(\frac{\mathbf{x} \bullet \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}\right) \quad (4.3)$$

where the $\|\cdot\|$ are the Euclidean norms. Pixel-to-pixel contrast is increased by making the spatial pixel size smaller, since we will be able to spatially resolve more objects that are different from each other. Thus, we see that decreasing the pixel size increases both pixel-to-pixel and wavelength-to-wavelength contrast, and reveals the fundamental tradeoff between spatial and spectral resolution inherent in spectral imagery.

4.1.6.2 Spectral Resolution

The objective of spectral image analysis algorithms is to identify and segregate materials based on their unique reflective properties when observed over a wide range of wavelengths. This implies that spectral channels must be sufficiently narrow and appropriately placed to resolve spectral features. Different materials exhibit different spectral features: certain paints and vegetation can be characterized by broad slowly varying spectral features while minerals and gases possess very narrow spectral features. The location of these features in the spectral channels varies for each class of material. Therefore, narrow spectral channels may be needed to resolve subtle features in differentiating similar spectra, and spectral channels that cover contiguous bands are needed to handle the expected variety of materials, since important features may be in different spectral locations for each material. In addition, narrow spectral channels that straddle the water-vapor absorption bands are important in estimating and correcting for the variable water vapor contained in the atmosphere (Shaw and Burke, 2003). In most cases, hyperspectral sensors oversample the spectral signal to ensure that any narrow features are adequately represented (Shaw and Manolakis, 2002).

4.1.6.3 Spatial and Spectral Resolution Tradeoff

As we increase the spectral resolution in an attempt to better discriminate between material classes, we must balance this with other design parameters, most notably the spatial resolution. When the achievable SNR is limited by the imaging process in the sensor and not the noise of the scene, the SNR^2 of the sensor grows in proportion to the product of the receiver aperture area d^2 , the area of the pixel on the ground (GRD^2), the time interval over which the signal is integrated at the detector, τ , and the scene radiance at the sensor, L , divided over the K spectral channels, giving the following proportionality relationship:

$$SNR^2 \propto d^2 \times GRD^2 \times \tau \times \frac{L}{K} \quad (4.4)$$

This illustrates the fundamental tradeoff between spatial and spectral resolution (Shaw and Burke, 2003). Thus, the desire for finer spectral resolution to help with improved material identification competes with the desire for finer spatial resolution to improve the spectral contrast. This tradeoff is fundamental to imaging spectrometer design.

4.1.6.4 Spectral Mixing

Virtually any spectrum found in a scene is a mixture of constituent materials due to the spatial resolution of the sensor relative to the spatial variability present in the ground scene. The sensor integrates the radiance from all materials on the ground surface that are recorded by the sensor as a single pixel. When we analyze a mixed spectrum, we want to know the constituent materials and their proportions. As pixel size becomes larger, spectral mixing is an increasingly important factor. Depending on the scale at which we are viewing the material, the mixing may be best described in either a linear or a nonlinear fashion. At a millimeter scale, intimate mixing occurs when light is partially absorbed by one material before entering another material. The Bouger-Beer-Lambert law gives the absorption at each wavelength and states that such mixing will be non-linear. At the scale of meters, a linear mixing model adequately describes the mixing. As we move from the laboratory scale to the scale of remote sensing, linear mixing begins to dominate for most scenes (Adams and Gillespie, 2006).

4.1.6.5 Noise and Variability Characteristics

The energy field arising from the earth is finite in magnitude, and the spectral image collection process divides this finite quantity spatially into pixels. The power level in each pixel is then divided into spectral channels. There is a tradeoff between spatial and spectral resolution as seen in equation 4.4. At finer spatial and spectral resolutions, less power is left to overcome the internal noise present in the sensor system, leaving us with a less precise measure of the signal level arriving from the surface. The random nature of a signal does not imply that it is simply noisy. Signal variation often has a structure about it that can bear information, which implies that definition of signal and noise is problem dependent (Landgrebe, 2003). While the term “spectral signature” may suggest a unique correspondence between a material and its reflectance spectrum, variability is observed in the reflectance spectra of most materials. Many mechanisms are responsible for this variability including uncompensated errors in the sensor, uncompensated atmospheric and environmental effects, surface contaminants, variation in the material such as age induced color fading due to oxidation or bleaching, and adjacency effects in which reflections from nearby objects in the scene change the apparent illumination of the material (in a wavelength-

dependent fashion). Seasonal variations also introduce large changes in the spectral character of a scene (Shaw and Burke, 2003). As Figure 4.3 shows, while the shapes of material spectra are fairly consistent, the amplitude varies considerably. Figure 4.3 shows multiple instances of vehicle paint spectrum retrieved from 114 fully resolved pixels in a

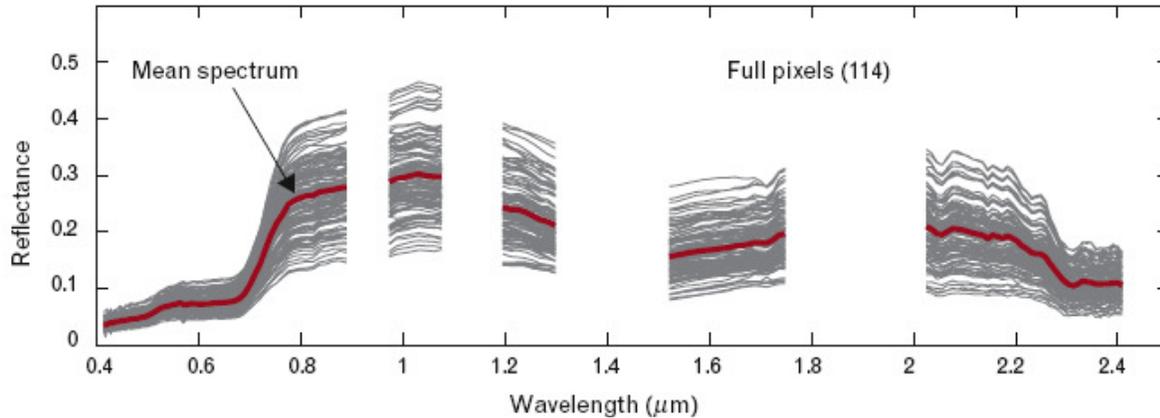


Figure 4.3: Spectral variability (from Shaw and Burke, 2003)

hyperspectral image. In an effort to exploit the spectral shape invariance, some of the more successful detection algorithms give more weight to the spectral shape than to the amplitude when determining the presence of a particular material in a pixel.

4.2 Spectral Imagery Application Areas

Different applications of spectral imagery seek to exploit the various types of information resident in the image. Shaw and Manolakis (2002) offer an organizational structure for the major groupings of spectral imagery application tasks. We will use their scheme in the next several subsections to provide a very brief overview of these applications.

4.2.1 Target and Anomaly Detection

The tasks in this application area are to search a spectral image for rare (known or unknown) spectral signatures. The availability of prior information about the target distinguishes target recognition from anomaly detection. Target detection algorithms search for known spectral signatures and operate by matching observations with reference spectra. Target matching approaches are complicated by the large number of possible objects of interest, the inherent

variability of the reflectance spectra of these objects (Figure 4.3), and the complications associated with accurately applying atmospheric compensation. Anomaly detection is characterized by the desire to locate and identify uncommon features in an image. Anomaly detection algorithms seek to distinguish observations of unusual materials from typical background without reference to the target signatures or target subspaces. Anomalies are defined as observations that deviate in some way from the neighboring clutter of background or the image-wide clutter background (Stein, et. al., 2002).

4.2.2 Unmixing

The task in the unmixing problem is estimating the fraction of the pixel area covered by each material in a mixed pixel (a pixel which contains a mixture of pure constituent materials). A fundamental question of unmixing is whether the mixture of spectral signature is formed by a linear or nonlinear process. Unmixing is an estimation problem that is a special case of the generalized inverse problem. System parameters are estimated by using one or more observations of a signal that has interacted with the system before arriving at the sensor. The process of unmixing involves three stages: dimension reduction, endmember determination, and inversion. The end result is an abundance image for each endmember representing the fraction of that particular material in each pixel. Unmixing algorithms may be organized by three criteria: how an algorithm interprets mixed-pixel spectra (statistical or deterministically), how an algorithm characterizes the randomness of the data (parametrically or non-parametrically), and the objective function used to optimize the algorithm (Keshava, 2003).

4.2.3 Change Detection

These applications involve finding significant changes (defined as being important to the user) between two images of the same region made at different times. The output of a change detection algorithm is a map of significant scene changes. These algorithms do not assume a target signature.

4.2.4 Classification

The classification task seeks to assign a class label to each pixel of the image. Landgrebe (2003) points out that successful classification strategies require that the classes be of informational value, exhaustive, and separable and modeled adequately. There must be enough training samples and they must be representative of the class intended. Classification algorithms pay no attention to resonance bands or spectroscopy per se, and are designed to compare vectors in a multi-dimensional spectral space. They compare image spectra with one another and group them according to similarity of predefined properties. The image analyst must determine the number and type of classes as well as quantitatively characterize these classes using spectral libraries or training data and ground truth information. The design of a good classifier requires a sufficient amount of training data for each background class. The natural criterion of performance is the minimization of the probability of misclassification errors.

The hierarchy of classifiers is based on the assumptions about relationships among the classes. The key variable controlling the selection of algorithm complexity is the number of training samples available by which to define each class in feature space. There are many classifiers from which to choose. Quadratic pixel classifiers are the most common. For more complex distributions, the data is broken into subclasses, each with a quadratic distribution. Non-parametric and iteratively trained algorithms, such as neural networks, can be made to perform well on individual data sets, but they need significant computation and larger training sets. The nature of the information being sought will determine the nature of the classification algorithm.

4.3 Spectral Image Target Detection

In this research, our focus is on the target detection task. At its most general, a typical detection task requires an algorithm to report the presence or absence of a target in an input image. The algorithm first computes a number that gives a measure of the evidence of the presence of the target, and reports target presence if the evidence strength is greater than a particular value. In any detection task there are some variables that affect the SNR in the

image, which, in turn affects the performance of the detection system (Kanungo, et. al., 1995).

The target detection task may take many forms. A target may be spatially resolved or unresolved, it may be rare or distributed widely throughout a scene, we may know the identity of a target and be seeking its location or we may know neither and suspect that an anomaly is present. To detect a target pixel against its background, the target must differ in brightness and/or its spectrum. Spectral contrast is a limiting factor in detecting a target against a background. Standard spectral classifiers do not consider spatial information, but treat each pixel independently. Sub-pixel detection seeks to determine the presence of a material of interest in a pixel when it is present in quantities of less than a fully resolved pixel. The limit of sub-pixel detection is the smallest fraction of a target endmember that can be measured to a specified degree of confidence in a given spectral mixture.

4.3.1 Reflectance Spectra

Since reflectance spectra are independent of the illumination, working in this domain provides the best opportunity to identify materials by matching them with spectra from a spectral library. The spectrum of the solar radiation reaching the earth is altered in a temporally and geographically dependent fashion due to the propagation of solar radiation through the earth's constantly changing atmosphere. These atmospheric modulation effects must be compensated in order to reliably recover the reflectance spectra of materials on the ground in a sunlit scene (Shaw and Burke, 2003). A simple and reliable means of transforming the observed sensor digital counts or radiance values into reflectances is called the empirical line method (ELM), and requires knowledge of reference objects such as reflectance calibration panels to be deployed in the scene. For each spectral band, ELM performs a linear regression to relate the observed sensor values of the calibration panel to the known reflectance value. The linear regression results in a gain, G , and offset, L_o , factor for each band, λ , that can be applied to a sensor radiance measurement, L_s , at every pixel (at the x, y spatial location) in the scene to yield a reflectance estimate, ρ , for that location:

$$\rho(x, y, \lambda) = \frac{L_s(x, y, \lambda) - L_o(\lambda)}{G(\lambda)} \quad (4.5)$$

The underlying assumption of ELM is that the atmospheric conditions over the calibration panels apply to the rest of the scene, and consequently ELM cannot account for any atmospheric variation across the scene.

If calibration panels are not available in a scene, then atmospheric compensation may be undertaken by estimating the gain and offset terms using naturally occurring scene materials or exploiting the statistical properties of the scene. The alternative approach is based on models of illumination and atmospheric scattering and absorption as a function of wavelength. Atmospheric transmittance and illumination depend on factors such as vertical temperature profile, water vapor concentration, concentration of mixed gases, concentration and types of aerosols, solar angle, cloud cover, shadowing, and viewing geometry. These physics-based models use band ratio techniques to quantify the effects of water vapor on hyperspectral radiance measurements made near the edges of known atmospheric water vapor absorption bands to estimate the column water vapor on a per-pixel basis (Shawn and Burke, 2003). Two of the more commonly used techniques are the atmospheric removal (ATREM) algorithm (Gao and Goetz, 1990) and the fast line-of-sight atmospheric analysis of spectral hypercubes (FLAASH) (Alder-Golden, et. al., 1998). Errors in estimates of environmental and sensor parameters may lead to significant errors in the estimate of reflectance spectral, which in turn lead to target detector performance loss due to signal mismatch. Local variations in illumination further complicate the conversion of radiance to reflectance (Stein, et. al., 2002).

Consideration must also be given to the reference library spectra against which matches are being performed in the target detection task. Factors that differentiate laboratory from remote sensor collected spectra include: geometric factors, calibration, atmospheric effects, spatial resolution, spectral resolution, impurities, and spectral variability. Spectra derived from images avoid issues of calibration and resampling of laboratory spectra to the channels of a particular imaging system. The main difficulty is to find image pixels which are pure samples of each material component.

4.3.2 Theory and Design of Detectors

The detection problem is formulated as a binary hypothesis test with two competing hypotheses: background only (target absent) or target and background (target present).

Given an observed random spectral pixel vector \mathbf{x} , we want to choose between two competing hypotheses:

$$H_0 : \text{target absent}$$

$$H_1 : \text{target present}$$

If we visualize the image in a two dimensional spectral feature space as shown in Figure 4.4,

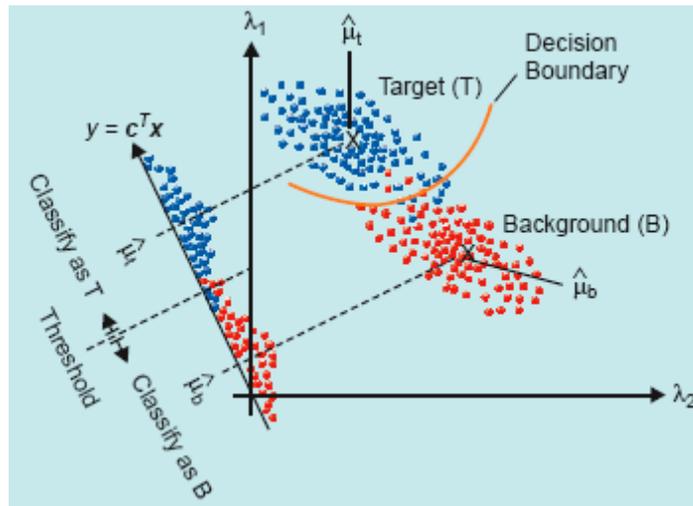


Figure 4.4: The target detection problem (from Manolakis and Shaw, 2002)

we note several interesting characteristics in this Figure. First, each point belongs to either a target (blue) or a background (red) class. In order to make a decision about class membership, we need to divide the feature space into two regions by some method. Second, in the ideal case, the target and background distributions would exist without any overlap and we would be able to make perfect decisions. However, in reality, they overlap, which results in classification errors. Third, the decision boundary will be a curve corresponding to a nonlinear detector, as shown by the orange line in Figure 4.4. Fourth, we can also make a decision by processing the pixel vector \mathbf{x} by a system that calculates a scalar $y = D(\mathbf{x})$ and then comparing y to a scalar threshold. This reduces our multivariate problem to a scalar one using a linear transform operator, $D(\mathbf{x})$, known as the two-class classifier, or discriminant function, statistics, filter, or detector. Figure 4.4 shows $D(\mathbf{x})$ as the operation $y = \mathbf{c}^T \mathbf{x}$, where \mathbf{c} is a linear filter (Manolakis and Shaw, 2002).

The $D(\mathbf{x})$ is derived as a likelihood ratio. Assuming that we have conditional probability density functions describing each hypothesis, $p(\mathbf{x}|H_0)$ and $p(\mathbf{x}|H_1)$, we obtain a likelihood ratio:

$$D(\mathbf{x}) = \frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} \quad (4.6)$$

If the ratio exceeds a specified threshold, then the target present hypothesis is selected as true. Thus, the likelihood ratio test or any monotonic function of it provides the information used to decide on the presence of a target. Detectors based on the likelihood ratio test have several advantages: they minimize the risk of incorrect decisions and lead to detectors that are optimum for a wide range of performance criteria (including maximization of separation between target and background spectra) (Manolakis, Marden, and Shaw, 2003).

In practice, the operator of the detector desires to set the threshold such that the number of detection errors is small and the number of correct detections is high. There is always a tradeoff between a low threshold to keep the probability of detection (PD) high and a high threshold to keep the probability of false alarm (PFA) low. This tradeoff is described by the ROC curve of Figure 3.13. If the conditional densities are completely known, we can choose the threshold to optimize the detector according to one of two criteria: the Bayes criterion, which chooses the threshold that leads to minimum overall error (both misses and false alarms), or the Neyman-Pearson criterion, which maximizes PD while keeping PFA under a certain predefined value. For target detection applications, the Neyman-Pearson criterion is most applicable, and the ROC curve of the optimum Neyman-Pearson detector provides an upper bound for the ROC of any other detector (Van Trees, 2001). In addition, practical target detection systems need to function without operator intervention to set the detection threshold in order to maintain a constant false alarm rate (CFAR).

In practical situations, the conditional densities are not known and must be estimated from the data. The generalized likelihood ratio test (GLRT) applies when the conditional probability densities depend on some unknown target and background parameters, estimated by their maximum-likelihood estimates. The GLRT leads to “adaptive” detectors that appear to work well in several applications. For low probability targets where there is insufficient training data for the target class, the GLRT approach is used for the development of target detection algorithms.

The ability of the models to capture the essential aspects of the data directly impacts the performance of the resulting detectors. The most successful target detection algorithms employ information about potential targets and backgrounds accurately and properly use

available information. Accurately modeling target variability in target detection algorithms leads to improved detection performance and accurate modeling of background variability leads to improved detector performance and facilitates the development of CFAR detectors (Manolakis, Marden, Kerekes, and Shaw, 2001). For full pixel targets, detection performance is mainly determined by the variability of target and background spectra since the spectrum observed by the sensor is either produced by target or background exclusively. For subpixel targets, the most important consideration is that the background spectrum adds an interference component to the variability of target and background spectra (Manolakis, Marden, and Shaw, 2003).

The analytical derivation of target detectors is often based on signal models involving multivariate normal distributions. However, the actual response of a detector to background pixels almost always differs from the theoretically predicted distribution for Gaussian backgrounds. Manolakis, Marden, and Shaw (2003) note that empirical distributions usually have “heavier” tails that strongly influence the observed false alarm rate of the detector. Departures from normality need to be understood, and one way to do so is to examine the more general case of the normal distribution known as the elliptically contoured distributions. A key characteristic of normal random vectors is the elliptical shape of their equal probability contours. The probability density function of $K \times 1$ random vector \mathbf{x} with an elliptically contoured distribution function depicts this:

$$f(\mathbf{x}) = (2\pi)^{-K/2} |\boldsymbol{\Sigma}|^{-1/2} \left[\int_0^{\infty} \alpha^{-K} e^{-\frac{1}{2}\alpha^2 d} f_{\alpha}(\alpha) d\alpha \right] \quad (4.7)$$

In this expression, K is the number of spectral channels, $\boldsymbol{\Sigma}$ is the covariance matrix of the data, the term in brackets is a positive, monotonically decreasing function for all K , d is the (squared) Mahalanobis distance of \mathbf{x} from mean vector $\boldsymbol{\mu} = \mathbf{0}$, $d = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$, and α is a random variable with PDF f_{α} that controls the density of the contours. These models may be used to select the threshold for CFAR detectors more accurately, develop detection algorithms that better exploit the statistics of spectral background data, and test the robustness of detectors designed under the normality assumption (Manolakis, Marden, and Shaw, 2003).

4.3.3 Detector Taxonomy

Manolakis and Shaw (2002) note three key factors in determining a taxonomy for target detection algorithms: 1) the type of model used for spectral variability (probability density, subspace, or linear mixing), 2) the composition of the pixel under test (pure or mixed), and 3) the mixed pixel model (linear mixing, which assumes deterministic endmembers, or stochastic mixing, which assumes random and independent selection of endmember spectra from multivariate normal distributions). We divide the discussion of specific detectors into those aimed primarily at full-pixel and sub-pixel detection tasks.

4.3.3.1 Full pixel

If we assume that we know or can estimate parameters of the target and background distributions, then we model the target absent and target present hypotheses with multivariate normal distributions having different mean vectors and covariance matrices, ($\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ for target absent and $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ for target present). Under this assumption, the Neyman-Pearson detector leads to a quadratic detector given by:

$$y = D(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) - (\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \quad (4.8)$$

This detector makes a decision by comparing the Mahalanobis distance of the observed spectrum between the centers of the target and background classes (the result is the curved decision boundary in Figure 4.4). If the target and background distributions have the same covariance matrix, then the detector becomes a linear processor known as Fisher's linear discriminant or a matched filter:

$$y = D(\mathbf{x}) = \mathbf{c}^T \mathbf{x} = \left[\kappa \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \right]^T \mathbf{x} \quad (4.9)$$

The linear filter is \mathbf{c} , and κ is a normalization constant. The output is the projection of the test pixel vector along the direction of the parameter (filter) vector, (shown in Figure 4.4). The optimum direction is the one that gives the best separability between the two classes (i.e. the best detector performance). The term "matched" implies that the detector evaluates the amount of correlation between the background-centered reference target signature and the test pixel spectrum in a whitened (symmetric covariance) space. The matched filter is also obtainable by maximization of a cost function. The resulting detector is:

$$y = D(\mathbf{x}) = \frac{(\mathbf{x} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)}{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)} \quad (4.10)$$

In the case of normal distributions, the performance of the matched filter is completely determined by the Mahalanobis distance between the target and background distributions. If the distributions have equal covariance matrices with uncorrelated components having equal variances, then the matched filter becomes a correlation detector (Manolakis, Marden, and Shaw, 2003).

A special case of the matched filter detector is the constrained energy minimization (CEM) detector. This detector obtained by minimizing the total energy or the output of the linear filter subject to the constraint that the output is one. CEM was derived from the minimum variance distortionless response beamformer in sensor array processing with the desired signature interpreted as the desired direction of signal arrival (Farrand and Harsanyi, 1997). The spectral image analog to the interference of multiple signals in array signal processing is the matched filter in a colored interference-plus-noise situation. When the CEM detector is normalized by the target energy, it provides the maximum likelihood estimate of the target abundance (Chang, et. al., 2000). Strictly speaking, the CEM is defined using data correlation matrices.

4.3.3.2 Subpixel

The classification of subpixel target detectors is created by the choice of mathematical model to describe background variability (subspace or statistical). The target spectral signature variability is described using the subspace model $\mathbf{S}\mathbf{a}$, in which $\mathbf{K} \times \mathbf{M}$ matrix \mathbf{S} contains all of the available prior information about the target, and \mathbf{a} is the $\mathbf{M} \times 1$ relative abundance of each of matrix \mathbf{S} 's, component column vectors (endmembers). The background variability can be described with either a subspace model (structured background) or a statistical distribution (unstructured background). Mixed pixels are usually modeled using the linear mixture model (Manolakis, Marden, and Shaw, 2003).

Unstructured Background Models

These are based on the hypotheses:

$$\begin{aligned} H_0: \mathbf{x} &= \mathbf{v} && \text{target absent} \\ H_1: \mathbf{x} &= \mathbf{S}\mathbf{a} + \boldsymbol{\sigma}\mathbf{v} && \text{target present} \end{aligned}$$

We assume that the additive noise has been included in the $K \times 1$ background vector \mathbf{v} . The background has the same covariance structure under both hypotheses but different variance, which is directly related to the amount of target in the pixel under test (also called the pixel fill factor). The GLRT approach leads to the adaptive coherence/cosine estimator (ACE) detector, which is shown in the case of the target subspace containing only one vector, \mathbf{s} :

$$y = D_{ACE}(\mathbf{x}) = \frac{(\mathbf{s}^T \boldsymbol{\Sigma}^{-1} \mathbf{x})^2}{(\mathbf{s}^T \boldsymbol{\Sigma}^{-1} \mathbf{s})(\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x})} \quad (4.11)$$

In whitened coordinate space, where the target covariance is of constant variance in each band and uncorrelated, the test statistic is the cosine squared of the angle between the test pixel and target subspace. The adaptive matched filter (AMF) is obtained as a special case of ACE for targets with amplitude variability (Manolakis, Marden, and Shaw, 2003).

Structured Background Models

The background variability is modeled by a subspace model and the two hypotheses are:

$$\begin{aligned} H_0: \mathbf{x} &= \mathbf{B}\mathbf{a} + \mathbf{w} && \text{target absent} \\ H_1: \mathbf{x} &= \mathbf{S}\mathbf{a} + \mathbf{B}\mathbf{a} + \mathbf{w} && \text{target present} \end{aligned}$$

The matrix \mathbf{S} is user-specified and represents the signal subspace, \mathbf{B} is determined from the data and represents the background subspace, and \mathbf{w} is white noise (which is the only source of randomness in this model). Manolakis and Shaw (2002) show that the GLRT approach leads to the adaptive subspace detection (ASD) algorithm.

4.4 Target Detection Performance Assessment

In general, the performance of detectors is based on their ability to enhance the “visibility” of the desired target and to accurately model the background statistics. A target becomes more visible to the detector when the background-target separation in detection statistic space increases. Manolakis, Marden, and Shaw (2003) discuss desirable target detection algorithm performance characteristics: a high PD, a low PFA, robustness to deviations from the assumed theoretical model, CFAR operation under the assumed statistical model, and a reasonable robustness to supplied and estimated parameters. There are several means by which to evaluate the performance of the target detection algorithm. Some of these include noting the target-background separation of the filter output distributions, the detected target

pixels at a given PFA, and the PFA to detect all target pixels. As a performance assessment tool, though, the ROC curve methodology is the most informative tool, as it is based on statistical criteria. ROC curves provide a means to evaluate detector performance or compare detectors independently of threshold selection. Any systematic procedure to determine ROC curves or the threshold requires specifying the distribution of the observed spectra \mathbf{x} under each of the two hypotheses. In most practical situations, the conditional probabilities needed for the likelihood ratio test depend on some unknown target and background parameters. Therefore, the ROC curves depend on the unknown parameters and it is almost impossible to find the detector whose ROC curves remain an upper bound for the whole range of unknown parameters (Manolakis and Shaw, 2002). There are some practical difficulties when trying to derive ROC curves for detection algorithms for hyperspectral images: 1) number of pixels in a hypercube limits the empirical PFA estimation (10^5 pixels will limit the PFA to no less than 10^{-5}), 2) the number of targets of a particular type or class in a scene is usually small, which limits the achievable accuracy and creates commensurately larger ROC curve confidence intervals (Kerekes, 2008), and 3) ground truth for the entire scene is usually limited, which makes confirmation of false alarms difficult (Manolakis, Marden, and Shaw, 2003).

4.5 Target Detection Performance Prediction

Kerekes (2006) motivates the need to predict (target detection) system performance with the following statement, “*An accurate understanding and ability to predict performance for a given system would be very useful not only during design and operation of the system, but could also possibly provide a metric for comparison with developing and evaluating new processing systems.*” The complexity and interdependence of the many aspects of the spectral imaging process make these goals challenging. The characteristics of the scene, the analysis task at hand, the intervening atmosphere, the sensor, the processing and analysis algorithms, and even the analysts’ expertise all affect the achievable performance. We examine tools for performance prediction for target detection in spectral imagery.

The forecasting of spectroradiometric system performance (FASSP) model, briefly described in Section 3.3.1.2, was developed specifically to explore spectral imaging system performance and parameter sensitivity. It considers the remote sensing process as a linear system propagating surface class statistics through the system in order to arrive at a measure

of performance. One application of this tool is to predict target detection performance for a specified scenario. Through the use of statistical descriptions for the target and background, and linear transformations to model the effects of the observing system and processing, the performance can be predicted analytically, rather than through a physics-based simulation. The underlying premises of the FASSP model are: 1) the various surface classes of interest can be represented by first and second order spectral statistics and 2) that the effects of various processes in the end-to-end spectral imaging system can be modeled as transformations and functions of those statistics and parameters. The model propagates the spectral statistics through the effects of the atmosphere, the sensor, atmospheric compensation, feature extraction techniques, and then applies a detection algorithm to convert the high dimensional statistics to a scalar test statistic (matched filter output) to which a threshold can be applied and detection performance computed. For subpixel target detection scenarios, the linear mixing model is employed and the pixel of interest containing the target is assumed to be a sample from a random process described by the area-weighted mixture of the target and background classes. The rest of the analytically-described scene is comprised of a number of homogeneous background classes, each covering an area percentage of the scene (Kerekes and Baum, 2002). Performance is depicted using ROC curves and plots of PD at a specified PFA plotted against the target fill fraction required to achieve that performance.

Schwartz, et. al. (1995) describe a statistical parametric multispectral sensor performance model that incorporates a mission flight model, a multispectral target and background signature model, a multispectral sensor model, and a multispectral target detection model. They demonstrate that the performance of mine detection algorithms depends on the statistics of the target and background signatures as processed by the sensor and presented to the detector. Thus, a key feature of the model is its ability to generate these statistics from first principles and process them using an appropriate sensor model.

4.6 Summary

This chapter has examined spectral imaging in detail. The physical basis, methods of image formation, and characteristics pertinent to spectral imagery were explored. Four major applications of spectral imagery were briefly discussed. Target detection in spectral imagery

Chapter 4. Background: Spectral Imaging

was closely reviewed, both from the theoretical view as well as the implementation of detectors for full pixel and subpixel detection problems. The methods of assessing and predicting target detection performance in spectral imagery were also briefly discussed.

Chapter 5

Prior Work: Spectral Image Quality

Having established the requisite background in image quality and spectral imaging, we now examine previous spectral image quality research. The reviews of traditional image quality and spectral imaging were necessary in order to appreciate the depth of research in photographic and panchromatic digital image quality antecedent to spectral imaging and the distinct characteristics of spectral images. Reconciling these two strands of inquiry may seem difficult, but the consideration of each is of paramount importance if we are to properly approach the topic of spectral image quality. We need the perspective gained by these inquiries in order to critically evaluate the current state of the art if we hope to advance it.

The key question in this critical evaluation must be, “Which spectral image quality approaches take the unique nature of spectral imagery and spectral image processing into account?” It seems logical to conclude that the methods that will advance the consideration of spectral image quality will be those that view spectral imagery for what it is (a different sort of image) and devise appropriate measures and models to do the job in this realm. We will use the observations from this Chapter’s review of methods to predict spectral image utility to formulate an approach that addresses some of the perceived shortfalls of existing methods.

The organization of this chapter attempts to mirror that of the review of traditional image quality approaches in Chapter 3. Figure 5.1 provides the organization for our examination of the spectral image quality approaches. Following an overview subsection, we discuss measures of spectral image fidelity, then measures of spectral image utility, and finally how some of the image fidelity measures are used by models to make predictions of the spectral image utility. The emphasis of our review of the spectral image quality approaches is on spectral image utility, as we believe that this component of spectral image quality is the most relevant to those interested in extracting information from spectral images.

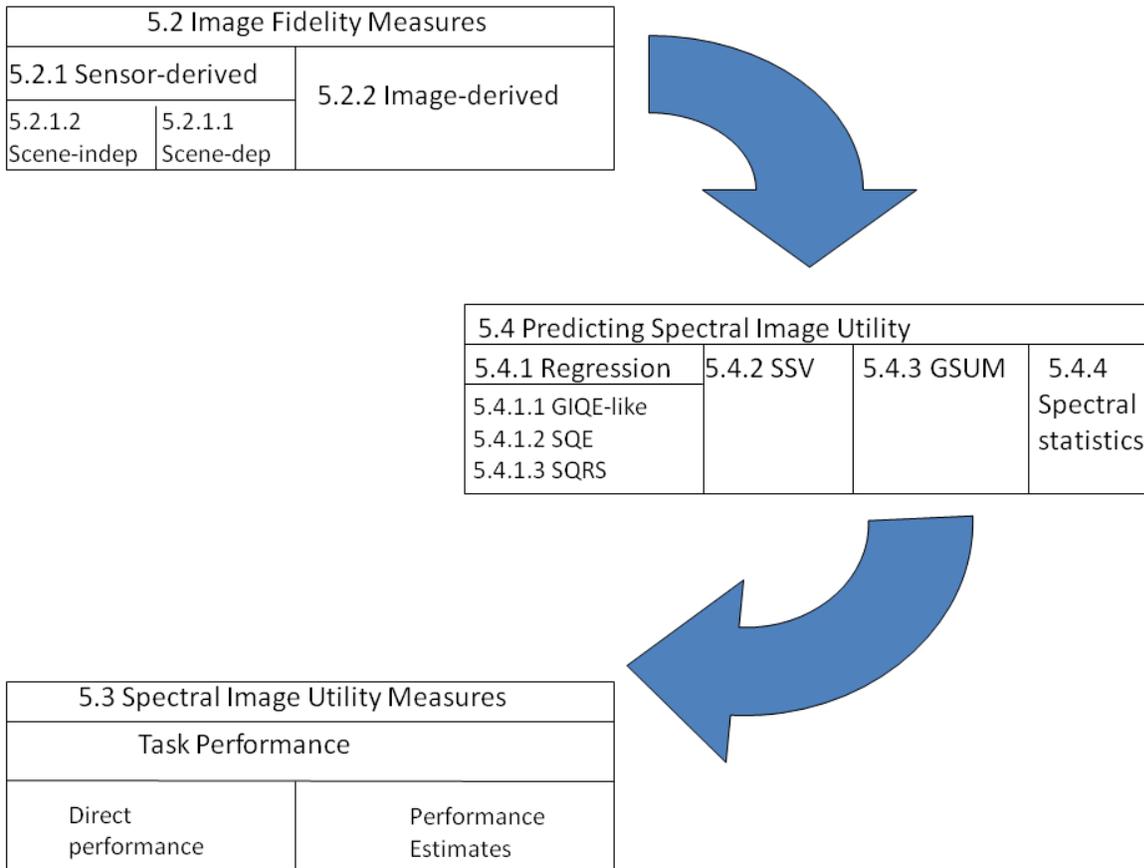


Figure 5.1: Organization of spectral image quality review

5.1 Overview

Regarding the issue of spectral image quality, Sweet, Granahan, and Sharp (2000) observe, “There is no objective definition of spectral image quality. It must be inferred from measurements of spatial resolution, calibration accuracy, spectral resolution, SNR, contrast, bit error rate, dynamic range, sensor stability, geometric registration, and other factors. While many aspects of sensor performance and image acquisition are measurable, a mathematical construction that incorporates these measurements into a quality rating is very difficult to derive. The quality of spectral imagery has not yet been described in a parametric manner. ...repetition of gray-scale quality measures over multiple bands may be useful, but it does not address spectral image quality.” Based on our reviews in Chapter 3 and 4, we see that traditional methods of predicting the utility of panchromatic images do not capture the radiometric and spectral aspects of spectral imagery. This is primarily because the utility of

panchromatic images is based on the perception of spatial patterns by a human observer, whereas spectral images are processed by computer algorithms based on the statistical or subspace characteristics of the image pixel vectors. The interaction between the spectral, radiometric, and spatial nature of a spectral image all contribute to its ultimate utility. Kerekes, Simmons, and Cisz (2005) describe the interaction between the spatial and spectral character and utility by pointing out that though an image with a small number of spectral channels but high spatial resolution may have high utility judged by someone looking at spatial information, an image with many spectral channels but moderate spatial resolution may have even higher utility when judged by an analyst looking at spectral information. The goal in reviewing spectral quality approaches is to gain an appreciation for how the various methods employ the constituent parameters of spectral imagery to form an estimate of utility.

5.2 Spectral Image Fidelity Measures

Many aspects of spectral image fidelity are the same as those already discussed for panchromatic images. Fundamentally, as with traditional image fidelity measures, the spectral image fidelity measures may be viewed as either sensor-derived or image-derived. The next two subsections describe each.

5.2.1 Sensor-derived

Image fidelity along spatial, radiometric, and spectral dimensions is measured using methods from traditional image fidelity, termed sensor-derived because they stem from measuring the characteristics of the sensing instrument. These metrics may be further divided into two categories, as with the traditional image fidelity measures: those sensor-derived metrics that depend on ground truth of some type in the scene and those that do not.

5.2.1.1 Scene-dependent

These sensor-derived metrics need a reference object in the scene in order to be calculated. Spatial, spectral, and radiometric fidelity are addressed by these measures. Spatial fidelity metrics include ground resolved distance (GRD), relative edge response (RER), and geolocation accuracy and are calculated for each spectral channel of the spectral image. The GRD is estimated by using the cross-section of a spatially uniform object of known size in the image and comparing it with the ground-truth size of the object to calculate the size of the

pixel projected on the ground. Spatial RER is derived by employing the MTF derived from the image. This is calculated from an image edge feature using the Fourier transform of the line spread function, obtained by differentiating the measured edge spread function. Geolocation accuracy is a measure of the ability to relate individual pixels to their absolute ground location, and requires that at least five points must have known map coordinates. Spectral fidelity metrics include the full width half maximum (FWHM) of the sensor spectral response function (spectral resolution), spectral RER, and spectral calibration accuracy. The FWHM is determined by fitting known narrow spectral atmospheric features to the recorded image spectra in order to determine the sensor spectral response. In order to calculate a spectral RER, a narrow (many times narrower than the sensor spectral response function) spectral feature is used. The ability of the measured spectrum to replicate the spectral ‘edge’ is captured in a method analogous to the spatial edge spread function and MTF measurement. The spectral calibration accuracy is calculated together with the FWHM, since the fit of the sensor response to spectral atmospheric features depends on absolute wavelengths assigned to each spectral bin, so the calibration error value is fit at the same time as FWHM. The radiometric calibration accuracy measures the error between an extracted spectral spectrum (radiance converted to reflectance) and the reflectance spectrum of the same material measured in the lab (Martin, Vrabel, and Leachtenauer, 1999).

5.2.1.2 Scene-independent

As with panchromatic imaging systems, scale and $\lambda FN/p$ are two measures of image fidelity that do not require reference objects in the scene being imaged. While scale does not have wavelength dependence, the $\lambda FN/p$ measure may be calculated with the center wavelength, λ , corresponding to the center of each spectral channel.

5.2.2 Image-derived

These approaches to image fidelity require degraded versions of a single image. A good image-derived image fidelity measure should react to the degradations, causing a decrease in image utility; it should not react if the end application of the image is insensitive to a particular degradation. These approaches are based on the digital image processing approaches to panchromatic image quality discussed in Section 3.1.2 in that they provide a measurement of the distance between the original hyperspectral image and the degraded one.

The degradations include additive white noise, spatial and spectral smoothing, and misregistration. These are intended to simulate various image chain effects.

Christophe, Leger, and Mailhes (2005) offer three categories of spectral image fidelity criteria. The first is taken from traditional image and video processing image degradation metrics which are applied directly to spectral images by performing the operations in each channel and then summing over all channels. There are eight metrics in this first group of criteria: mean squared error (MSE), root MSE (RMSE), relative MSE, maximum absolute difference, percentage maximum absolute difference, mean absolute error, SNR, and peak SNR (PSNR). These were described in Section 3.1.2. The second group of criteria addresses the spectral dimension of the images by treating them as a collection of spectral vectors. There are four criteria in this group: maximum spectral similarity, maximum spectral angle, maximum spectral information divergence, and minimum Pearson correlation. The maximum spectral similarity (MSS) between the original image, $I(x,y,\lambda)$ and the degraded image, $I_D(x,y,\lambda)$ is defined as:

$$MSS = \max_{x,y} \left\{ \sqrt{RMSE(I(x,y,\cdot), I_D(x,y,\cdot))^2 + (1 - corr(I(x,y,\cdot), I_D(x,y,\cdot)))^2} \right\} \quad (5.1)$$

where $corr$ indicates the Pearson correlation between the two images:

$$corr(I(x,y,\lambda), I_D(x,y,\lambda)) = \frac{\frac{1}{K-1} \sum_{k=1}^K (I(x,y,\lambda) - \mu_I)(I_D(x,y,\lambda) - \mu_{I_D})}{\sigma_I \sigma_{I_D}} \quad (5.2)$$

The vectors μ and σ (of dimensionality of K spectral channels) correspond to the mean and standard deviations taken over all image spatial dimensions. $I(x,y,\cdot)$ and $I_D(x,y,\cdot)$ refer to an element of the pixel vectors at x,y spatial coordinates and spectral channel λ drawn from the image and its degraded version, respectively. The maximum spectral angle is defined as:

$$MSA = \max_{x,y} \left\{ \cos^{-1} \left(\frac{\sum_{k=1}^K (I(x,y,\lambda) - \mu_I)(I_D(x,y,\lambda) - \mu_{I_D})}{\sum_{k=1}^K I(x,y,\lambda)^2 \sum_{k=1}^K I_D(x,y,\lambda)^2} \right) \right\} \quad (5.3)$$

The maximum spectral information divergence is defined as:

$$MSID = \max_{x,y} \left\{ \sum_k (\rho - \rho_D) \ln \left(\frac{\rho}{\rho_D} \right) \right\} \quad (5.4)$$

where ρ corresponds to the image or degraded image normalized by the norm of the length of the pixel vectors at each spatial location. The last group of criteria is derived from variants on Wang and Bovik's (2002) universal image quality measure, Q , from equation 3.18. These include the minimum Q between original and degraded image over all spatial coordinates or over all spectral coordinates, and the product of these.

5.3 Spectral Image Utility Measures

In the literature review, only the task performance measures discussed in Section 3.2.1 are applied to spectral image utility. While information theoretic assessment of spectral images has been examined by Aiazzi, et. al. (2001), this approach focused on the effects of compression, and not on the spectral image applications of interest for this research. Thus, we divide the spectral image task performance measures into two categories, as in Section 3.2.1: direct performance measures and performance estimates.

5.3.1 Direct Performance Measures

This group is the most commonly encountered and can be calculated directly from the image. The specific measures vary depending on the application, but all are quantified using statistical means. For classification applications, the utility is measured in terms of a misclassification error rate. For detection applications, the spectral image utility is measured in terms of the PD and PFA, and often expressed as a ROC curve. These tools were introduced in Section 4.4 for quantifying target detection algorithm performance in spectral imagery.

5.3.2 Performance Estimates

The second group for quantifying utility is based on estimating rather than calculating performance measures. NIIRS ratings are a popular means of quantifying the interpretability (utility) for panchromatic images using a performance estimate approach. The Multispectral Imagery Interpretability Rating Scale (MS IIRS) was developed to provide a structure for distinguishing varying levels of multispectral imagery interpretability. The NIIRS format serves as a framework for MS IIRS as a tool for making quantitative judgments about the potential interpretability of an image. As with NIIRS, MS IIRS is based on exploitation tasks

(criteria) which indicate the level of information that can be extracted with an image of given interpretability level. It represents a measure of the information potential (interpretability) of an image. It relies on the experience of an image analyst to imagine how well criteria would be rendered if those features were present in the image to be rated. The scale does not address hyperspectral imagery as these images are almost exclusively exploited by automated algorithms (IRARS, 1995). The criteria that compose the scale are tasks that benefit from the exploitation of spectral information and are shown in Table 5.1. MS IIRS has not gained

Rating Level 1

Distinguish between urban and rural areas.
Identify a large wetland (greater than 100 acres).
Detect meander flood plains (characterized by features such as channel scars, ox bow lakes, meander scrolls).
Delineate coastal shoreline
Detect major highway and rail bridges over water (e.g., Golden Gate, Chesapeake Bay)
Delineate extent of snow or ice cover.

Rating Level 2

Detect multi-lane highways.
Detect strip mining.
Determine water current direction as indicated by color differences (e.g., tributary entering larger water feature, chlorophyll or sediment patterns).
Detect timber clear-cutting.
Delineate extent of cultivated land.
Identify riverine flood plains.

Rating Level 3

Detect vegetation/soil moisture differences along a linear feature (suggesting the presence of a fence line).
Identify major street patterns in urban areas.
Identify golf courses.
Identify shoreline indications of predominant water currents.
Distinguish among residential, commercial, and industrial areas within an urban area.
Detect reservoir depletion.

Rating Level 4

Detect recently constructed weapon positions (e.g., tank, artillery, self-propelled gun) based on the presence of revetments, berms, and ground scarring in vegetated areas.
Distinguish between two-lane improved and unimproved roads.
Detect indications of natural surface airstrip maintenance or improvements (e.g., runway extension, grading, resurfacing, bush removal, vegetation cutting).

Rating Level 4 (Cont.)

Detect landslide or rockslide large enough to obstruct a single lane road.
Detect small boats (15 - 20' in length) in open water.
Identify areas suitable for use as light fixed-wing aircraft (e.g., Cessna, Piper Cub, Beechcraft) landing strips.

Rating Level 5

Detect automobile in a parking lot.
Identify beach terrain suitable for amphibious landing operation.
Detect ditch irrigation beet fields.
Detect disruptive or deceptive use of paints or coatings on buildings/structures at a ground forces installation.
Detect raw construction materials in ground forces deployment areas (e.g., timber, sand, gravel).

Rating Level 6

Detect summer woodland camouflage netting large enough to cover a tank against a scattered tree background.
Detect foot trail through tall grass.
Detect navigational channel markers and mooring buoys in water.
Detect livestock in open but fenced areas.
Detect recently installed minefields in ground forces deployment area based on a regular pattern of disturbed earth or vegetation.
Count individual dwellings in subsistence housing areas (e.g., squatter settlements, refugee camps).

Rating Level 7

Distinguish between tanks and three-dimensional tank decoys.
Identify individual 55-gallon drums.
Detect small marine mammals (e.g., harbor seals) on sand/gravel beaches.
Detect underwater pier footings.
Detect foxholes by ring of spoil outlining hole.
Distinguish individual rows of truck crops.

Table 5.1: Multispectral Image Interpretability Rating Scale (from Irvine, 1997)

widespread use as a means of capturing the utility of spectral images. However, it is common in the spectral image quality literature to equate performance with NIIRS. The development of spectral image utility prediction models based on the GIQE paradigm leads to expression of the utility in terms of NIIRS.

5.4 Predicting Spectral Image Utility

Shen (2003) offers motivation for the ability to predict spectral image utility, “As hyperspectral remote sensing technology migrates into operational systems, there is an urgent need to understand the phenomenology associated with the collection parameters and how they relate to the quality of the information extracted from the spectral data for different applications. If such relationships can be established, data collection requirements and tasking strategies can then be formulated for these applications.” Simmons, et. al. (2005) describe the requirements of the ideal spectral image utility predictor. An ideal utility predictor would apply to panchromatic systems as well as those with many channels, apply to any exploitation method (visual or computer-aided), would work for targets with broad spectral features (like vegetation) as well as those with narrow features (minerals or gases), would be relevant to applications where spatial detail is more significant as well as those where spectral information is the primary factor, and would be appropriate for subpixel as well as fully resolved targets. This list of characteristics will be helpful to keep in mind as we consider the current literature on the topic of spectral image utility prediction. There are four categories into which current approaches may be classified. The first is the regression approaches, which seek to predict a measure of utility from an empirically-derived expression made of image fidelity measures. The second is the spectral vector approach, which considers the ability of image spectra to satisfy certain criteria. The third seeks to combine spectral and spatial information in a manner more aligned with how an image analyst reaches conclusions. The fourth uses spectral statistics in parametric descriptions of the image and propagates these through a model of the image chain and application-specific processing algorithm to make a prediction of the spectral image utility.

5.4.1 Regression Approaches

These seek to predict image utility by using an empirically derived expression that relates sensor or image parameters to utility. In general, approaches based on a regression will specify a utility metric, analyze a large number of images with varying image fidelity parameters, and then form a regression between the utility measure and fidelity parameters. The expressions derived in this manner show utility to be dominated by spatial resolution as expressed in the GRD, with the value of the image utility measure going up with decreasing

GRD. Additionally, image utility decreases with decreasing SNR, and increases with increasing spectral channels. These results tend to confirm intuition about how the behavior of image fidelity measures relates to the predicted image utility. Regression approaches appear to be most applicable in facilitating trade studies between image utility and sensor/image parameters that describe a class of sensors or images.

5.4.1.1 General Image Quality Equation (GIQE)-like

Martin, Vrabel, and Leachtenauer (1999) build on the approach used to develop the GIQE for panchromatic imagery. They conjecture that spectral image utility may be viewed on two levels: 1) the probability of success for materials identification and 2) the usefulness of the product in solving an information requirement. Material identification is accomplished by matching a spectrum from the image with a known spectrum from a reference library. The probability of correct materials identification is a function of the accuracy of signature definition (the degree to which signatures for all possible materials are known), sensor performance (spatial, spectral, and radiometric fidelity measures), analysis system performance (the algorithms used to extract the spectral characteristics of the pixel and compare these to a spectral library), the sample abundance, and the decision criteria (the measure of required similarity between the measured and library spectra). The usefulness of the image in solving information requirements is accomplished by human image analyst performing image interpretation (understanding that it can only be done with three bands simultaneously using image-space analysis). Their approach first objectively calculates the quality of an image using a notional equation involving the image fidelity terms, then subjectively determines the utility through human analyst measurement, and finally relates these two metrics to obtain an objective utility metric via multiple regressions. The spectral image fidelity metrics described in Section 5.2 are used as the first step of the process to obtain a quantifiable measure of image fidelity.

5.4.1.2 Spectral Quality Equation (SQE)

Shen (2003) relates image fidelity parameters not to a human-derived estimate of image interpretability, such as NIIRS in the GIQE development, but rather to a direct performance measure. Her reasoning is that because spectral data is processed with algorithms that are judged in a statistical sense by PD and PFA at various thresholds, performance estimates such as NIIRS are not meaningful for spectral images. The spectral quality equation (SQE)

relates four sensor-derived (scene dependent) image fidelity parameters (spatial resolution, spectral resolution, SNR, and scene complexity) to the probability of correct detection for the object/anomaly detection application at a given probability of false alarm. Scene complexity is measured by the average (across all spectral channels) standard deviation of the image pixel values, σ . The probability of correct detection (PD) at a specified PFA is the quantitative measure for image utility in this approach. Object detection was performed by applying linear unmixing to each cube using background materials as endmembers. A threshold was applied to the resulting residual image to produce a detection map for each data cube. The thresholds were empirically chosen to attain a constant PFA for all cubes. The PFA was defined as the number of false alarm pixels divided by total number of image pixels. The associated PD was computed as the number of correctly detected objects in the scene divided by the total number of objects. A function expression relating GRD, spectral resolution ($\Delta\lambda$), SNR, and scene complexity to PD was obtained by performing multiple linear regressions with PD as the dependent variable and various functions of the four parameters as the four independent variables for two hyperspectral images, resulting in 120 data points. Least squares fitting of all the cases resulted in equations for two PFAs. The resulting expression for PFA = 1×10^{-3} has the following form:

$$P_D = 6.25339 + 0.403898 \log_{10} \left(\frac{1}{GSD^2} \right) + 0.195825 \log_{10} \left(\frac{1}{\Delta\lambda} \right) + 0.120574 \log_{10}(SNR) + 2.43465 \log_{10} \left(\frac{1}{\sigma} \right) \quad (5.5)$$

The goodness of the regression equation was judged by its multiple linear correlation coefficient, F-statistic, and RMSE between all samples and a subset of samples.

5.4.1.3 Spectral Quality Rating Scale (SQRS)

Kerekes and Hsu (2004a and 2004b) modeled their approach after the NIIRS GIQE development. A number of analyses were conducted with spectral images (output of the FASSP model for target detection scenarios) of varying image fidelity parameters, followed by the development of regression equations relating the spectral image fidelity parameters to a spectral quality rating (utility measure) for specific applications. The surface of constant utility represents the regression equations, which describe the tradeoffs between image fidelity parameters to achieve the same utility in quantitative terms. Image fidelity parameter combinations that lie on the surface described by the spectral quality equation will have

equivalent utility. By generating many scenarios, the triplet of image fidelity parameters (GRD, number of channels, and SNR) that achieved a specified PD at a fixed PFA served as a point on the constant utility surface for a single target/background pair. Figure 5.2 depicts the constant utility surface obtained in this manner. The surfaces resulting from all

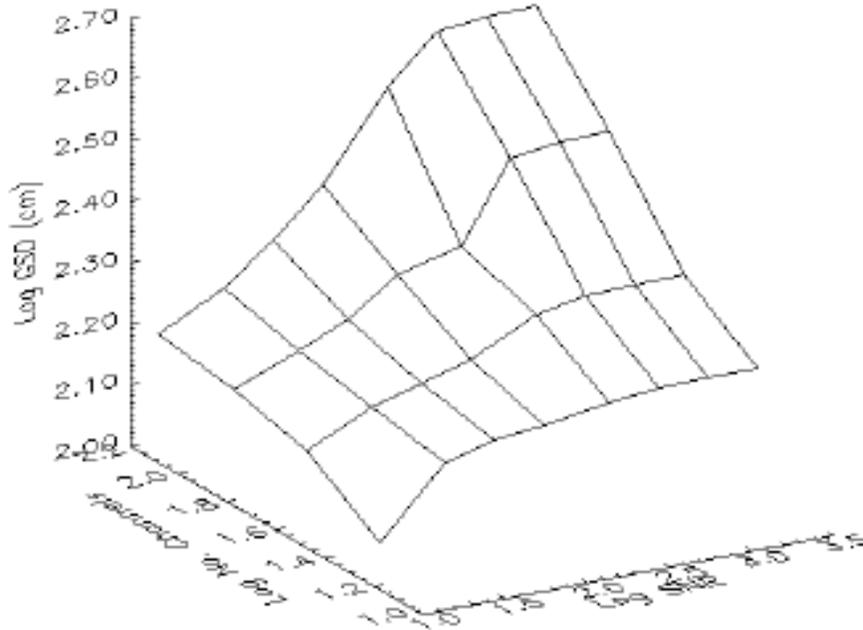


Figure 5.2: Surface of constant utility (from Kerekes and Hsu, 2004a)

target/background combinations were calculated and a regression to a linear function of the logarithmically transformed parameter values was performed. The resulting equation denoting the spectral quality rating scale (SQRS) is:

$$SQRS = 9.65 - 3.22 \log_{10}(GRD) + 0.44 \log_{10}(SNR) + 0.81 \log_{10}(K) \quad (5.6)$$

GRD is the ground resolved distance in centimeters, and K is the number of spectral channels. The equation was constrained so that if the number of spectral channels was set to one, the equation would predict a value close to that predicted by the GIQE.

The SQRS metric was adjusted to account for specific targets in target detection tasks by taking into account a measure of separability between the target and background mean vectors. This metric was called the signal-to-clutter ratio by Kerekes and Hsu (2004b) and is the seen to be the Mahalanobis distance between target and image means:

$$SCR = \sqrt{(\mathbf{t} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{t} - \boldsymbol{\mu})} \quad (5.7)$$

In this equation, \mathbf{t} is the mean vector of the target, $\boldsymbol{\mu}$ is the image mean vector, and $\boldsymbol{\Sigma}$ is the image covariance matrix. The Mahalanobis distance is a measure of similarity between a group of values with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ (both estimated from the image) and the target spectral vector. A large SCR indicates that the target vector is very dissimilar to the image parameters. This assumes that the image may be parameterized by its mean and covariance. It does not take the variability in the target vector into account. But, this is the basic form that the spectral matched filter is based upon, thus it should be very helpful in giving us insight into the behavior of a particular image and target combination. The SQRS that incorporates the SCR is:

$$SQRS_{SCR\text{-detection}} = \frac{10.6 - 1.6 \log_{10}[T] - 3.3 \log_{10}[GRD(cm)] + 1.6 \log_{10}[SCR]}{10} \quad (5.8)$$

T is the threshold on the normalized output test statistic, y , for the spectral matched filter that leads to a specified false alarm rate on the image background. Because the mean and covariance have the dimensionality of number of spectral channels, this parameter is included (in feature space). The SNR is included since the image covariance includes sensor noise as well as other sources of variability. This is different than original SQRS in that it is tuned to a specific target.

5.4.2 Spectral Vectors

Another approach to predicting utility is not to attempt to use regression equations to relate image fidelity parameters to an image utility measure such as NIIRS or PD, but to start by defining utility differently. Sweet, Granahan, and Sharp (2000) define an image of high utility as one that contains separable classes that are spectrally very similar. They propose a spectral similarity scale based on an objective methodology. Vector differences are described by distance and shape in order to quantify similarity. Distance is measured in the Euclidean sense between vectors using correlated variables and shape is measured using correlation. For any pair of reflectance spectra, these two difference measures constitute a two-element vector called the difference vector. The absolute magnitude of the difference vector is the difference magnitude, and these constitute the spectral similarity scale. The minimum value is zero, maximum is $\sqrt{2}$, and small values on this scale mean similar spectra:

$$\text{spectral similarity} = \sqrt{d_e^2 + \hat{r}^2} \quad (5.7)$$

The term d_e is the average distance between the two vectors and K is the number of spectral channels:

$$d_e = \sqrt{\frac{1}{K} \sum_{i=1}^K (x_i - y_i)^2} \quad (5.8)$$

The correlation coefficient squared compares the shapes of vectors since subtracting the means removes any bias terms and dividing by the standard deviation removes gain factors:

$$\hat{r} = 1 - r^2 = 1 - \left(\frac{\frac{1}{K-1} \sum_{i=1}^K (x_i - \mu_x)(y_i - \mu_y)}{\sigma_x \sigma_y} \right)^2 \quad (5.9)$$

Subtracting the squared correlation coefficient from one ensures that a small \hat{r} implies similar spectra.

The utility of an image is obtained by relating the minimum similarity value of the image to the similarity value required to separate specific spectra in that image. It is assumed that if the similarity value of a pair of spectra is larger than the minimum image similarity value, then, in general, the spectra could be separated. A strength of this methodology is to identify the capability of an image to differentiate spectra of interest (Sweet, 1999). This approach appears to be most useful in predicting image utility for classification tasks.

5.4.3 Spectral-spatial

In an attempt to apply image utility prediction to more than just a subset of spectral imagery applications and to better capture the situation in which fewer spectral channels may provide better utility in an application that is more spatially oriented, Simmons, et. al. (2005) propose an approach called the general spectral utility metric (GSUM). GSUM attempts to capture the spatial and spectral information inherent in the image. It is based on the expectation that spatial and spectral information will combine to give greater image analyst confidence in an assessment than from either type of information alone.

Simmons, et. al. (2005) assess the confidence of an image analyst in performing a specified task using a spectral image. They assume that spectral and spatial information are largely separable with both contributing to the overall utility of the image. As long as the

fundamental information used to determine spatial utility (size and shape) is different from that used to determine spectral utility (spectral signature) the information will not be redundant even though other factors (SNR and MTF) may have an impact on both sources of information. We currently have separate metrics that will give us spatial and spectral information regarding the satisfaction of a particular exploitation task (NIIRS for visual/spatial exploitation and PD as a confidence value that indicates the likelihood that a particular signature is that of a desired target). NIIRS values can be related to image analyst confidence for performing specific tasks (called essential elements of information). The relationship between spatial resolution units required on target to accomplish detection, recognition, and identification tasks and the confidence is described by the Johnson criteria, from equation 3.24.

The approach combines the spatial and spectral information in a common term (confidence) to give an overall confidence in performing the specified task. Combining the spectral and spatial confidence values is accomplished with the following formula:

$$C_{TOTAL} = 1 - (1 - W_{spatial}C_{spatial})(1 - W_{spectral}C_{spectral}) \quad (5.10)$$

As one approaches a confidence value, C , of 1.0, using either spatial or spectral information separately, one is likely to remain highly confident of the answer regardless of any lack of supporting information from the other side. If both confidence values are near zero, the total confidence should be near zero. Intuitively, the combined confidences that are greater than zero would be expected to be greater than either single confidence. The weighting functions, W , provide a means for adjusting confidence based on the scenario specific implications that might reduce confidence from the predicted confidence. Spectral confidence will generally not compare equitably with spatial confidence, and the weighting function is included to transform them into confidence values that can be compared to spatially-derived confidence. This approach appears to be most useful in predicting utility for a specific image, where utility is analogous to image analyst confidence.

5.4.4 Spectral Statistical Analytical Model

This approach to predicting spectral image utility is performed on a scenario-driven basis that does not attempt to generalize to any image with a single expression. FASSP has been described in Section 4.5 as a tool for predicting target detection performance of a specific

scene, sensor, and processing configuration. We also saw in Section 3.3.1.2 that FASSP was a parametric analytical model that did not generate an image per se, but a parametric version of a scenario as an input and some measure of utility as an output. FASSP can be used to predict the utility of a spectral image, and does so by operations on a parametric version of the image. This approach is simpler because it corresponds to a single run of FASSP scenario rather than many runs that are used to provide points on a constant utility surface to generate a spectral quality equation. This offers a utility prediction on a per scenario basis, though at the expense of not including generalization in a single expression. The analytical performance model approach is useful for studying the effect on utility of variations in image chain parameters for a notional imaging scenario.

5.5 Summary

This chapter followed the organization for reviewing traditional image quality metrics established in Chapter 3. Spectral image fidelity and utility measures were presented and found to be comparable to those for traditional image quality metrics, with the exception of the spectral-specific measures. The methods for assessing spectral image utility were discussed, which all involve application of computer information extraction algorithms. Methods for predicting spectral image utility were viewed as falling into one of four categories: regression based, spectral vector-oriented, spectral-spatial combined confidence, and spectral statistical analytical models.

While each of the spectral image quality prediction approaches discussed in Section 5.4 has merit and applicability in certain situations, none of the methods addresses how to predict the utility of a *specific* spectral image for subpixel target detection applications. Several of the approaches require knowledge of the collection parameters (GRD, spectral resolution, and SNR), which, in general, may not be available. Also, none of the approaches offer a method of equitably comparing the spectral image utility predictions with assessed utility. In Chapter 6, we propose an approach to predicting the spectral image utility for the specific application of detecting sub-pixel targets using the spectral matched filter. Our methodology predicts the likelihood of finding a synthetically implanted target in a target-free image in advance of applying the detector. It differs from the FASSP analytical model-based approach to prediction in that it begins with a real image rather than a statistical

description of a notional imaging scenario, thereby capturing the effects of the scene, atmosphere, and sensor in estimated statistical parameters. It is similar to FASSP in that it uses transformations of multivariate spectral statistical parameters, rather than the data itself, by the detector into parameters that describe the scalar output test statistic. Hence, the prediction assesses detector performance in a more computationally efficient manner than actually applying the filter to the data.

Chapter 6

Approach

This chapter describes the actions and methods required to accomplish each of our stated research objectives. We begin in Section 6.1 by providing a framework for image quality that is broad enough to encompass spectral imagery and the specific application of target detection. In Section 6.2, we define our image utility metric. Then we describe the general operation of the target implant image utility assessment method in Section 6.3. Section 6.4 addresses the image-derived image utility prediction method. We then turn to addressing more specific implementation details surrounding the design of the various experiments and demonstrations undertaken to illustrate the behavior of the utility metric. These are the comparison of different images using their image utilities (Section 6.5), the specifics of characterizing all aspects of image utility prediction (Section 6.6), and the implementation of methods to calculate the sensitivity of utility to image chain parameters (Section 6.7). In Section 6.8, we provide the details of the real and simulated spectral images that we use to generate the results of Chapter 7.

6.1 A Framework for Image Quality

We believe that several frameworks for considering the image quality problem may be adopted. The first and simplest is the image chain paradigm. This construct was used in Chapter 1 to discuss the concepts of image fidelity and utility. The difficulty of establishing a framework for image formation based on a simple image chain formulation is captured eloquently by O’Sullivan, Blahut, and Snyder (1998), *“Even the terms source, sensor, and image can be slippery; our understanding of these terms is closely tied to and colored by our view of a particular physical problem. It is not yet common practice to study problems of image formation in terms of an abstract formalization that is not connected to a specific physical problem.”* While the information-theoretic framework is fairly general in scope, it does not seem to adequately address all elements of the imaging process. It also does not deal with the assessment and prediction of image quality. The task-based framework proposed by Barrett and Myers (2004) is robust and leads to the objective assessment of image quality. There are four elements to their framework: 1) the specification of a task in quantifiable

terms such as a test statistic or parameter estimate, 2) the description of object classes and imaging process in order to describe the data, 3) selection of the observer, representing the means by which tasks are accomplished, and 4) the choice of a figure of merit. This framework is shown in Figure 6.1.

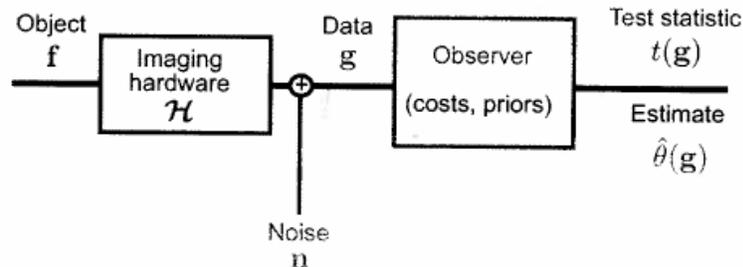


Figure 6.1: Task-based framework for objective image quality assessment (from Barrett and Myers, 2004)

The first step in applying the task-based framework to a specific imaging problem is to develop a model for the objects under investigation, both the deterministic aspects and object variability. Next, a model for the acquired data is generated, applying all knowledge of the image formation process to include the deterministic mapping of the object through the imaging system and system noise sources. An observer is selected in the next step. The ideal (Bayesian) observer uses all statistical information available regarding a task to maximize task performance, and the ideal observer provides an upper bound on observer performance. The observer applies one or more operations to the data in order to facilitate inference. In classification tasks, the output of the observer operations is a test statistic, whereas in estimation tasks, the output is an estimate of the parameter of interest. Finally, selection and application of a figure of merit indicates how well the observer performs at the task.

We believe that the task-based approach with some modifications is the most appropriate unifying framework. Figure 6.2 shows our modified task-based framework. In its basic form, the framework looks like the task-based framework of Figure 6.1. We add the semantic definitions of image fidelity and utility at the appropriate locations and explicitly show the choice between the activities of assessing or predicting image fidelity and utility. These additions are important because they allow the many image quality metrics and models to be put into the most appropriate “pigeonhole.” We further refine the framework by adding choices that are pertinent to the optical remote sensing community. These include the imaging system type (spectral or panchromatic), the type of observer (human image analyst,

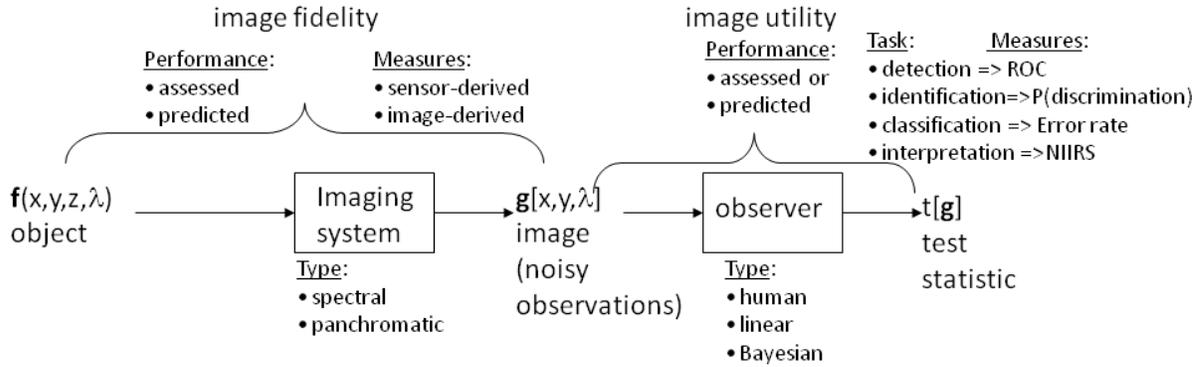
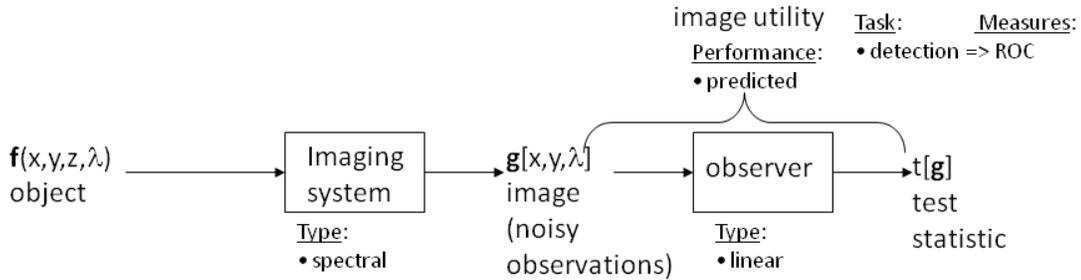


Image Quality Framework



Framework applied to Target Detection in a Spectral Image

Figure 6.2: Framework for image quality with spectral image utility shown as an example

linear, or Bayesian), task (detection, identification, classification, or interpretation), and measures for the particular task (ROC analysis, probability of discrimination, error rate, or NIIRS for utility and either sensor- or image-derived measures for fidelity). The bottom part of the figure shows the specific choices in the modified task-based framework that accommodate spectral image utility prediction. The issue of image fidelity is not addressed, because we are concerned with utility. The task is detection and we apply a linear filter that produces a scalar test statistic as the observer.

The spectral image utility prediction methods that we reviewed in Section 5.4 are addressed by this framework. The inputs to the observer are not necessarily observations of the image, but rather image fidelity parameters, such as the sensor GRD, SNR, and spectral resolution. The observer transforms these inputs into some form of utility, depending on the individual method’s definition of utility. The observer is essentially the model employed by the method to make the prediction, whether it is embodied as a regression equation, the spectral similarity calculation of equation 5.9, the semantic transform of GSUM, or the

transformation of statistical parameters into a test statistic by FASSP. Thus, the point is that the framework is nothing more than an input-output relationship in which methods of assessing utility operate on real data, whereas methods of predicting utility operate on some representation of that data (statistical parameters) or representations of the data that apply to many datasets (sensor parameters).

6.2 Defining a Spectral Image Utility Metric

As we have established in Chapter 5, although very helpful in assessing the utility of panchromatic images, the NIIRS does not capture the radiometric and spectral aspects of spectral imagery. This inadequacy is primarily due to the fundamental differences in exploiting the information inherent in each type of image. The information in panchromatic images is extracted by the perception of spatial detail and patterns by a human observer. By contrast, spectral images are processed by computer algorithms operating on the spectrum associated with each pixel individually and exploiting the statistical or subspace characteristics of the image pixel vectors to uniquely identify the material in the corresponding ground resolution cell. Thus, the interaction between the spectral, radiometric, and spatial nature of a spectral image contributes to the ultimate utility of a spectral image, and a utility metric should ideally capture this interaction.

If we adopt the perspective that image utility is another name for the performance metric of a specific application, then we arrive at a spectral image utility metric by using the measures of performance routinely employed to evaluate spectral processing algorithm performance. The simplest method to assess spectral image utility is by applying an information exploitation algorithm to the image. Thus, performance measures such as probability of detection for target detection tasks and classification accuracy for classification applications can be used for quantifying spectral image utility. We note that there is no analogous NIIRS for spectral imagery (aside from the multispectral image interpretability rating scale, which does not apply to the higher dimensionality hyperspectral imagery).

In our application area of interest, subpixel target detection, the performance of an algorithm operating on a specific image for a given target is typically described using the receiver operating characteristic (ROC) curve. The ROC curve expresses the probability of detection for a range of probabilities of false alarm. The probabilities plotted in the ROC

curve are derived from the output of the target detection algorithm. The ROC curve will serve as the point of departure for developing our spectral image utility metric. Our early approach to defining utility was to sample the ROC curve by using the probability of detection at a specified probability of false alarm. This is intuitively pleasing because it addresses the analyst’s question, “At a given probability of false alarm (or the number of false alarm pixels that we are willing to accept), what will the probability of detecting the target in this image be?” Figure 6.3 shows this concept using a notional ROC curve. While

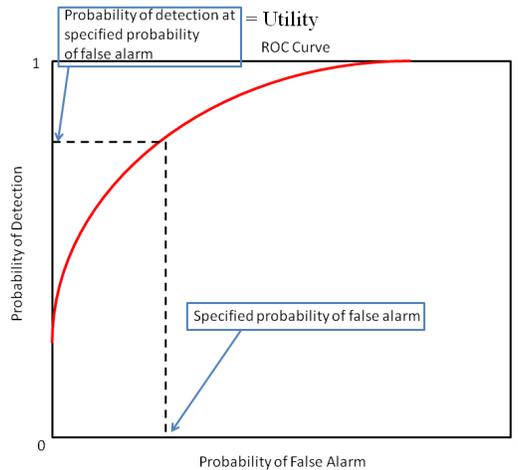


Figure 6.3: Probability of detection at specified probability of false alarm as a utility metric

this definition of utility is simple, it has a major shortcoming. It does not capture the shape of the ROC curve, which is an important indicator of how the target detectability grows as a function of probability of false alarm. We could have a situation in which two very differently shaped ROC curves attain the same probability of detection at false alarm. Suppose that one is convex-shaped and the other is concave-shaped in the PFAs between 0 and the specified PFA. If we were to examine a probability of false alarm lower than our specified PFA, we would discover that one utility metric using our original definition is lower than the other. Clearly, we would like to be able to capture the ROC curve shape as a means of comparing the ROC curves. We do not want to have to sample the ROC curve at multiple PFAs in order to form such utility metric. Our solution to this problem is depicted in Figure 6.4, which shows the definition of spectral image utility adopted throughout this research. The figure shows that we are interested in the area under the ROC curve up to a specified PFA as our metric of image utility. The area captures the distinctive shape of a ROC curve and is a statement about the detection performance at a range of PFAs smaller

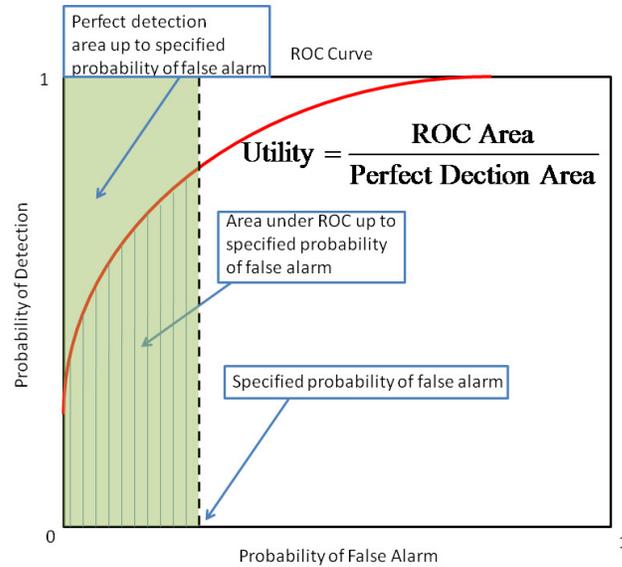


Figure 6.4: Definition of utility using the normalized area Under the ROC curve

than the specified PFA. Note that the metric is normalized by the area associated with the perfect detection (1.0) up to the specified PFA. This normalization allows us to compare the areas under different ROC curves on a scale between 0 and 1.

The term “image utility” quantifies the ability of an image to satisfy performance requirements for a well-defined task. More generally, it describes the ability of the image to deliver information about the object being imaged. Thus, any metric of image utility is fundamentally an application performance measure. The subtle difference between an image utility metric and an application performance measure is based on one’s perspective: image utility focuses on labeling a specific image’s usefulness in a particular application whereas a performance measure focuses on describing the performance of the application across many images.

6.3 Assessing Spectral Image Utility

Having established the definition of the appropriate utility metric in the previous section, we now address the issue of how to assess the utility metric and summarize it in a meaningful fashion to facilitate the assessment of image utility by image analysts. Specifically, we examine the target implant method of creating a binary hypothesis test for assessing the utility, the various implementations of detection algorithms in this scheme, and the means of summarizing the utility metric. The process of *assessing* the image utility simply implies that

some information exploitation algorithm is directly applied to the image in order to glean the desired information from the image. The important question is the means by which we generate the ROC curve needed to give us the information about arbitrary images.

The assessment of spectral image utility is accomplished by applying an information extraction algorithm to the image and noting the results. Because the MS IIRS is not used in the spectral image exploitation community, it is not realistic to say that the interpretability of a spectral image can actually be assessed by an image analyst. Utility assessment was described in Section 4.4 for the assessment of target detection task performance. More generally, application-specific algorithms produce the direct performance measures described in Section 5.3.1.

We note that the target implant method of assessing target detection performance is a method of assessing spectral image utility which we will make use of in our experiments. This method was developed S. Rotman and introduced in Cafer, et. al. (2002). The idea is to synthetically implant a target spectrum in every pixel of the image in order to generate the target present hypothesis.

6.3.1 Target Implant Method

In order to generate a ROC curve, the output of the detection filter is needed for pixels that contain the target and pixels that do not contain the target. These two situations may be called the target present and target absent cases. However, in practical application to arbitrary images, the target present case will be unattainable unless we have ground truth about the target locations in the image. Further, if the target is only present in a small number of image pixels, then the estimate of the probability of detection will be unreliable. Finally, if the target is not present in the image, then we *do not have* a target present case and cannot evaluate detection performance. Given our goal of assessing the utility of *any* image, these constraints must be considered the norm. In order to overcome these challenges, we adopt an idea from Cafer, et. al. (2007) to *fractionally implant* a target spectrum in every spatial pixel of the image in order to generate a target present case. We call this method of assessing image utility for subpixel target detection applications the *target implant method*. It is based on the assumption that subpixel targets will mix linearly with background pixels.

The target implant method poses the question “What if every pixel in the image

contained the same subpixel target?” to generate the target present hypothesis. While this may be improbable from the perspective of considering where targets might truly be found in images, from the perspective of evaluating the utility (the usefulness for detecting a target) of an image, it creates a ROC curve that represents a summary of the overall detectability of the implanted target across the entire image. This is exactly the characteristic that makes this method so appropriate as a means of assessing image utility. The target implant method offers a flexible means to assess the performance of detection algorithms for a wide variety of targets and enables the calculation of a spectral image utility metric for images with no targets, limited targets, or no target ground truth. Figure 6.5 provides an overview of the target implant method for assessing spectral image utility in the target detection task. We see

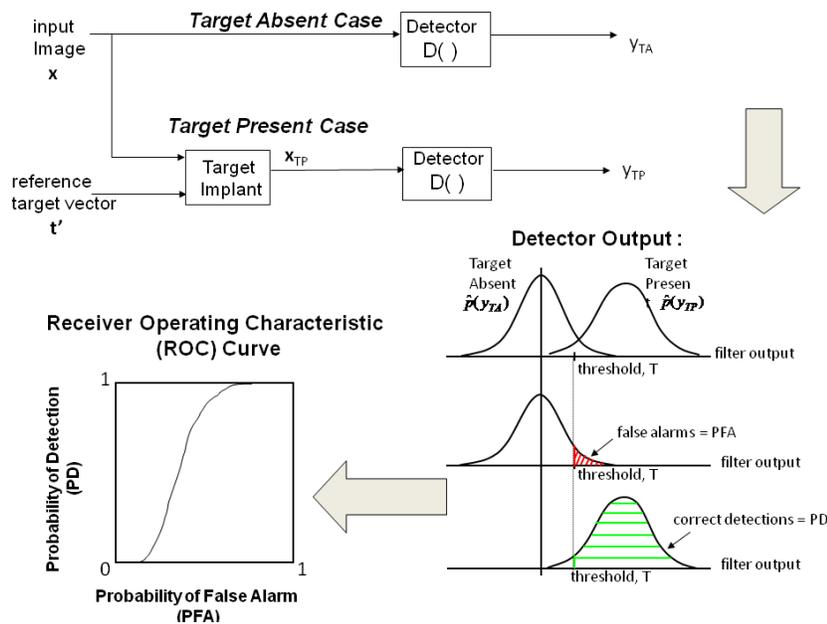


Figure 6.5: Target implant method of assessing spectral image utility

that the detection problem is posed as the classical binary hypothesis test. The top left part of Figure 6.5 illustrates the operations necessary to obtain the filter output for each of the hypotheses (cases). Let the reflectance domain spectral image have K spectral channels, so that a pixel at a given spatial location in the image is represented as the $K \times 1$ vector, \mathbf{x} . We create the target absent case by applying the arbitrary target detector operator, D , to each pixel of the image:

$$y_{TA} = D(\mathbf{x}) \tag{6.1}$$

The scalar y_{TA} is the output in the target absent case at one spatial pixel location.

The target present case is created by applying the detector to each spatial pixel location of the original image in which the $K \times 1$ target vector has been implanted, resulting in the scalar output y_{TP} :

$$y_{TP} = D(\mathbf{x}_{TP}) \quad (6.2)$$

The implantation of the target is accomplished in a fractional manner (replacement, not additive) using the subpixel target implant fraction f in each pixel of the image as:

$$\mathbf{x}_{TP} = f\mathbf{t}' + (1-f)\mathbf{x} \quad (6.3)$$

The $K \times 1$ implanted target pixel vector \mathbf{x}_{TP} is created by first realizing a $K \times 1$ random vector \mathbf{t}' from a normal random process described by the statistical parameters of the $K \times 1$ target mean vector, \mathbf{t} , and $K \times K$ covariance matrix, Σ_T . These parameters are drawn from a reference library created by careful collection of known target pixel vectors from HYperspectral Digital Imagery Collection Experiment (HYDICE) imagery exhibiting unimodal normal statistics as described in Kerekes and Baum (2002). This target vector variability is a departure from the target implant method described in Cafer, et. al. (2007), which treats the target deterministically rather than stochastically. This modification has been adopted to better reflect the variable nature of target vectors. The random target vector \mathbf{t}' is then mixed *fractionally* with every data pixel to the specified subpixel mixing fraction f as shown in equation 6.3.

The subpixel mixing fraction may be given a physical meaning because it represents the ratio of target to image pixel area projected to the ground. If we assume a square target and square image pixels, then the fraction may be expressed in terms of the linear dimension of the target, l , and the linear dimension of the ground resolved distance (GRD).

$$f = \frac{l^2}{GRD^2} \quad (6.4)$$

We adopt this interpretation of the target implant fraction because it is intuitive to the analyst to think of a target in terms of a physical size rather than a fraction of a pixel. Also, having a target size allows direct comparison of image utilities between images having different GRDs.

The right side of Figure 6.5 shows the target absent and present probability density functions (PDFs), $\hat{p}(y_{TA})$ and $\hat{p}(y_{TP})$ respectively, which are estimated from the detector output histograms for the entire image. Integrating the PDF between a given threshold value,

T , and infinity yields a probability. The integral of the target absent PDF is called the probability of false alarm (PFA), representing the probability that a pixel will be classified as a target when it is really not, and the integral of target present PDF is called the probability of detection (PD), representing the probability that a pixel is correctly classified as a target.

$$PFA = \int_T^{\infty} \hat{p}(y_{TA}) \quad PD = \int_T^{\infty} \hat{p}(y_{TP}) \quad (6.5)$$

In target detection, it is the relative relationship of these two probabilities which expresses the goodness of a particular detector or detection scenario, with the most desirous situation being one in which a high PD is achieved at low PFA. The ROC curve, shown on the bottom left side of Figure 6.5, captures this relationship. The ROC curve is obtained by plotting PD against PFA for each threshold setting. A perfect ROC curve would consist of a PD of 1 for all PFAs, a situation created by target absent and present PDFs with no overlap. At the other extreme, a useless ROC curve would consist of a straight line between the PD, PFA pairs (0,0) and (1,1), and correspond to complete overlap of the target absent and present PDFs.

Using the target implant method to assess the utility of an image, the resulting ROC curve is a statement about the overall detectability of the implanted target across the entire image. While the target implant method addresses concerns about the lack of ground truth and limited target pixels in the image and offers a flexible means to assess the performance of detection algorithms for a wide variety of targets, there are some requirements and limitations that should be highlighted. The method requires a reflectance image in order to employ the reference spectra from the target library. This requirement could be alleviated by using the radiance image and forward propagating the reference target spectrum into radiance space using a radiative transfer model of the atmospheric conditions at the time of image acquisition. We choose to work in the reflectance space, since this is the most straightforward approach. Assuming that we work in reflectance space, the target reflectance spectrum needs to be resampled to match the spectral characteristics of the image. If the image characteristics are unknown, or if the image suffers from distortions such as artifacts, then the assessed utility using this method will be inflated, since these spectral distortions cannot be applied to the target spectrum. This is because the target present case is representing a situation in which an undistorted reference target is implanted into a distorted image, making the detection easier, and thus resulting in a higher image utility. Thus, the requirement is that the

reflectance spectral image be as spectrally accurate as possible to allow matching with the reference spectrum. Another limitation is that utility will be underestimated in an image which contains significant amount of target pixels. This is due to the impact of the early false alarms that the target pixels will cause, which will degrade the resulting utility.

6.3.2 Subpixel Target Detection Algorithms

Obviously, a key parameter in the assessed utility of a spectral image is the specific detection operator, $D(\cdot)$. The choice of detector is driven by image analysis requirements and performance will be determined by target and background variability, pixel composition (pure or mixed), and how the detector accounts for these factors. In the spirit of affording maximum flexibility to the analyst in defining utility, we incorporate three detectors into the assessment methodology. The detectors require little a priori knowledge about the image, but do require that a reference library containing a target vector is available.

The first detector we consider is the spectral matched filter (SMF), because it is linear, reliable, and simple. The SMF is similar to the constrained energy minimization (CEM) filter of Farrand and Harsanyi (1997), but differs in that it uses the covariance matrix and subtracts the image mean vector, $\boldsymbol{\mu}$, from the target and background vectors. This detector is constructed with the $K \times K$ inverse image covariance matrix $\boldsymbol{\Sigma}^{-1}$ and a $K \times 1$ target mean vector \mathbf{t} from the spectral library:

$$D_{SMF}(\mathbf{x}) = \frac{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{t} - \boldsymbol{\mu})}{(\mathbf{t} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{t} - \boldsymbol{\mu})} \quad (6.6)$$

The filter operates on each image pixel (the $K \times 1$ pixel vector \mathbf{x} with the data mean vector subtracted), and creates a scalar result representing the relative degree to which the pixel vector matches the target vector.

The second detector is the nonlinear adaptive coherence/cosine estimator (ACE), which is derived from the generalized likelihood ratio test approach from Manolakis (2005).

$$D_{ACE}(\mathbf{x}) = \frac{\left((\mathbf{t} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^2}{\left((\mathbf{t} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{t} - \boldsymbol{\mu}) \right) \left((\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)} \quad (6.7)$$

This equation is the same as equation 4.11 except that the signal vector is denoted as \mathbf{t} and the image mean, $\boldsymbol{\mu}$, is subtracted from both target and data pixel vectors. The numerator of this detector is the squared Mahalanobis distance (which is taken to be a square distance, as

discussed in Section 4.3.2) between the demeaned image pixel and the demeaned mean target vector. Both the SMF and ACE employ this distance, but the action of the denominator leads to different interpretations for each detector. Manolakis (2005) shows that in a whitened space, the SMF represents the distance between an image pixel vector and the target subspace whereas the ACE may be thought of as the angle between them.

Unlike SMF and ACE, the third detector is not designed to account for the variability of the image background or target. It treats the image and target vectors deterministically and assumes that they are spectrally pure. It is also simple and may be applied without any estimation of image statistics. It is the popular spectral angle mapper (SAM) of Yuhas, et. al. (1992) expressed here as:

$$D_{SAM}(\mathbf{x}) = \cos^{-1} \left(\frac{\mathbf{t}^T \mathbf{x}}{(\mathbf{t}^T \mathbf{t})^{1/2} (\mathbf{x}^T \mathbf{x})^{1/2}} \right) \quad (6.8)$$

The output of the detector is the angle in radians between the image test pixel and the mean target vectors. Note that the image mean vector is not subtracted as in SMF and ACE. We expect that SAM will have difficulty in our application of subpixel target detection, but include it as a reference baseline. In typical applications of SAM, the output of SAM is between 0 and 1, and a small angle between target and test pixel vectors implies that the two vectors are very similar. In our implementation, we subtract the actual SAM output angle (converted to degrees) from 180 so that we create a target absent situation that has a smaller mean than the target present situation. In this manner, the output of SAM is consistent with that of ACE and SMF.

6.3.3 Summary Utility Metric

The target implant method of assessing utility offers great flexibility because a wide variety of target, target implant fraction (target size), and algorithm choice combinations may be considered for a single image. Each unique combination represents a different *target detection scenario* and results in a unique ROC curve, which summarizes the probability of detecting the specified target over all false alarm probabilities across every pixel of the image. In this manner, each ROC curve describes the overall utility of the image for a particular target detection scenario. Because the ROC curve is generated by the target

implant method, which gives independent and equal consideration to every image pixel, it is a desirable means of assessing the utility of many images on a consistent basis.

Our stated goal is to assess the utility of any spectral image for the subpixel target detection application, which implies that we would like to consider more than just one target detection situation in arriving at a utility metric. The target detection parameters that we control in the target implant method are the specified PFA at which we want to operate the detection threshold, the particular target we seek, the amount of target implanted in each pixel (which may be considered to be the linear dimension of a square target), and the detection algorithm we use. We seek a means of summarizing a range of target detection scenario parameters into a single metric that will describe how useful a particular image is for the subpixel target detection task.

In order to summarize utility over a range of PFAs, we employ the area under a ROC curve over a PFA interval of interest. The area under curve (AUC) that we introduced in Section 6.2 is a widely used figure of merit for detection performance in medical diagnostics applications (Metz, 1986). Recall that we normalize this area by dividing it by the area under the perfect detection scenario. In target detection for spectral images, the operational range of PFA will tend to be small, since only a handful of false alarms are acceptable amongst a relatively large number of image pixels. In order to better represent this desire to operate at low PFAs, we apply a weighting function $z(PFA)$, to the calculation of the AUC. A simple weighting function is a rectangular window that applies equal weight to all PFAs in a desired operating range (from the lowest achievable PFA to the specified PFA) and zero elsewhere. The lowest achievable PFA corresponds to a single false alarm, and the probability associated with this is the reciprocal of the number of image pixels. Any weighting function may be applied to the ROC curve during the utility calculation to meet the specific requirements of the particular target detection scenario, offering great flexibility. The integral version of the utility is represented as:

$$Utility(\mathbf{t}, f, D(\mathbf{x}), PFA) = \frac{\int_0^1 z(PFA) \cdot PD(\mathbf{t}, f, D(\mathbf{x}), PFA) dPFA}{\int_0^1 z(PFA) dPFA} \quad (6.9)$$

The $z(PFA)$ weighting function is applied to the ROC curve and perfect detection area and defines the range of PFAs that we would like to emphasize in the utility calculation. Besides the rectangular window from low to specified PFA, other weighting functions that we investigate are the decreasing exponential, the Gaussian, and a Dirac Delta function that will pick out the PD associated with a specific PFA (which yields our original utility metric of PD at PFA shown in Figure 6.3). Note that the utility in equation 6.9 is a function of the target type, \mathbf{t} ; implant fraction, f ; detector, $D(\mathbf{x})$; and the PFA by virtue of PD dependence on these factors. We emphasize that the assessed utility is a unique combination of these target detection parameters, and will spend considerable time in Chapter 7 showing how the interaction of these parameters produces different image utilities. In practice, a discrete approximation to equation 6.9 is calculated.

Using equation 6.9 we may generate the utility associated with our PFA range of interest for one target, one fraction, and one detector. If we vary the target, fraction, and detector parameters over a range to create a more robust target detection scenario, we would like to be able to summarize the utility over this range in order to make a statement about the generalized utility of the image. In order to give maximum flexibility to the analyst in defining utility based on their unique requirements, we offer the ability to condense the utilities resulting from a range of parameters to a single summary metric or to leave the ROC areas as an ensemble from which the analyst can select parameters of most interest. The simplest approach to attaining a summary metric is to form a weighted summation of utilities over all L targets, M implant fractions, and N detectors. This situation for the summary utility, U , is shown as:

$$Utility_{summary}(Utility(\mathbf{t}, f, D(\mathbf{x}), PFA)) = \sum_{k=1}^N w_k \sum_{j=1}^M w_j \sum_{i=1}^L w_i Utility(\mathbf{t}_i, f_j, D_k(\mathbf{x}), PFA) \quad (6.10)$$

The summary utility metric is a function of the utility for the combination of detection scenario parameters, which have been explicitly indexed using i for the target, j for the implant fraction, and k for the detector. Selection of appropriately normalized weighting functions w_i , w_j , and w_k , to achieve the desired selectivity in targets, fractions, or detectors of interest allow the analyst to tailor a utility metric best suited to his or her purposes. In this research, we only use simple rectangular windows since these weighting functions to give us

the average value of several parameters, although any type of weighting function can be employed, as we discussed with the specified PFA weighting function. The final result is a single scalar that represents the degree of utility of a particular image in fulfilling analyst-defined target detection scenario(s).

6.4 Predicting Spectral Image Utility

There are several methods of predicting the utility of an image, as we saw in our discussion of current approaches to predicting spectral image utility in Section 5.4. While each method has its strengths and applicability, there are also characteristics which detract from their usefulness in predicting utility based on our definition. The regression methods rest upon using scalar sensor- or image- derived parameters to yield a scalar to quantify the utility of the image based on some sort of spectral quality equation. This requires something to be known about the collection parameters (GRD, spectral resolution, and SNR), which, in general, may not be available. In addition, even if the only scene dependent parameter is the scene standard deviation averaged over all bands, as with the SQE method, the equation is not built using the same information about the image that the spectral matched filter will be using. In most cases, the parameters used for prediction are sensor-derived, although the scene is the significant driver of algorithm performance for target detection tasks. The spectral vector approach uses the magnitude and angle difference between two spectral vectors to characterize the utility of an image, but these are not what SMF is based on. The spectral statistical analytical model approaches are used to predict the performance for a target in a notional scene using statistical descriptions of that scene. The problem is that these scene descriptions are not associated with the real image for which we are seeking to predict the utility.

We have two requirements to establishing a spectral image utility prediction methodology. The first is that we need to be able to compare predicted utility directly to the target implant method assessed utility in order to determine the accuracy of our predictions. Our assessed utility is specific to the image in question, so must our predicted utility. The second is that we seek to predict image utility in a manner that is easier than assessing the utility. When we use the target-implant method of assessing image utility, we must run the target detection algorithm for every pixel in the image twice (once for each of the two

hypotheses). This could consume significant computing resources depending on the number of spatial pixels and the length of the spectral vectors involved.

6.4.1 Image-Derived Prediction Method

The above two requirements lead to an approach we term the image-derived utility prediction approach. This approach performs the same transforms and basic operations as the target implant method of utility assessment using the binary hypothesis test approach. Rather than operate on the actual image pixel vectors, the utility prediction is made by operating on statistical parameters that represent the image. The image-derived approach to spectral utility prediction uses the first- and second- order statistics of the image as the parameters that will be transformed. The parameters can be used to completely describe a distribution of pixel vectors assumed to have normal distributions and, with the addition of an additional parameter, elliptically contoured T-distributions that are more applicable to complicated spectral imagery.

By using the image-derived statistical approach, we fulfill the two requirements enumerated above, and we see four benefits in adopting this approach. First, we directly compare the prediction with the assessment. Second, we know that our predicted utility is specific to the image in question because it takes the scene composition into account. Third, we inherit the sensor-derived parameters when we estimate the statistics directly from the image. The scene and atmosphere effects are also included in our estimated parameters, relieving us of the need to know these about each image. Fourth, since we are dealing with the SMF, which is not based on spatial information, we may use purely statistical parameters to describe the image, since they are independent of spatial information in the image.

The basic flow of the image-derived statistics image utility prediction method is shown in Figure 6.6. Note the similarity in the binary hypothesis test structure of the target-implant method in Figure 6.5. In calculating the predicted image utility, three statistical models are employed to describe the distribution of the scalar test statistic at the output of the filter. The predicted target absent and present distributions at the filter output are then transformed into PD and PFA using CDFs in the same manner as the assessed utility described in Section 6.3. The assumption is that the statistical distribution of the image is linearly transformed by the action of the filter to a scalar distribution. Thus, we must use

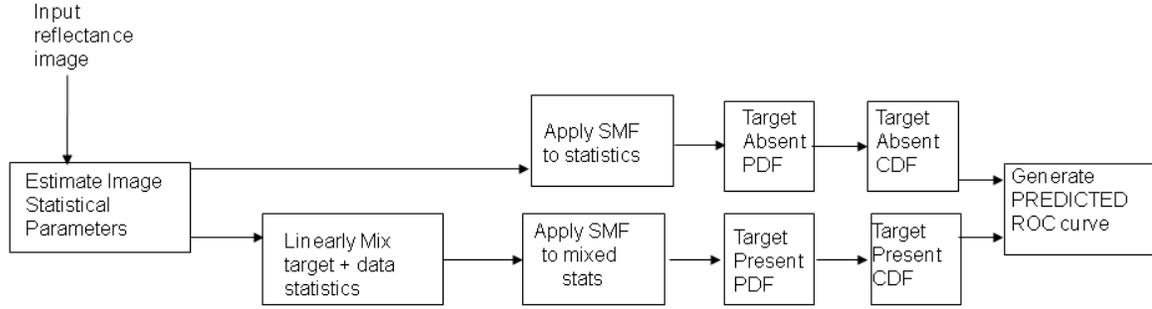


Figure 6.6: Flow diagram for image-derived spectral image utility prediction

detection filters that are linear, such as the SMF in the utility prediction. In this research, we do not implement a methodology to predict the utility for situations where nonlinear detectors such as the ACE and SAM. This would involve the transformation of a random variable using a nonlinear function.

6.4.2 Statistical Models

Three statistical models are employed with which to generate the scalar distributions that model the actual assessed target absent and target present filter outputs. The first statistical model is a Gaussian having the parameters described by the transformed mean and covariance of the entire (global) image. The second model is created by a weighted sum of the individual spectral class test statistic distributions. In the equations that follow, the global Gaussian model is treated as a special single class case of the class sum model, in which the subscript i refers to the spectral class number. The transformation from multivariate to scalar statistics for the target absent case of this model is:

$$\begin{aligned}\theta_{TA_i} &= \mathbf{w}^T (\boldsymbol{\mu}_i - \boldsymbol{\mu}) \\ \sigma_{TA_i}^2 &= \mathbf{w}^T \boldsymbol{\Sigma}_i \mathbf{w}\end{aligned}\tag{6.11}$$

where θ_{TA_i} is the target absent distribution scalar mean, $\boldsymbol{\mu}_i$ is the multivariate class mean, $\boldsymbol{\Sigma}_i$ is the class covariance, and $\sigma_{TA_i}^2$ is the target absent distribution variance. The test statistic output distribution is then the weighted sum of Gaussian distributions, each parameterized by θ_{TA_i} and $\sigma_{TA_i}^2$. The target present situation is created by fractionally mixing the mean vectors and covariance matrices of the target and background based on the target implant fraction, f , and then applying the matched filter vector \mathbf{w} to the target present multivariate statistics to obtain their scalar versions:

$$\begin{aligned}\boldsymbol{\mu}_{TP_i} &= f\mathbf{t} + (1-f)\boldsymbol{\mu}_i \quad \rightarrow \quad \theta_{TP_i} = \mathbf{w}^T (\boldsymbol{\mu}_{TP_i} - \boldsymbol{\mu}) \\ \boldsymbol{\Sigma}_{TP_i} &= f^2\boldsymbol{\Sigma}_t + (1-f)^2\boldsymbol{\Sigma}_i \quad \rightarrow \quad \sigma_{TP_i}^2 = \mathbf{w}^T \boldsymbol{\Sigma}_{TP_i} \mathbf{w}\end{aligned}\tag{6.12}$$

Vector $\boldsymbol{\mu}_{TP_i}$ is the i^{th} class mean mixed with the FASSP reference library target mean vector, \mathbf{t} , θ_{TP_i} is the target present distribution scalar mean, $\boldsymbol{\Sigma}_{TP_i}$ is the i^{th} class covariance mixed with the target covariance, $\boldsymbol{\Sigma}_T$, and $\sigma_{TP_i}^2$ is the target present distribution variance.

The third statistical model uses the elliptically contoured T-distribution to represent a scalar test statistic PDF of the form:

$$f_t(y; \theta, \sigma^2, M) = \frac{\Gamma(\frac{1+M}{2})}{\sigma\sqrt{\pi M}\Gamma(\frac{M}{2})} \left[1 + \frac{1}{M\sigma^2} (y - \theta)^2 \right]^{-\frac{1+M}{2}}\tag{6.13}$$

in which Γ is the gamma function, M is the degrees of freedom (DOF) parameter, θ and σ^2 represent either the target absent or target present parameters from equations 6.11 and 6.12, and y is the test output threshold variable. In the scalar filter output space, the elliptically contoured T-distribution is noteworthy because the degree of freedom parameter has the ability to control the heaviness of the distribution tails. Target detection filter output distributions using real spectral imagery typically display heavy tails, which are not adequately represented by a normal model. In section 7.2, we discuss methods to estimate the optimal degree of freedom that will produce minimal error between predicted and assessed utilities.

The linear transformation from multivariate to scalar dimensionality preserves the Gaussian and elliptically contoured distributions. The question of interest in this research is how closely we can model the actual filter output with an analytical form described by Gaussian and elliptically contoured distributions using the estimated parameters. We know that a global Gaussian distribution is not going to be the best model to describe a complicated hyperspectral scene. Performing a summation of individual class Gaussian distributions may get us closer to the actual complexity.

We introduced another model for the predicted utility after making some observations about the behavior of the utility at different implant fractions. At 100% fractional implant, the target present situation in equation 6.12 will have the mean and covariance of the target, which is normal, whereas at very small implant fractions, the target present distribution will

look very much like the target absent distribution. We also observe that the T-distribution typically does the best job at modeling the target absent distributions. Using these observations, we created a composite model of the target present CDF, PD_C , based on a linear mixing of the normal target present CDF, PD_N , and the T-distribution target present CDF, PD_T according to:

$$PD_C = f \cdot PD_N + (1-f) \cdot PD_T \quad (6.14)$$

6.5 Experiments to Demonstrate the Viability of the Utility Metric

We describe the experimental design needed to demonstrate the viability of our utility metric and our approach to assess the utility of images. We begin with the simple and move to the more complicated. First, we want to demonstrate the formation of the utility metric and some of the simple characteristics of assessed image utility. Our image must be in reflectance space in order for us to employ the reference library target spectra descriptions. In general, we will not have knowledge of the image collection parameters needed to apply a radiative transfer model to forward propagate the target into radiance space. We do not explore the utility calculation in radiance space in this research, although doing so would be a logical direction for further investigation. We must have enough knowledge of our image spectral characteristics in order to resample the 10 nm spectral resolution of the reference target spectrum to match the image spectra. In the case of multispectral imagery, we must apply the spectral response function of the sensor to the target spectrum and its statistics. In order to make comparisons between images of different GRDs, we would like to know the GRD of the image for which we are assessing the utility so that we can express the target implant fraction as a target linear dimension.

Having established the ability to compute a utility for an image, we then look at different operating points and where the unique combination of target type, target size, specified PFA, and detector type places the utility in the vast space of possible utility values. We will work with a baseline of five targets selected from the FASSP reference library, representative of the range of detection difficulty. We choose to consider several (usually ten) target sizes that cover the fractional implant of 10% to 100% for a given image. We will usually operate with a specified PFA of 5×10^{-4} in all utility calculations, but explore

variations in this between 5×10^{-5} and 1×10^{-3} when looking at the impact of specified PFA on the utility. Finally, we observe the utility associated with the SMF, ACE, and SAM algorithms, but will most often use the SMF as the baseline detection algorithm.

We increase the complexity of the parameters in the next set of experiments. Instead of considering the results associated with single values of parameters, we look at two parameter settings and observe the effect on utility. This is done for the scene composition, GRD, spectral resolution, and SNR. These parameters require manipulation of existing images to achieve the effect of different images resulting from the different parameter settings. We make use of real and simulated images to assist with this investigation.

The previous two paragraphs described the examination of utility resulting from static cases of single parameter values. In the next series of experiments, we consider a more dynamic situation in which a range of parameters (target type, target size, specified PFA, and detection algorithms) are considered and the results are summarized using weighting functions that provide summary utilities resulting from several parameter variations. These are the situations which are more realistic for an analyst, since it is the exploitation of specific information that is the goal of the image analyst.

In the last experiment, we consider the utility associated with several different spectral images. We apply the methods of Section 6.3.3 to summarize the utility over four target types and the ten target sizes. We only employ the spectral matched filter in these utility assessments. We look at different images made with different sensors. Then we look at images of different scenes made by the same sensor to investigate the role of the scene composition. Finally, we examine the utility of the same scene made by different sensors to isolate the effect of image acquisition parameters on the utility.

The goal throughout these experiments is to show that the utility metric and the target implant method provide us with assessed utilities that meet with our expectations and intuition. Because we cannot analytically predict detection performance, we use this empirical approach to learn more about the behavior of the utility metric.

6.6 Experiments to Characterize the Accuracy of Predicted Spectral Image Utility

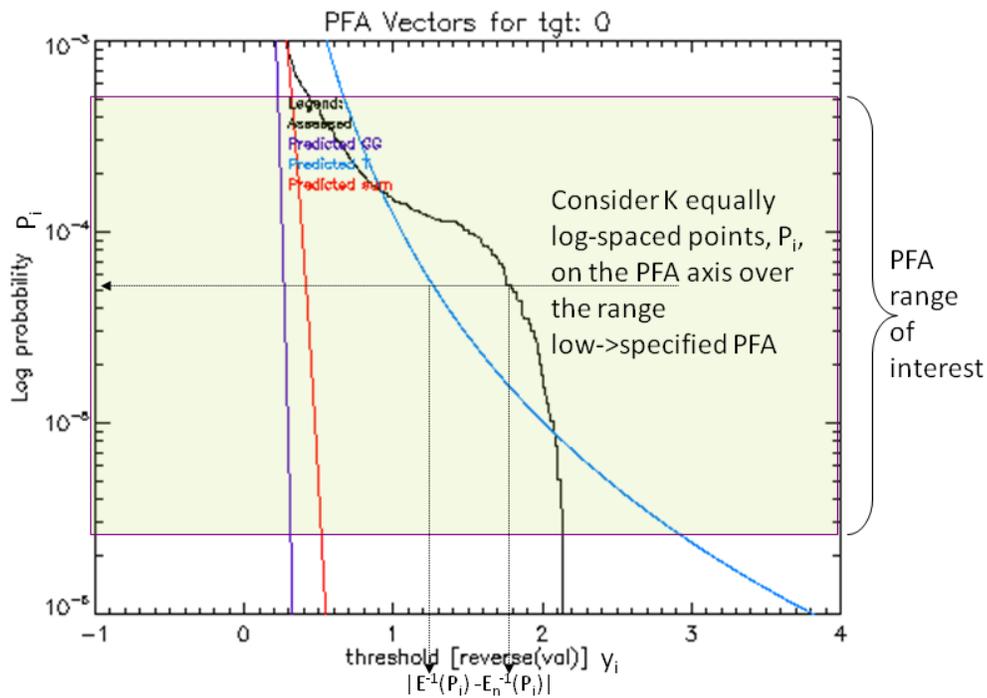
Another of our research objectives is to explore how closely our image-derived prediction of utility matches the target implant method assessment of utility. We do this in Section 7.2 using three real images. We establish the basic behavior of the predictions using the target absent and present distributions, ROC curves, and plots of utility versus target size, as we did with the assessed utility. We then measure the goodness of the utility predictions relative to the assessments. In order to do this, we need an appropriate metric to quantify the goodness of fit between the predicted and assessed utilities. This is presented in Section 6.6.1. In Section 6.6.2, we discuss how we will examine the sensitivity of the prediction relative to the assessment as model parameters are varied. In general, we want to compare the goodness of the prediction stemming from each of the three background models as a function of their parameter sensitivity. We vary the number of classes used for the sum of Gaussian class background model, the number of degrees of freedom used in the elliptically contoured T-distribution model, and the sample support used to estimate parameters in the normal model. Finally, though not directly pertaining to the accuracy of the image-derived prediction approach, we seek to put our method in the context of other spectral image quality methods. Section 6.6.2 describes how we implement a fair comparison between these methods.

6.6.1 Metrics for Prediction Accuracy and Performance

In quantifying the accuracy of the utility predictions relative to the target implant method of utility assessment, we have three options as to where we can make the comparison. We can do it at the filter output distributions, at the CDFs that describe the probabilities up to certain thresholds, or once the utility metrics have been calculated. We choose the later two, because the probabilities that constitute the ROC curve are insightful and the utilities are of primary interest to us.

At the probability level, we employ a metric presented by Manolakis and Marden (2004) called the *exceedance metric*. We use the exceedance metric as a measure of distance between our predicted CDFs and the assessed CDFs for both the target absent and target present situations, and apply the exceedance metric in selecting the optimal degree of freedom parameter for the T-distribution model.

The plot in Figure 6.7 is the inverse CDF plotted against the threshold associated with the target absent probabilities (probability of false alarm) depicted as the filter output



Single value of exceedance metric = Absolute error between assessed and predicted threshold values for a given probability, P_i

Figure 6.7: Exceedance metric

threshold, y , is swept from right (large) to left (small) values. The inverse CDF is called the probability of exceedance, $E(y)$, and is defined as:

$$E(y) = 1 - F(y) \tag{6.15}$$

where $F(y)$ is the CDF. In Figure 6.7, the threshold values, y_i , that are compared to calculate this metric are found by applying the inverse of the exceedance curve. The inverse of the exceedance function associated with the probability of interest, P_i , is:

$$E^{-1}(P_i) = \{y : 1 - F(y) = P_i\} \tag{6.16}$$

There are two arrows on the Figure that indicate the threshold values associated with a specific probability P_i for the assessed and predicted inverse CDFs. While the figure shows only one probability P_i , in practice, K values will be used to compute the exceedance metric over a probability range of interest. The metric used to determine the goodness of fit of a predicted distribution to the assessed PFA over the K probabilities is the exceedance metric:

$$D = \frac{1}{K} \sum_{i=1}^K |E_n^{-1}(P_i) - E^{-1}(P_i)| \quad (6.17)$$

The subscript n denotes the assessed value. We use the exceedance curve to place emphasis on the shape of the distribution tails. Particularly with the false alarms, this is where we are very interested in trying to achieve a good match between prediction and assessed situations. We restrict the range of PFA to span the lowest achievable PFA (one pixel out of the total number of image pixels) up to our specified PFA value. We choose this range because this is the range used to calculate the utility in the area under the ROC curve. Note that equation 6.17 is the mean absolute error, with the mean taken over the points sampled. We typically use $K=200$ points which are equally spaced on a log scale.

When comparing the assessed and predicted utilities, the absolute error is employed and is calculated as:

$$Utility\ Error = |Utility_{assessed} - Utility_{predicted}| \quad (6.18)$$

We choose this metric because it is simple, and since each utility results from a unique combination of parameters, it avoids losing information by summarizing it over a range of parameters. We will use the utility error plotted against a range of target implant sizes as a means of comparing the accuracy of different prediction methods.

Another metric to measure the accuracy of predicted utility is the distance of a predicted and assessed utility pair from the 1:1 line (the line connecting pairs (0,0) and (1,1)) in a scatter plot of assessed and predicted utility. The points on the scatter plot are generated by considering a range of different target types and sizes, each combination mapping to an assessed and predicted utility pair on the scatter plot. This is shown in Figure 6.8 for a single assessed and predicted utility pair generated by a unique combination of image chain parameters. The 1:1 line represents the perfect prediction, in which any point on the line results from the assessed and predicted utilities being identical. The minimum distance to this perfect utility pair is taken to be the metric of utility accuracy. The summary accuracy metric for all assessed and predicted pairs of utilities is computed by taking the mean over all distances. This metric is chosen because it better expresses the distance of a prediction from the perfect prediction. Originally, we employed the linear correlation coefficient in these scatter plots, but found that it was not reporting on the accuracy of the prediction, only the degree of linearity of the pairs of utilities.

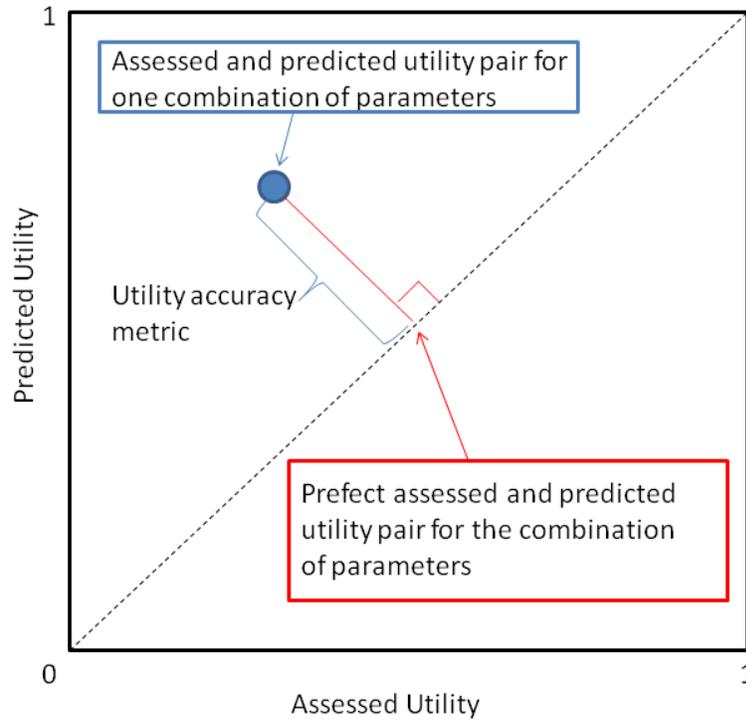


Figure 6.8: Distance metric for prediction accuracy

6.6.2 Experiments to Explore Sensitivity of Prediction Accuracy to Model Parameters

The approach to considering the prediction accuracy sensitivity to different prediction model parameters is straightforward. We create multiple predicted utilities using a range of parameter values. We then calculate the utility error using equation 6.18 for each of the utilities generated by a different parameter value. We plot the utility error against the parameter values, normalize the parameter axis by the maximum value, and then calculate the instantaneous slope at a mid-range parameter value. Figure 6.9 shows the approach. We are not attempting to characterize the sensitivity over the entire range because we want to restrict our attention to those relatively small perturbations of the parameter about the baseline of parameter settings. We have normalized the parameter axis in an attempt to allow equitable comparison between parameters that are in different units. The selection of the parameter range is somewhat subjective, and has been guided by experience with the behavior of the predicted utility. In general, we try to select a range that is representative of

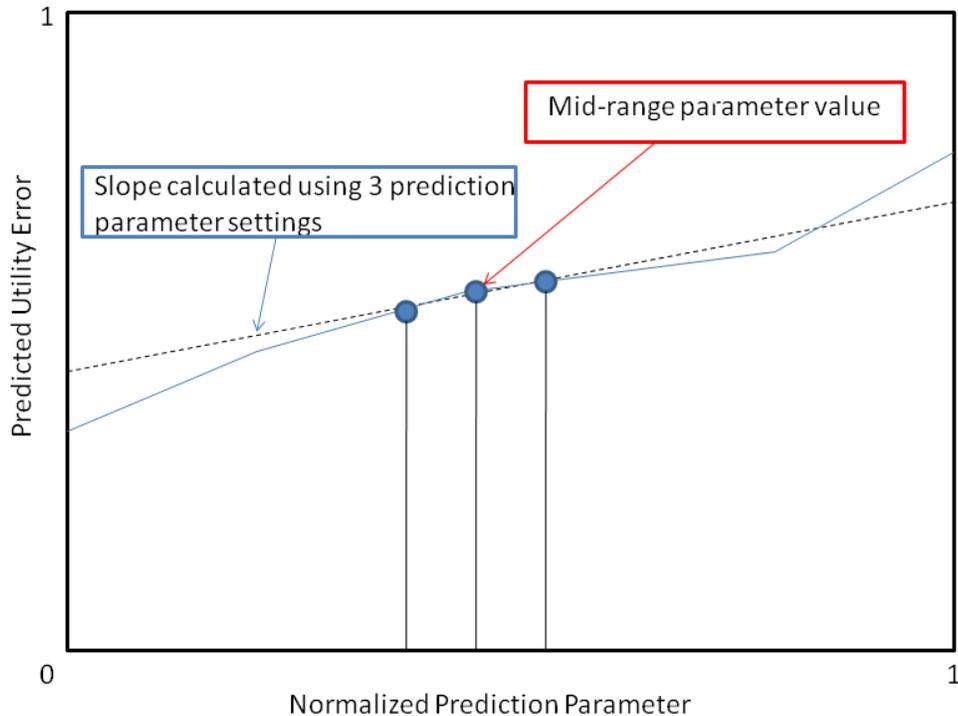


Figure 6.9: Method for calculating prediction accuracy sensitivity to prediction parameter variations

normal conditions, with the midpoint of the range reflecting a baseline situation from which we depart.

The first prediction parameter we investigate is the T-distribution degree of freedom parameter. The range of this parameter is chosen as 1 to 10. This is guided by our observations that for most targets, the degree of freedom that produces the minimal utility error is usually less than 5. The predictions with different degrees of freedom are generated by supplying the prediction model with the specified degree of freedom and recording the predicted utility. The second prediction parameter is the number of classes used in the sum of Gaussian classes model. We use four different numbers of classes (4, 8, 12, and 16) for this investigation, and obtain class statistics associated with classes resulting from a k-means unsupervised classification. We assume the baseline of 8 classes for the slope calculation. The third parameter considered is the sample support used to estimate the image mean and covariance for the normal model. We create the different statistics by sampling the image using a random uniform distribution. We use 100, 10, 1, and 0.1 % of the original image pixels as the parameter values. Even though we think of full image sample support as the baseline, in this case, 10% is the midpoint for the slope calculation. This reveals the

somewhat ad hoc nature of this approach to sensitivity, which leaves room for future improvement.

We also perform benchmarking of the time required to run predictions and assessments by capturing the total elapsed time in all (built-in and user defined) operations used in our Interactive Data Language (IDL) code. We are interested in this aspect because we want to examine the trade off between prediction accuracy and prediction speed.

6.6.3 Experiments to Compare Image-derived Utility with Other Spectral Image Quality Methods

This section seeks to place our methods of assessing and predicting utility into the context of the larger body of spectral image quality methods. Section 5.4 reviewed the spectral image quality methods from a theoretical perspective. Our focus in this section is on practical implementation issues pertaining to demonstrating the various spectral image quality techniques along with our spectral image utility measures. An important issue that we spend considerable effort addressing is to make the comparison as fair as possible. This is complicated by the fact that the different methods were established to operate using varying amounts of input information. In an attempt to better understand how we might be able to render a valid comparison, we categorize the spectral image quality prediction methods into three categories. These do not correspond to the organization of Section 5.4 because there we reviewed them based on the derivation of the approach, whereas here, we are interested in an efficient grouping for actually implementing the methods.

- Target type-specific methods. These methods require input of a specific target spectrum in order to evaluate the image utility.
- Target type and size specific methods. These methods are based on specific spectral information as well as spatial information about the size of the target relative to the spatial resolution of the sensor.
- Target type and size independent methods. These methods rely on parameters that describe the sensor and the image, but do not take a particular target into account.

Our assessment and prediction methodology depends on the specification of three major target detection parameters: the target type, the target size, and the specified PFA over which to calculate the utility. In order to achieve a fair comparison with methods that do not take these parameters into account, we need to devise a consistent strategy. It would be most

consistent to run a scenario of a single target at a single fraction and a fixed specified PFA. The full pixel fraction 1.0 seems the only fraction that takes out any arbitrariness, since the other spectral image quality methods are not built specifically on the subpixel target detection task, like our methods. The assessed utility will be treated as the “ground truth” in this comparison. The following sections will look more closely at the implementation of techniques within each category. Within each of the three categories, methods are compared by noting their respective utility scores. In cases where all of the methods produce utilities between 0 and 1, we plot the utilities generated by the predictions with the target implant assessed utility. We do not attempt to quantify prediction accuracy with distance from the 1:1 line as in the previous section, but we do find the linear correlation coefficient as a means of uncovering any interesting trends between the spectral image quality methods and our assessed utility.

6.6.3.1 Target Type Specific Methods

The spectral similarity vector (SSV) approach only requires a target and an image to return a figure of merit for image quality. The simplest implementation of this method is to calculate it using the image and target mean spectra. A more complicated implementation is to treat SSV as another detector in the target-implant methodology of utility assessment (much like SAM). We opt for the simpler method, since this is more in keeping with the spirit of its intended application. Application of this metric to any image and target pair is very straightforward.

The spectral quality rating scale (SQRS) metric is adjusted to account for specific targets as discussed in Section 5.4.1.3 using the signal-to-clutter ratio. We indicate this metric as SQRS_SCR for comparison purposes. Thus, we need to adjust the assessed utility metric to equitably compare to this metric. We use the same target type, and capture the threshold value associated with achieving a specified PFA. This threshold is empirically derived from the assessed target absent distribution of the target implant method. The SQRS_SCR metric is divided by 10 to yield a value between 0 and 1.

6.6.3.2 Target Type/Size Independent Methods

The two methods considered in this section are the SQRS and the spectral quality equation (SQE). The SQRS was implemented using equation 5.6. The SNR used is a notional quantity for the sensor type, the GRD is based on the particular image acquisition parameters (altitude) and sensor parameters (spatial resolution), and the value N is the number of spectral channels across the entire spectral range of the sensor in feature space (those channels free of atmospheric absorption effects). The metric is also divided by 10 to yield a value in the range of 0 to 1. SNRs were obtained by reviewing pertinent literature and finding values that seemed to be in the middle range of the published values.

The SQE approach in equation 5.5 is implemented with the average spectral resolution, average scene standard deviation, σ , in units of HYDICE scaled radiance units, SNR, and the GRD. The equation was derived by performing object detection on each of 120 HYDICE radiance cubes altered spatially, spectrally, and radiometrically from the baseline images. It is clear that because of the different conditions used in deriving this equation, obtaining an equitable comparison with the image-derived methods of predicting utility are challenging.

6.6.3.3 Target Type/Size Specific Methods

The general spectral utility metric (GSUM) method incorporates spatial and spectral confidence into a measure of spectral image utility. In order to realize an analogous situation using our assessed and predicted metrics, we need to note a few points. The GSUM spatial confidence term is based on N , the number of pixels of the minimum target dimension. Thus, we translate this to our paradigm by noting that if we seek a target implanted at a fraction of 1.0, this corresponds to $N=1$ pixel. We use the detection task from the Johnson Criteria. In equation 3.24, if we use a baseline of 1 pixel per cycle for N_{50} , the full pixel target will give a 50% spatial confidence, because we only have one cycle on a target. For the spectral confidence, we use PD at specified PFA from an assessed ROC curve of the target implant method for a full pixel target. We obtain this by sampling our assessed ROC curve for the specific target that we are studying to provide consistency between our utility methods and GSUM.

6.7 Experiments to investigate utility sensitivity

The challenge in exploring the sensitivity of utility to various image chain parameters is the extremely complex interaction of many image chain parameters in a seemingly simple target detection task. This is compounded by the fact that the utility metric is based on a method that relies on a hypothetical situation of a target implanted in every image pixel for the target present case. Despite these challenges, our goal is to make an initial foray into developing a methodology to better understand how our image utility metric behaves in response to changes in different image chain parameters. This is an important understanding, because we need to be able to have some confidence in our utility metric. We would like to know those situations in which it works well, those in which it is prone to give erroneous answers, and those parameters to which it is highly sensitive. Only when the metric is well understood and reliable should it be used to investigate many scenarios and draw conclusions about image utility. Thus, we assign a high priority to being able to calculate utility sensitivity. In the previous section, we briefly enumerated experiments to investigate the sensitivity of utility error to three prediction parameters. Here, we look only at the assessed utility, and we seek to examine in detail the sensitivity of the utility to as many image chain parameters as we can devise experiments to investigate. Figure 6.10 shows some of the image chain parameters that we believe will impact the utility. There are four major categories of image chain parameters: scene, image acquisition, preprocessing, and information exploitation. In our specific application, the information exploitation parameters are those relating to the target implant method of assessing image utility for subpixel target detection using the spectral matched filter. Figure 6.10 is by no means intended to be taken as a complete image chain, but rather representative of the many parameters and complexity inherent in the target detection task.

In our exploration of utility sensitivity to image chain parameters, we try to be very consistent in the approach. In practice, each image parameter requires a slightly different approach in order to vary the parameter to obtain a plot of utility versus parameter. Some require an image for each parameter setting. Others require an adjustment of a scaling factor in the utility calculation. There is much room for creative implementation, and we have devised simply one possible way to accomplish an equitable comparison. There is an infinite

space of parameters and parameter settings, and thus a very difficult problem to approach in an empirical fashion, as this research undertakes. The next section outlines our general conceptual approach to sensitivity analysis, then we briefly look at specifics of implementation for each of the major parameter groupings.

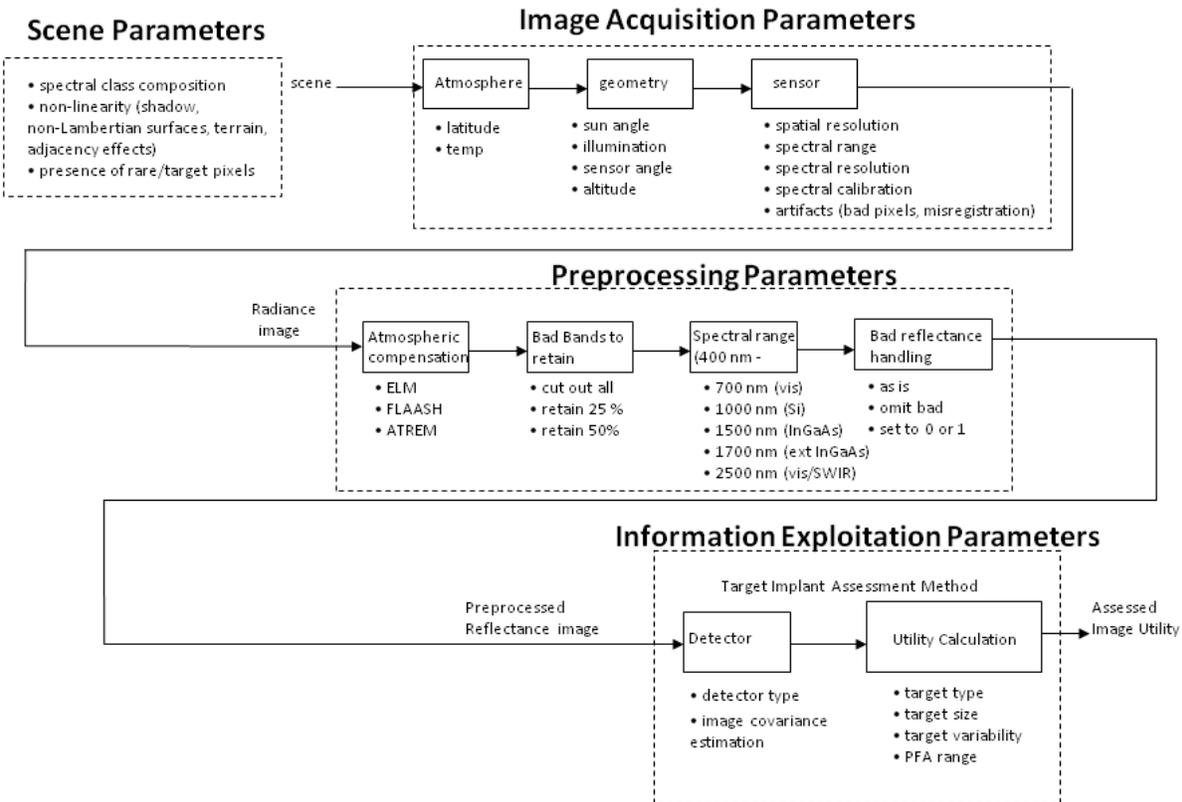


Figure 6.10: Some of the image chain parameters that impact assessed utility

6.7.1 Conceptual Approach

We would like to investigate the impact of a single parameter independently on the utility. Our first task is to devise experiments that will isolate the parameter of interest. This implies that we either manipulate the image to change the parameter so that no other parameters are changed or make the changes to the parameter directly. The tact we follow depends on which set of image chain parameters we are examining. If we are looking at the scene and image acquisition parameters, then we need to generate different scenes and images to reflect the change of the parameter. If we are looking at the preprocessing and information exploitation parameters, we can make the change directly to how we process the image and calculate the utility. This difference in approach reflects the amount of control that the image analyst has

over image chain parameters at various points in the image chain. Obviously, the analyst has no control over the scene, so he must find images that contain the scene content of interest. The control over image acquisition parameters is likewise limited, although some manipulation of these parameters is possible after the image has been acquired as long as the radiometric, spectral, and spatial characteristics are properly adjusted to reflect the change in one parameter. The alternative is to acquire several images of the same scene with different sensor characteristics, but unless the imagery collection activity is specifically designed to produce temporally and spatially co-registered images with different sensor settings, it will not be feasible. The image analyst has many options available in preprocessing parameters, and we can explore them relatively easily. In our target implant method of assessing utility, we have shown that there is much flexibility that the image analyst has in designing the specific target detection scenario.

Once we have adjusted the parameter of interest, we next need to decide on an appropriate range of parameter settings that represent excursions from some baseline condition of the parameter. Figure 6.11 depicts the notional process that we use in examining

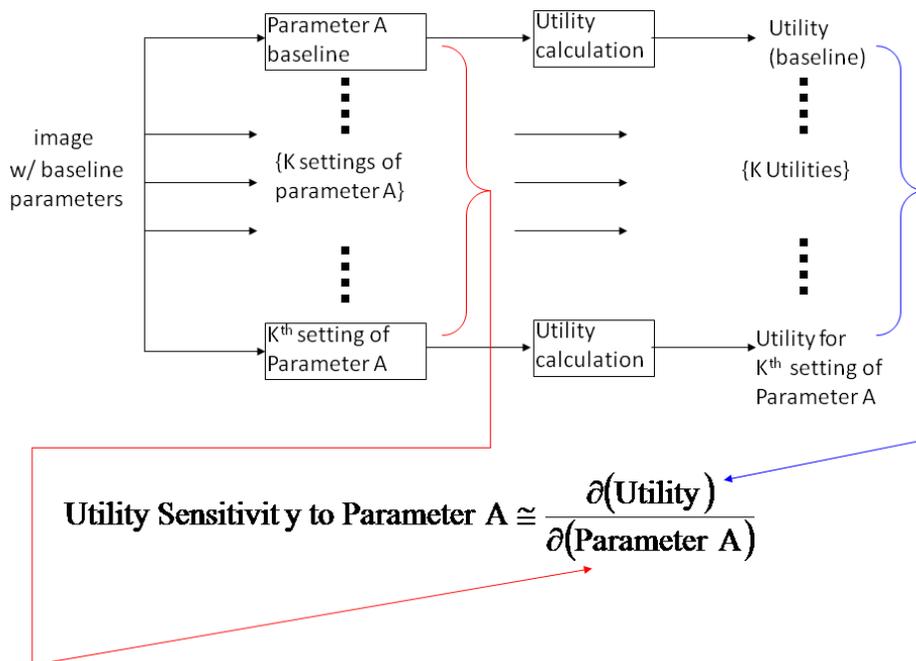


Figure 6.11: Conceptual approach to defining utility sensitivity to one parameter

the utility sensitivity to one parameter. Assuming that we have a baseline parameter defined and K parameter settings, we then calculate the K+1 utilities associated with these different parameter settings. The final step is to define the sensitivity of the utility to this parameter.

Ideally, we would like to think of sensitivity as a partial derivative that describes the change in utility with respect to the change in the parameter. In implementation, we follow the approach described in Section 6.6.2 for the utility error sensitivity to the prediction parameters. We take the simple approach of defining the utility as the slope of the plot of utility versus the parameter range in the vicinity of the baseline parameter setting. This implies that care must be taken in selecting the range of the parameters considered so that the baseline setting corresponds to mid range and that the specific parameter settings are in relatively small increments, since we will only be using the points adjacent to the baseline for this calculation. We also normalize the parameter axis by dividing by the maximum value so that we can put all parameters on a common quantifiable basis.

6.7.2 Baseline images

We choose to work with a real image and a set of simulated images in this investigation of utility sensitivity. The real image is a well-characterized hyperspectral image for which we have good ground truth and good atmospheric compensation. It was acquired under ideal conditions and sensor parameters are well documented. The simulated images are also hyperspectral images meant to simulate the sensor characteristics of the real scene, but with varying the noise characteristics and the collection altitude.

In each case, we assess the utility using the target implant method for a baseline situation that includes the ELM atmospheric compensation method, retains all retrieved reflectances even if they are negative or greater than one, and operates on channels not impacted by atmospheric transmission bands. It considers a single target implanted at a single implant fraction (target size). Utility is calculated with one detection algorithm operating at a single specified PFA.

6.7.3 Scene Parameters

We look at the sensitivity of the utility metric to targets resident in the image that the detector is seeking. Because our definition of utility is based on the target implant method, and not on the traditional approach to target detection, the target absent hypothesis will be impacted if targets we seek are present in the image. We investigate this effect by taking out all known man-made targets already present in the real image (using the HYDICE canonic data set MIT

Lincoln Laboratory (2004)) and then incrementally add them back to the image. We do this for situations in which the only target in the image is the one we are seeking and where there are multiple interfering targets present. We also examine the spectral composition of the scene by creating a hypothetical scene using spectral classes extracted from our test images and presenting individual classes and mixtures of classes to the utility calculation. We use the number of classes and the Mahalanobis distance as a quantification of the parameter of spectral composition. We do not explore the impact of non-linear effects in the scene. Pixels that are in shadow, non-Lambertian surfaces, or on sloped terrain are not considered, although their impact could be significant.

6.7.4 Image Acquisition Parameters

We use the simulated images to examine the effect of different GRD and SNR, since we have images representing different settings of those parameters. We explore spectral resolution through an averaging of contiguous spectral channels, and use the resulting spectral resolution as the parameter we plot utility against. We adjust the resulting images so that we have a constant SNR in each of our three cases of spectral resolution. The adjustment is achieved by dividing the image values by the square root of the number of contiguous bands aggregated to achieve the coarser spectral resolution.

We next examine simulated image acquisition artifacts of two types. The first is spatial misregistration in the sensor between the NIR and SWIR channels. This is accomplished by introducing a fractional pixel shift to the right in the image between the two sets of channels. The fractional pixel shift is simulated by oversampling the image pixels by a factor of ten and then introducing the spatial shift and interpolating back to the original spatial resolution. This is a crude approximation to the subtle effects of misregistration, but it provides a controllable degree of spectral mixing due to adjacent pixel spatial overlap. It provides a general sense of utility sensitivity to this sort of artifact. We next deal with spectral shift as an artifact. Although it will not impact when ELM is used, the possibility exists with other atmospheric compensation methods that the spectral shift in the image will cause problems with matching to the reference spectrum, which is unshifted. This implementation shifts the whole spectrum by some fraction of a channel bandwidth. The target introduced is likewise shifted in the target present case to make a realistic situation, but

the detection filter is based on the unshifted reference spectrum of the target. The unrealistic part is that we are operating on the 145-channel spectrum, which was derived by cutting out spectral channels impacted by atmospheric absorption. In reality, any spectral calibration issues would impact the radiance image (including the bad bands) and this would include shifting bad bands into good bands. Shifts from 0.1 to a full channel are considered. We do not consider the effects of radiometric calibration on our utility calculation.

6.7.5 Preprocessing Parameters

The atmospheric compensation methods we consider are ELM, FLAASH, and ATREM. We assess the utility for the reflectance images generated using these different methods. One drawback of this implementation is that we have no way to “distort” the target present case since we cannot bring the reference spectrum into the same FLAASH or ATREM reflectance space as we can with the linear ELM transform. Thus, we have a target present case for ATREM and FLAASH in which the target present case has a good reflectance spectrum implanted in every pixel, and the very different nature of the spectra will cause the target present situation to be (artificially) easier to detect than it would be had we matched the target and image spectra. We do investigate the effect of ELM parameter distortions on the utility, and we can distort the target present situation to match the target absent. We know that errors in the ELM come from the method we use to select image pixels from within each calibration panel, the selection of ground truth spectra, the determination of the set of calibration panels to use in the ELM, the resampling of the ground truth spectra to match the airborne sensor’s center channel wavelengths, and the spatially nonuniform atmosphere over the scene. To investigate possible ELM parameter errors, we create a situation in which ELM gain and offset are varied independently up and down by 10%. While it would be nice to show some type of spatial dependence in the gain and offset (retrieve reflectance in different parts of the image using different gains and offsets to simulate the effect of a nonhomogeneous spatial distribution of atmosphere over the scene) it is difficult to implement, and the utility metric aggregates over the entire image, independent of spatial position. We use the deviation between the retrieved spectrum and ground truth as a measure of parameter change. A source of error in ELM that we do not specifically investigate is mismatch between the reference spectra and the observed radiance values. We could look at

the effect of using reference spectra that are deviating from the mean reference reflectance as we calculate the ELM parameters, and see what effect this has on utility. This would be like altering gain and offset simultaneously.

We examine the spectral range by setting an upper limit on the channels included in our image. We plot utility against the five settings of the upper limit of the spectra. We also look at how the definition of bad bands affects the utility by including a varying number of bad bands in the image and target spectra. The percent of bad bands included is used as the parameter being controlled, and ranges from all of the bad bands to none of them in four increments. We consider how the analyst might deal with reflectances that might be negative or greater than one in the image. The three cases are to include these bad reflectance values in the image, to set them to 0 and 1, or to exclude all pixel vectors that contain these values from further processing. The utility for each case is noted, but the sensitivity cannot be easily quantified, so we only observe the effect on utility.

6.7.6 Target Detection Scenario Parameters

We examine the target variance by applying a scaling factor to the target covariance matrix. The scaling factor is varied about the baseline of 1 in five steps to investigate this effect. The target type is investigated by considering 19 different targets and then using their Mahalanobis distance from the image mean as a measure of their uniqueness. Target size is examined by plotting the utility against ten increments of the target size, translated into fractions ranging from 10% to 100% of the image pixel area. Ten settings of specified PFA are considered and the utility is plotted against these to ascertain its sensitivity.

6.8 Image Descriptions

A total of six HYDICE images are used in this research. Four of these were collected in 1995 in very well-controlled experiments at sensor altitudes of 5,000, 10,000, and 20,000 ft, resulting in GRDs of 0.8m, 1.6m, and 3.2m. Figure 6.12 shows the images from Forest Radiance I Runs 05, 07, and 09, which are 320 x 1280 pixels in spatial extent. The images have 145 and 144 (for Run 07) of the original 210 spectral channels retained for the processing. The spectral resolution varies from 3 to 17 nm, and an average of 10nm is employed for our purposes. These spectral

channels retained for processing correspond to those not impacted by atmospheric absorption. The spectral range for HYDICE is 397 – 2496 nm. The primary atmospheric compensation method employed in this research was ELM, as calibration panels were deployed in both scenes,



Figure 6.12: HYDICE Forest Radiance I Run 05, 07, and 09 images

although FLAASH and ATREM reflectance cubes were also available in the HYDICE Canonical Data Set. The excellent ground truth characterization and much-studied nature of these particular images allows utilization of existing target masks. Because the images both contained the target types from which the reference library spectra were derived, target masks were used for some of our experiments to know exact spatial location of the target pixels or to exclude all target pixels from utility processing. There are numerous targets and panels with well-characterized ground truth measurements deployed in the images, which were taken within hours of each other. The red lines in Figure 6.12 denote the area in Run 07 and Run 09 corresponding to the Run 05 image.

Figure 6.13 shows an image from Desert Radiance II run 03 with 320 x 960 spatial pixels and 144 bands retained out of the original 210. This image has a GRD of 0.8m.



Figure 6.13: HYDICE image from Desert Radiance II Run 03

Two other HYDICE images are included, although the exact collection conditions were not available to us. These are labeled as Terrain and Urban in Figure 6.14. These images are in reflectance and use 145 spectral channels for utility calculations.



Figure 6.14: HYDICE Terrain (left) and Urban (right) images

Four images from three different hyperspectral sensors are considered in order to validate the ability to assess and predict utility for imagery collected by any type of imaging spectrometer. The Modular Imaging Spectrometer Instrument (MISI) (Feng, 1995) image of the Rochester Institute of Technology (RIT) campus was atmospherically calibrated using ELM and a 332 x 1272 pixel spatial subset of the roll-corrected version is used for this experiment. The spatial resolution is approximately three meters. Only the 35 channels in the visible portion of the spectrum are used, equating to a spectral range of 408 - 738 nm with a nominal average bandwidth of 10 nm. Application of the bad bands

list results in 31 bands used for this analysis. Figure 6.15 shows a false color composite of the MISI image:



Figure 6.15: MISI image of RIT Campus

Two Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) images of Lunar Lake, NV, and Moffett Field, CA, are used (Vane, et. al. 1993). They have been atmospherically compensated to reflectance units. Of the original 224 bands, 187 were retained after applying the bad bands list. Average spectral resolution is 10 nm covering a spectral range of 370 – 2506 nm and the spatial resolution is 20 m at 20 km sensor altitude. Figure 6.16 shows these images, which are 640 x 512 spatial pixels.

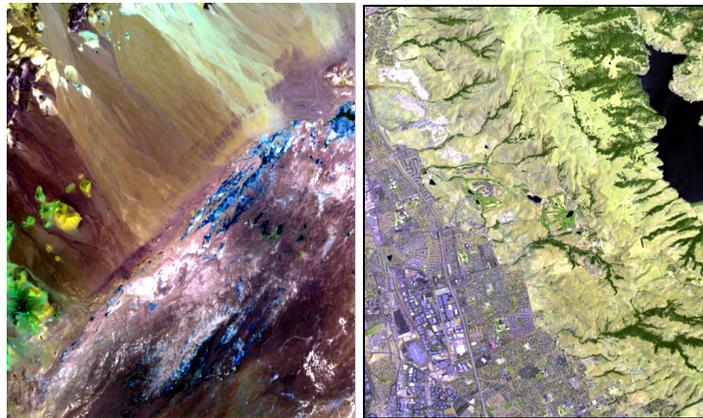


Figure 6.16: AVIRIS Lunar Lake (left) and Moffett Field (right) images

We examine another sensor, Hyperspectral Mapper (HyMap) (Cocks, et. al., 1998), which has a spectral range of 454 – 2496 nm in 15 nm channels and a spatial resolution of three meters. We use a 512 x 512 pixel subset of Cooke City, MT, in Figure 6.17. The image has 126

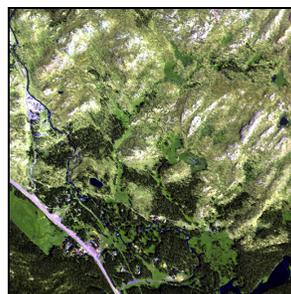


Figure 6.17: HyMap image subset of Cooke City

spectral channels, which were atmospherically corrected with the HYCORR algorithm. Of the 126 channels, we retained 96 to avoid the atmospheric absorption effects.

We use two multispectral images to investigate the application of the utility metric in images having large spectral channel bandwidths. The first is an image made with the Quickbird sensor (Digital Globe, 2008). The Quickbird image is 2000 x 2000 pixels and was taken from the larger 7099 x 7131 image of the Esparanza Fire in California on 27 October 2006. It is 4 spectral channels with a GRD of 2.4 m. The spectral response functions of this sensor along with the image are shown in figure 6.18. The spectral response function for each of the spectral channels is superimposed on a HYDICE spectrum for comparison. The image was converted to apparent reflectance using the AdVANTAGE algorithm (Beaven, et. al., 2003). This method is like an unsupervised two-point ELM that develops a linear relationship between scene data units and ground reflectance using in-scene derived vegetation and shade spectra. The average spectrum of the pixel spectra having the highest normalized difference vegetation index (NDVI) values is taken as the in-scene vegetation spectrum, and is matched with the most likely reference vegetation spectrum. The dark spectrum is determined from the minimum radiance value in each spectral band.

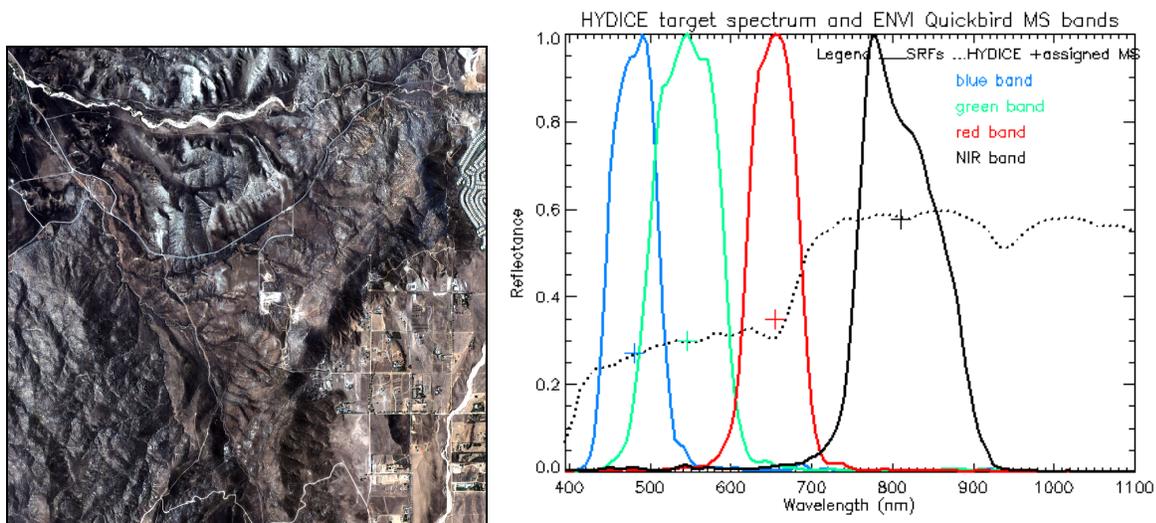


Figure 6.18 Quickbird image and spectral response functions

The second multispectral image is of Cuprite, Nevada. It was taken by the ALI (Advanced Land Imager) sensor (Bicknell, et. al., 1999) on 4 Mar 2002. There are 9 spectral channels, covering 430-2350 nm. Figure 6.19 shows the 580 x 500 spatial subset of the image and the spectral response function for this sensor. The image has a GRD of 30 m, representing

the largest that we consider. This image was transformed into reflectance by application of the Quick Atmospheric Correction (QUAC) algorithm (Bernstein, et. al., 2004). This method uses the standard deviation of scene endmembers as the gain term and the darkest observed values in each band as the offset term.

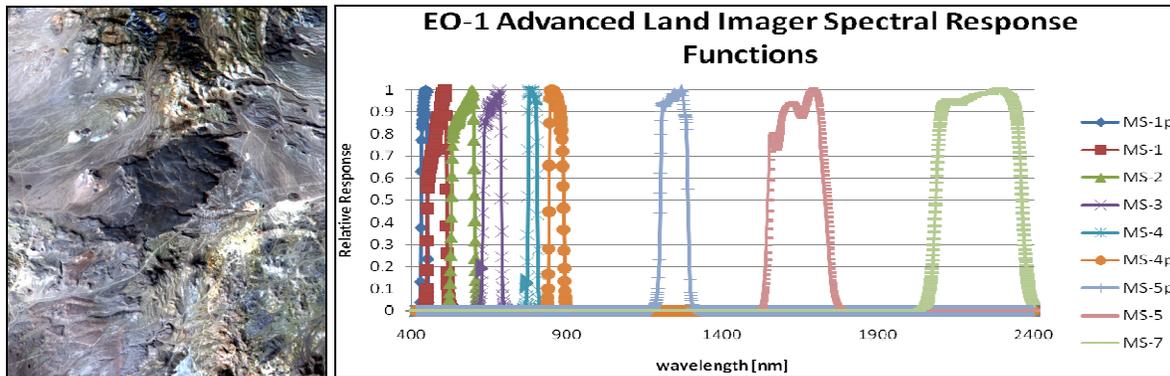


Figure 6.19: ALI Cuprite image and spectral response functions

Nine synthetic hyperspectral image generated using the Digital Imaging and Remote Sensing Image Generation (DIRSIG) model (Schott, et. al., 1999). These images are part of the DIRSIG Megascene, and were generated with differing sensor characteristics as part of an image analyst test conducted at RIT in 2006 (Kerekes, et. al., 2006). They have been formed using HYDICE sensor specifications for the spectral characteristics. The GRD of the images are 2, 4, and 8 m, and three SNR cases are considered: 10, 100, and 1000. Figure 6.20 shows these nine images with associated image parameters for reference. The 2 m GRD images are actually 400 x 400 pixels, the 4 m are 200 x 200 pixels, and the 8 m GRD are 100 x 100 pixels. The resulting images in Figure 6.20 have been made the same size, to emphasize the changes in image content due to GRD and SNR.

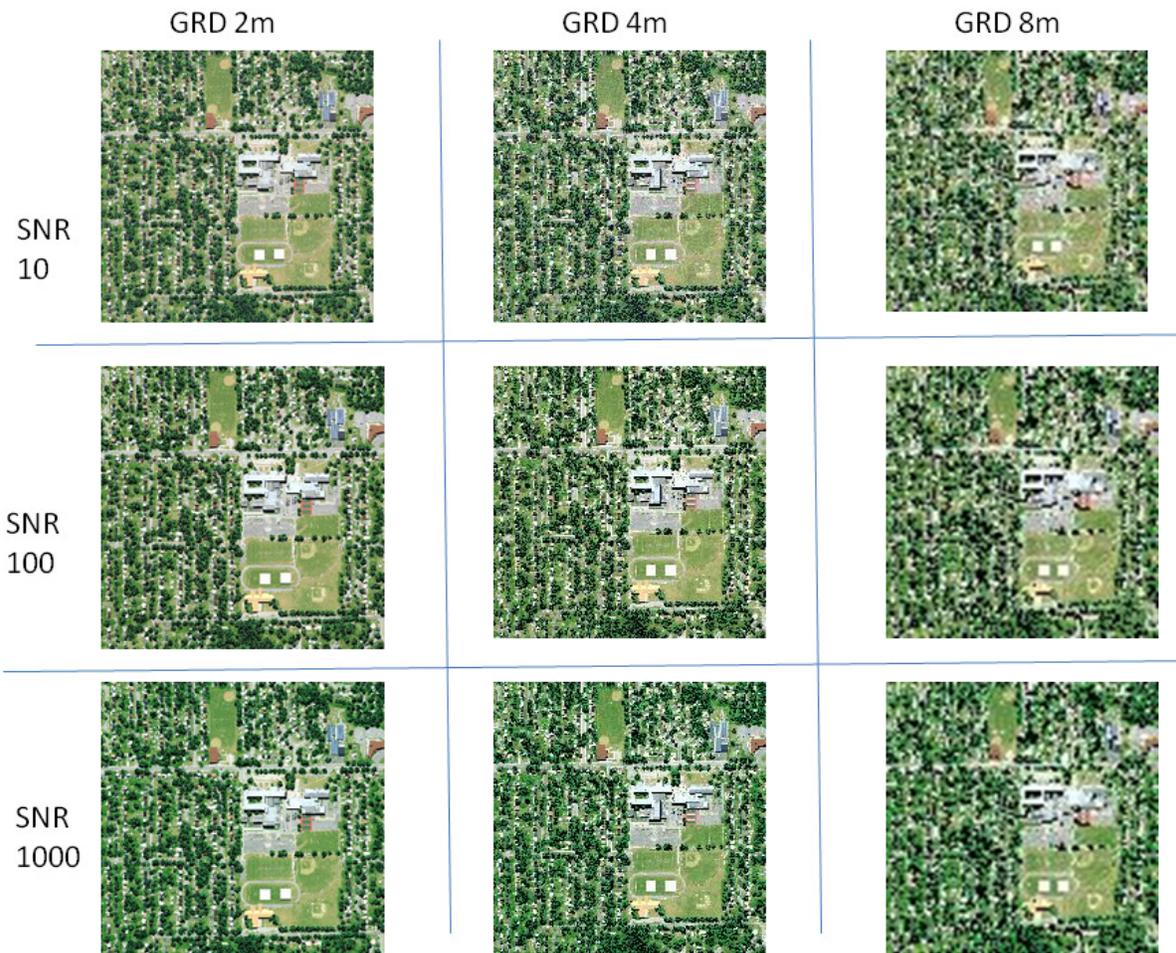


Figure 6.20: DIRISG Megascene image analyst images

Chapter 7

Results

This chapter endeavors to fully describe the character of the utility metric and demonstrate its application to a variety of spectral images. It is a collection of outcomes of empirical investigations into behavior of the utility metric in several target and image combinations. There are three major groupings of results, corresponding to each of our research objectives. The first group introduces the utility metric, shows its basic nature and behavior in different situations, and the versatility that the metric affords the image analyst in fulfilling a range of information exploitation tasks. The second group addresses the prediction of the utility metric using statistics estimated directly from the image, introduces metrics for describing the goodness and accuracy of these predictions, and then explores the sensitivity of the predictions to various factors and some tradeoffs inherent in employing the predictions. Included in this group is a comparison of our image utility assessment and prediction methods with other spectral image quality methods. The third group explores the sensitivity of the utility metric to variations in image chain parameters. In all cases, the results seek to thoroughly depict the situation and offer explanation for what is shown. In cases where further investigation is needed, this fact is highlighted.

7.1 Demonstrate Viability of Utility Metric

In this section, we show that our utility metric is an appropriate measure of the usefulness of a spectral image for subpixel target detection. In order to do this, we first illustrate how the utility metric works when applied to real spectral images. We then show how it behaves in response to changes in the specific target detection scenario, scene, and sensor involved in forming the spectral image. Having established the basic utility metric behavior, we extend the discussion to demonstrate the utility metric versatility and ability to fulfill image analyst information exploitation requirements to the desired level of detail or summary. Finally, we demonstrate the application of the utility metric to different images as an example of one way in which this utility metric might be applied by image analysts. We summarize this section with the salient characteristics that we have observed as well as the capabilities and limitations of assessing utility using the target implant method.

7.1.1 Utility Metric Formation

We must gain an understanding for how the utility metric is formed in order to appreciate how the metric behaves. The first part of forming the utility metric is the estimation of probability density functions (PDFs) at the filter output for the target absent and target present hypotheses. These estimated PDFs are illustrated using the target absent and target present distributions resulting from the filter output for the Forest Radiance I Run 05 image using the spectral matched filter designed for the c6 (radar-absorbent tent material) target. The target is assumed to be a square of 0.7 m linear dimension, translating to a 0.76 target implant fraction for this 0.8 m GRD image. Figure 7.1 shows these two distributions over all filter output values considered (lower left) and at a more restricted range of filter outputs (upper right). We see that the mean of the target absent distribution (depicted in black) is

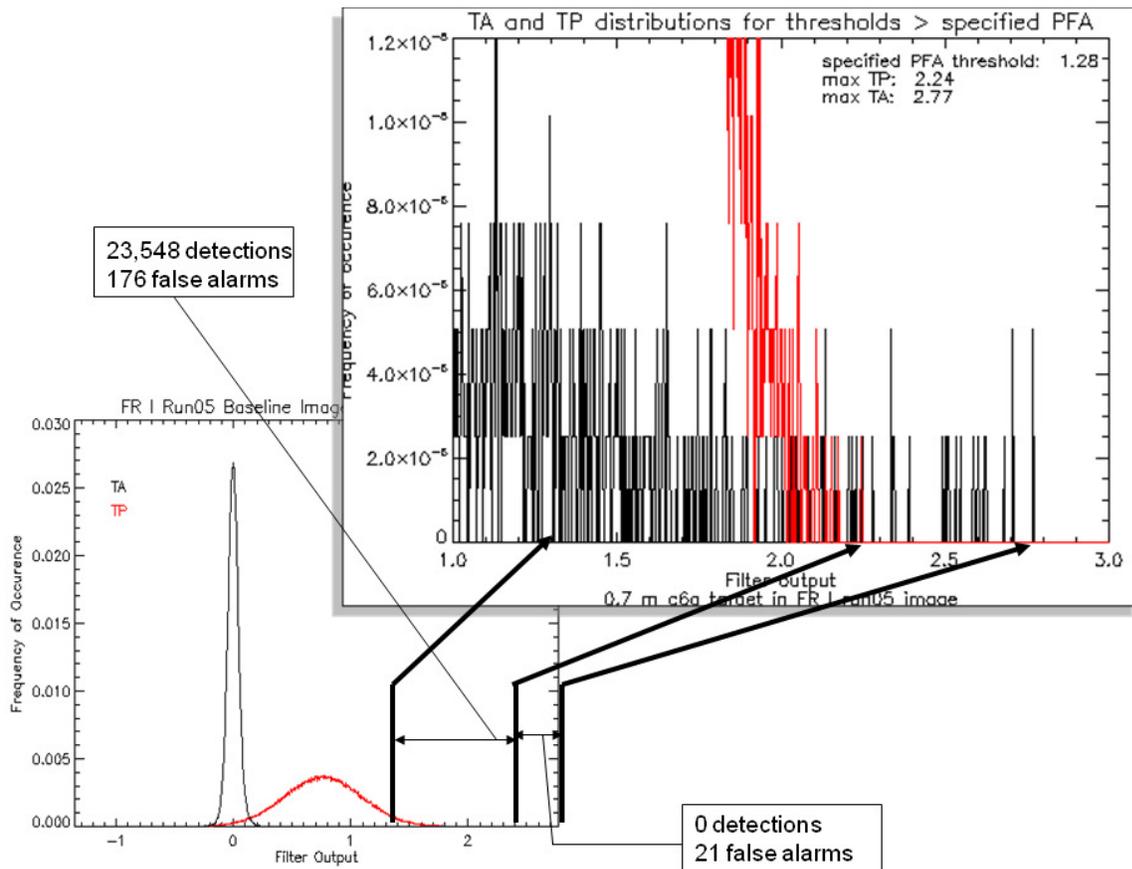


Figure 7.1: Complete and detailed views of the spectral matched filter output target absent (TA) and target present (TP) distributions for Forest Radiance I Run 05 0.7 m c6 target

zero, since the multivariate image pixel vectors input into the spectral matched filter are demeaned using the global image mean. We readily confirm that the scalar target absent distribution is also zero mean. The target present distribution (depicted in red) has a mean corresponding to the fraction of the target implant, which in this example is 0.76. The variance associated with the target present filter output is clearly much larger than that of the target absent distribution. This is due to the inherent variability of the target pixels implanted in the image to create the target present case. In this case, the multivariate covariance of the c6 target is significant, and produces a scalar output that has a correspondingly large variance. This particular target and image example is selected to depict the importance of the earliest false alarms in determining the detection performance. The original image has 172 c6 target pixels in it. The presence of those target pixels in the “target absent” image has an impact on the formation of the utility metric that is informative. Note that the largest target absent filter output (value 2.77) is larger than the largest target present filter output (value 2.24). These two points are annotated in Figure 7.1 with heavy lines. Also shown with a heavy line is the filter output threshold associated with achieving a specified PFA of 5×10^{-4} (value 1.28). A more detailed view of the right tail of the target absent and present distributions reveals that there are 21 pixels which register as false alarms corresponding to filter outputs greater than the largest value of the target present distribution. As the threshold is progressively made smaller than 2.24, more of the target present distribution pixel filter outputs are counted as correct detections until the number of false alarms (197 false alarms) corresponding to the specified PFA is reached. In the detailed view of Figure 7.1, it may be seen that a single false alarm contributes a probability of 2.5×10^{-6} to the probability of false alarm, corresponding to a single pixel out of the total number of image pixels.

The next step in the formation of the utility metric is to calculate the probabilities of false alarm and detection associated with every threshold setting and then plot these two probability vectors against each other to yield a ROC curve. Figure 7.2 shows the individual PFA and PD vectors and the resulting ROC curve as well as the normalized area under the ROC curve, which we adopt as the utility metric. The first 21 false alarms encountered before the first detection account for a PFA of 5.6×10^{-5} , as noted in the bottom plot of Figure 7.2, which shows the probabilities plotted as a function of the threshold setting. We call these the probability vectors, PFA shown in black and PD in red. At the threshold

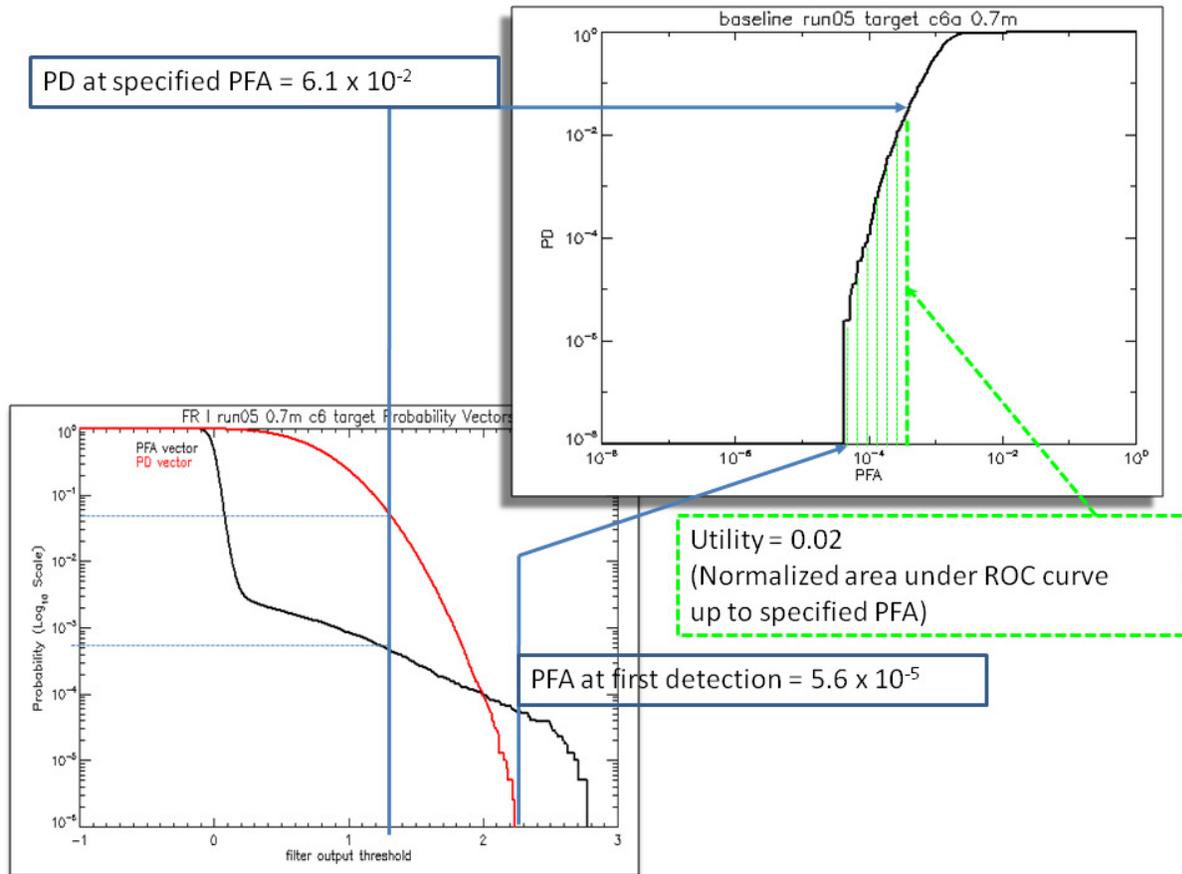


Figure 7.2: Receiver operating characteristic curve and utility metric formation for FR I Run 05 with 0.7 m c6 target

associated with the specified PFA, 1.28, we see that the PFA vector is at the specified PFA, 5×10^{-4} , and the PD vector has attained a value of 6.1×10^{-2} . The ROC curve in the upper right part of Figure 7.2 shows the PD and PFA vectors plotted against each other on a log-log scale. We see that the ROC curve has a negligible PD until the first detection. The first 21 false alarms, which correspond to thresholds larger than the largest target present filter output, have effectively “pushed” the ROC curve to the right and therefore increased the PFA at which the ROC curve begins to climb into non-negligible PD values. The utility for this particular case is shown as the dashed green lines in the upper part of Figure 7.2. This is the normalized area under the ROC curve between the lowest possible PFA (2.5×10^{-6}) and the specified PFA (5×10^{-4}). The reason for choosing this rather poor detection (and utility) example is to illustrate the ability of the utility metric to capture the “history” of the relation between false alarms and detections in a single scalar that may be used to compare different detection situations. It is clear that in this case, the utility is significantly hurt because of the

early false alarms which push the ROC curve to the right (larger PD attained at larger PFA). We know that in this particular situation, the false alarms are being caused by actual pixels corresponding to the target c6 material.

We further emphasize this important feature of the utility metric by showing a simple example in which the area under the ROC curve nicely captures the detection performance of a particular target/image/detector combination. Figure 7.3 shows the detailed target absent and present distributions for two target detection situations on the left. Target c6, seen at the

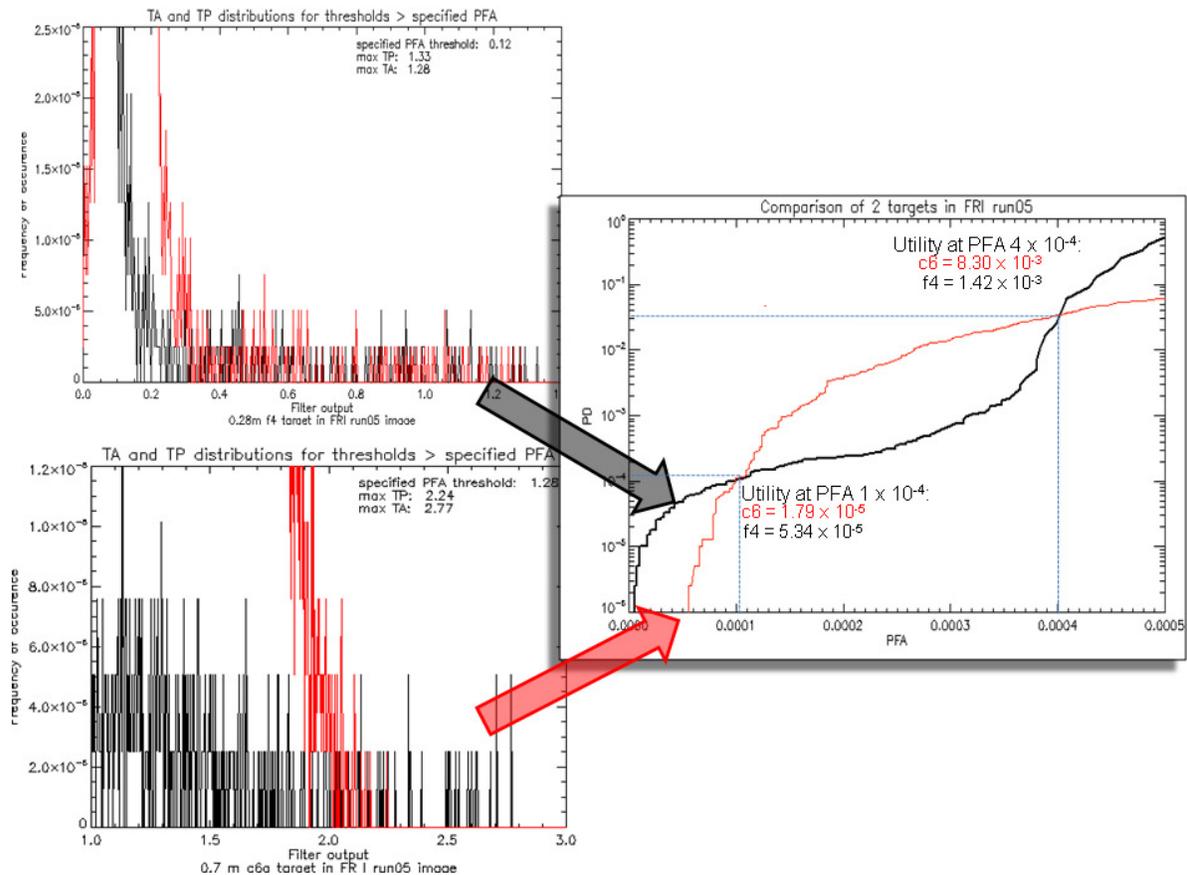


Figure 7.3: Area under ROC curves for two target detection scenarios

bottom left of Figure 7.3, has substantial (21) false alarms before the first detection is achieved, whereas target f4 has only two false alarms before the first detection (top left). The growth of the ROC curves in the right part of Figure 7.3 depicts the effect of these early false alarms. If the ROC curves were to be sampled for a PD at a given PFA of 1×10^{-4} , both would appear to produce a PD of 1×10^{-4} . However, if the area under the ROC curves is taken into account, it is clear that the f4 target ROC curve would be termed “better” since it

would produce a larger area up to the PFA of 1×10^{-4} . At a PFA of 4×10^{-4} , while the PD for both ROC curves is the same, the area associated with the c6 target is larger than that for f4, due to the shape of the ROC curve in reaching a particular PFA. Thus, while we typically think of a PD at a specified PFA as a good indicator of performance, it is clear that this metric of performance does not convey the subtleties associated with detection at very low PFAs. The normalized area under the ROC curve allows us to quantify the shape of the ROC curve as part of the utility metric, so that more convex ROC curves are rewarded for attaining a relatively high PD for a given PFA. Additionally, because we do not merely sample the ROC curve at a specific PFA, we can capture the behavior of the PD over a range of PFAs.

7.1.2 Simple Observations of Utility Behavior

We seek to better understand how utility responds to different target detection scenario parameters. To illustrate this, we look at the effects of detection algorithm, target type, and target size. In Figure 7.3, we saw the combined effect of the specified PFA, target type, and target size on the utility. A lower specified PFA (1×10^{-4} in Figure 7.3) results in a lower utility because of the correspondingly larger threshold associated with the lower specified PFA. A more difficult target (target c6 in Figure 7.3) results in a target present distribution with a greater amount of variance about the mean. A larger target results in a target present distribution with a mean further from the target absent distribution mean than a smaller target. Each of these effects will be considered separately in the subsequent sections.

7.1.2.1 Detector Type

The detector type plays a critical role in the assessed utility for a particular image. We consider three detectors, the spectral matched filter (SMF), adaptive coherence estimator (ACE), and spectral angle mapper (SAM). As we saw in Figure 7.1, the spectral matched filter responded strongly to the target c6 pixels in the image, producing target absent filter outputs larger than the largest target present filter outputs for 21 of these pixels. The ACE detector responds even more strongly to the c6 pixels in the image, assigning 70 of them values larger than the maximum target present filter output. The ACE algorithm shows more sensitivity to a covariance estimate that includes target spectra than the SMF. The result is that the utility reported by the ACE is lower than that of the SMF for this situation. At the other extreme is the SAM, which is not as sensitive to the presence of target pixels in the

image as SMF and ACE. It reports a significantly higher utility than either of the other two detectors. Figure 7.4 shows the target absent and present distributions for these three detectors in the left column and the resulting ROC curves (plotted on a linear PFA axis

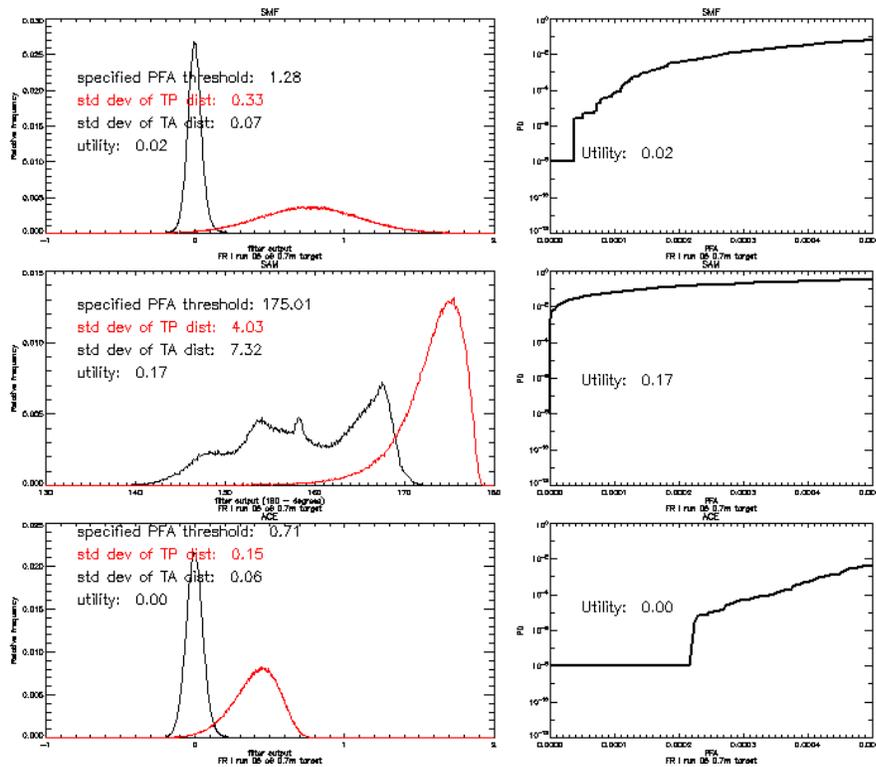


Figure 7.4: Distributions (left) and ROC curves (right) for the 0.7 m c6 target in the FRI Run 05 image for SMF (top), SAM (middle), and ACE (bottom) detectors

up to the specified PFA and a log PD axis) in the right column. The target absent distribution of the SAM detector is very interesting and its multimodal nature leads us to surmise that the angle between c6 and major groupings of pixels are very well defined. The target present distribution of the SAM detector is unimodal because the implanted target statistics, which are normal, are dominating the image statistics at the implant fraction of 0.76. The target present distributions of the ACE and SAM both exhibit a left skew, although it is unclear why. Only the SMF target present output mean corresponds to the fractional implant of the target (0.76). This is because the SMF is a linear filter, whereas the ACE is a nonlinear filter. Clearly, the first 70 false alarms in the ACE detector output cause the utility of the image assessed using ACE to be so low by pushing the curve to the right (larger PFA values). In this example, where a significant number of c6 pixels exist in the target absent image, the utility based on ACE reports a low utility because it is very sensitive to the presence of target

pixels when the image covariance was formed. SMF is not as sensitive to this, and SAM is very insensitive, since it is not based on a covariance estimate. This illustrates two important points about the utility metric. First, it will produce very different results depending on the type of detection algorithm upon which it is based. Second, because of the nature of the target implant method, if the image contains the target that is being sought by the detector, the utility will actually be reported as lower when based on detectors that are very good subpixel detectors (ACE and SMF) than on detectors that are not good for subpixel detection (SAM).

As a counter example that shows more intuitive behavior, we look at the same c6 target at a much smaller fraction (0.063) in a different image with the same three detectors. This image is a DIRSIG image which has the same spectral resolution as the HYDICE image previously considered, but has a 4 m GRD. It is important to note that the original image does not contain any c6 pixels. Figure 7.5 shows the distributions associated with this image along with the utilities in each case and the threshold corresponding to the specified PFA. We see that now ACE produces the highest utility (0.52), whereas SAM produces the lowest (0.01), and SMF is between these (0.16). These utilities confirm our intuition of the relative ranking of the performance of the target detectors because this image-target combination represents a more “ordinary” situation than the one previously presented. Note how the target present distribution for the SAM output more nearly resembles that of the target absent here than in Figure 7.4 because of the low implant fraction. Also shown to the right for each algorithm are the first 20 false alarm pixel vectors with the image and target means. This gives a sense for the types of spectra that are responsible for early false alarms which drive utility. We see that the first 20 false alarm pixel vectors (which correspond to the specified PFA for the number of image pixels) for each of the detectors are very different. The SMF false alarm pixel vectors are very variable, whereas the SAM vectors appear to be the same type of material, and the ACE vectors come from two distinct spectral groups. Thus, we see that the image-detector interaction plays a large role in the utility assigned to an image.

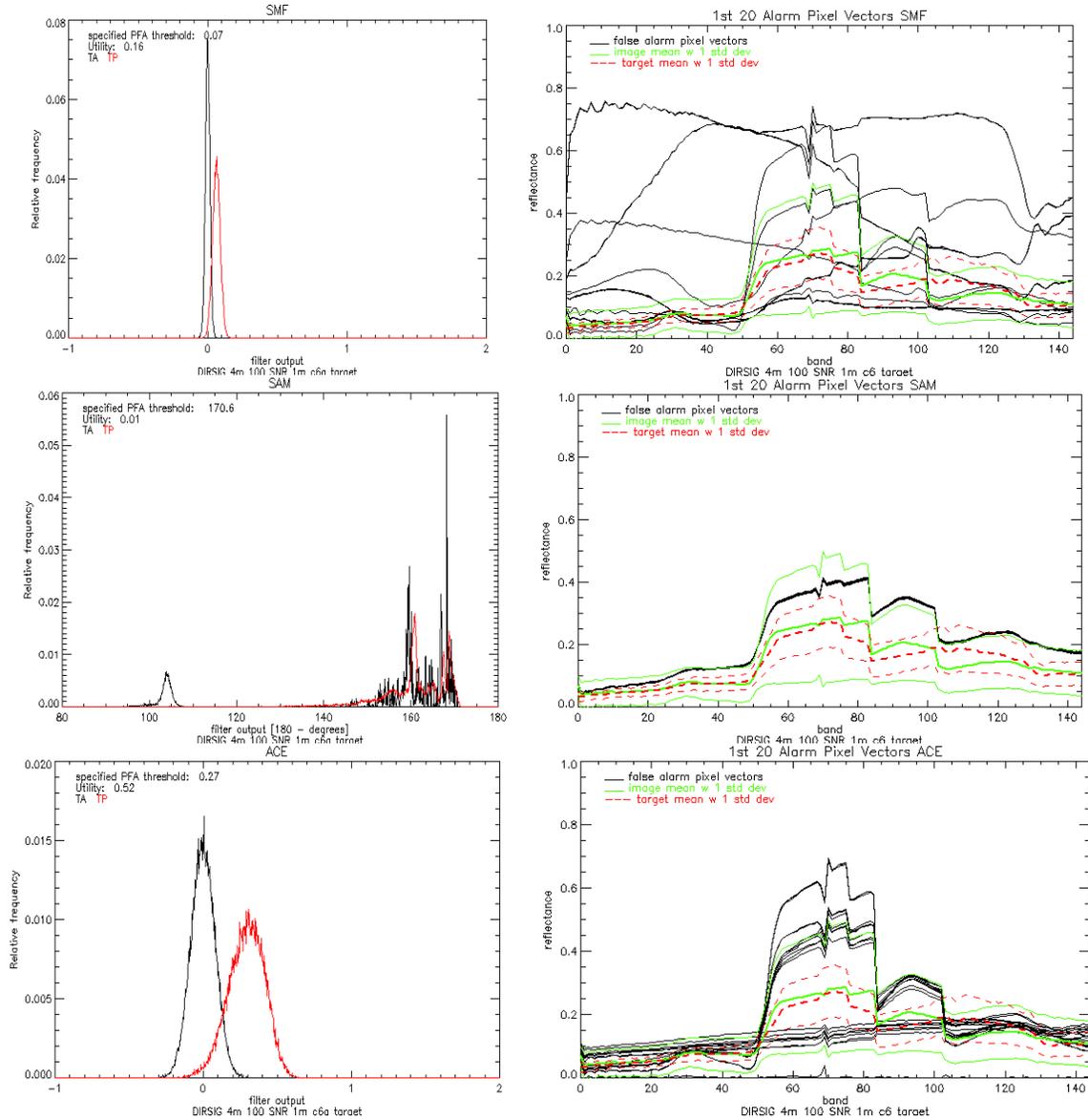


Figure 7.5: Distributions (left) and Spectra of First False Alarms (right) for 1 m c6 target in the DIRSIG 4 m 100 SNR image for SMF (top), SAM (middle), and ACE (bottom) detectors

7.1.2.2 Target Type

The type of target that the detector has been designed to detect plays an important role in the assessed utility of a particular image. This is due to the subtle interaction between the target and image in multivariate spectral space. If we were to repeat the utility assessed for FR I Run 05 with a different 0.7 m target, we would get different utility results, as shown in Figure 7.6. The left side shows the c6 target utilities, whereas the right shows the utilities for

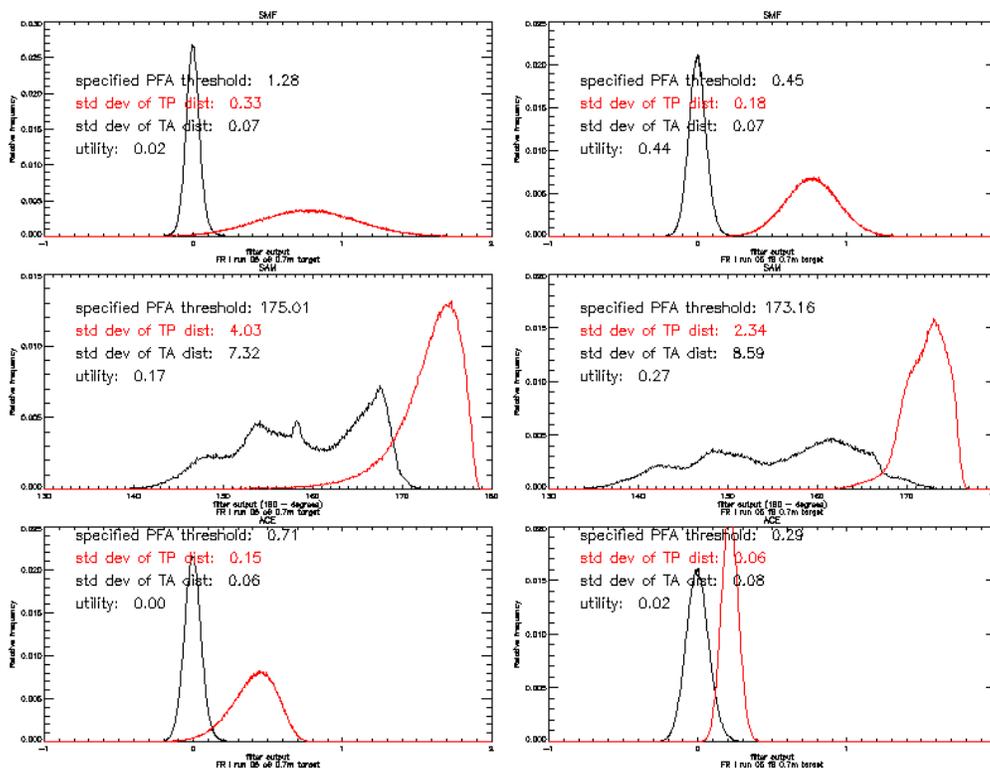


Figure 7.6: Distributions and ROC Curves for the 0.7 m c6 (left) f8 (right) targets in the FRI Run 05 Image for SMF (top), SAM (middle), and ACE (bottom) detectors

target f8, which is a tan fiberglass panel. The f8 target type manifests itself in the target present output distribution with smaller variances, thus allowing more PD to be built up as the threshold is swept from large to small values, and results in higher utilities. For the f8 target, the utility associated with SMF is the highest (0.44), with SAM next (0.27), and then ACE (0.02). The fact that ACE reports such a low utility for this target is surprising, but there are several pixels that are spectrally similar to this material in the image, which we believe has negatively impacted the ACE-based utility assessment.

In an attempt to better understand what causes different targets to be more detectable and thus report a higher utility than others, we examine several aspects of the target-image interaction. Figure 7.7 shows the image and target spectra, represented with means and standard deviations in 145 channel spectral space. The vector operation in 145-channel space does not lend itself to intuition, and it is difficult to discern that target f8 will be easier to detect than target c6. In general, we observe that the c6 spectrum looks more like the image mean over the entire Vis/NIR range of the spectrum, but this is a somewhat subjective claim.

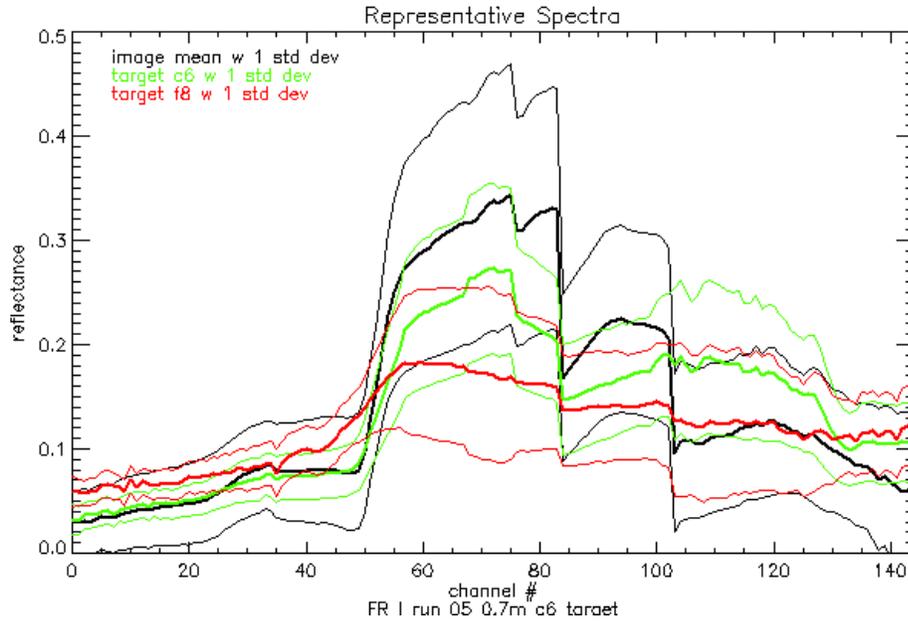


Figure 7.7: Comparison of image and target spectra

In order to better appreciate the transformation of multivariate spectral vectors from this multidimensional space into the scalar space of the filter output, we examine the details of the filter operation. The spectral matched filter is a linear filter that uses a filter vector formed from the inverse covariance matrix of the data and the demeaned target. If we examine the weights that this filter applies to each channel of input pixel vectors, we gain some insight into why certain targets behave differently than others from a utility perspective. Figure 7.8 shows the formation of the filter vectors for each of the two targets. On the left are the mean target vectors that have had the image mean subtracted from them. These vectors are then matrix multiplied with the inverse covariance matrix, which is shown in the center of Figure 7.8. The significant values are on the main diagonal, with large values corresponding to channels of small variance in the image covariance matrix. The product of the matrix multiplication is shown on the right of Figure 7.8 as the filter vector that forms the linear operator of the SMF. The results are not easily compared to the target spectral characteristics, since the weighting is determined by the “sifting” effect of the large amplitude inverse variances on the target vectors. The filter vector large amplitudes appear to correspond roughly to areas of significant amplitude change in the demeaned target vectors, channel 100 for the c6 target and channel 40 for the f8 target. In general, the c6 filter vector appears to have a larger number of significant weights compared to f8. This may be

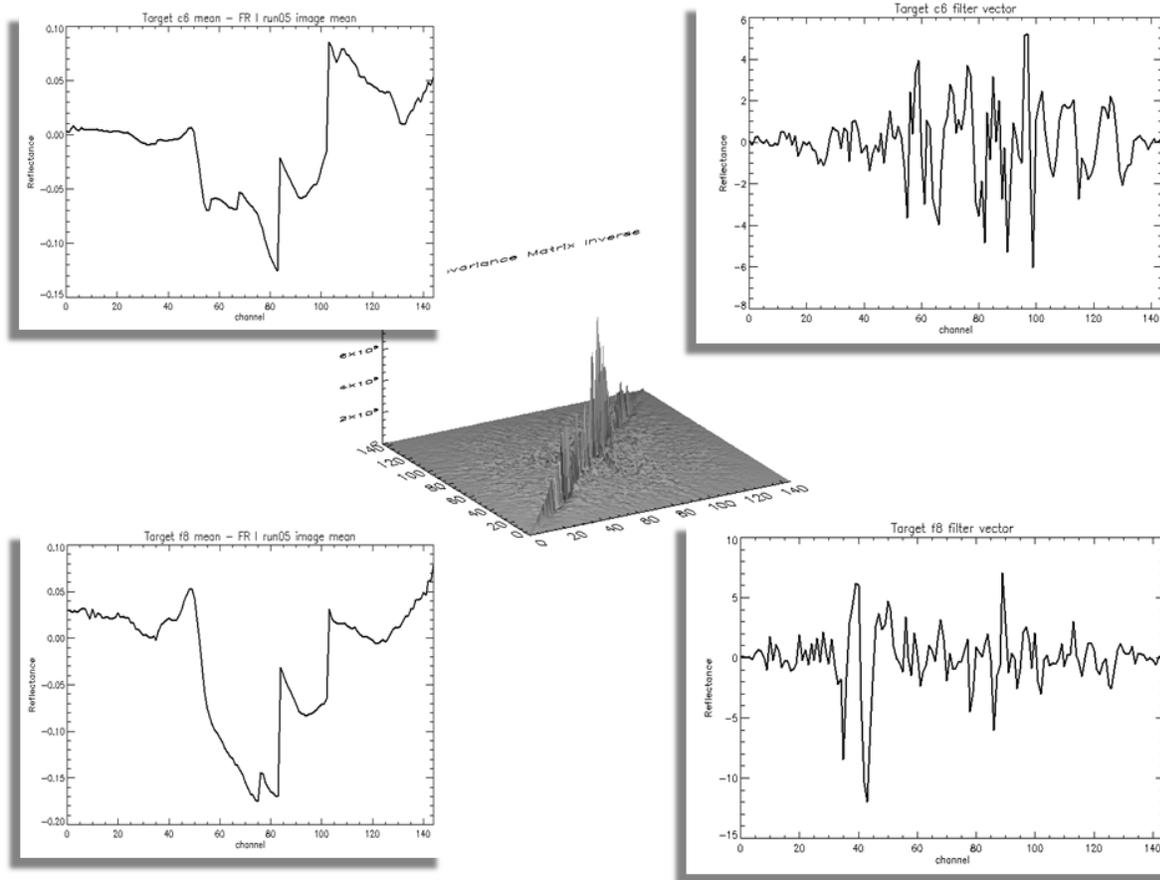


Figure 7.8: Target c6 and f8 demeaned target vectors (left), FR I Run 05 inverse image covariance matrix (center), and resulting spectral matched filter vectors (right)

responsible for producing the larger variance for c6 (0.33 versus 0.18 for f8) in the target present filter output of Figure 7.6, although the intrinsic multivariate variability of target c6 is the more likely explanation for the large variance.

We have looked for other ways that might characterize the difficulty of the target for a given image. One of these is the so called signal-to-clutter ratio (SCR), introduced in equation 5.7 and predicated on the Mahalanobis Distance between the target and image means. This is of limited value, however, because it is meant to measure statistical distances when assuming normal unimodal distributions, and spectral imagery is known to not be normally distributed. In Table 7.1, we include the SCR for five targets relative to the FR I Run 05 image to show that based on the three targets we have seen thus far, the utility metric and SCR correspond reasonably well in characterizing the difficulty of detecting a particular target in a given image. The SCR shows the c6 to be more difficult to detect than the f8 target in the FR I Run 05 image.

Target	SCR for FR I run 05
vf 124 (green camo tank)	9.2
c6 (green camo tent)	13.9
f8 (tan fiberglass)	14.7
f4 (green camo nylon)	47.5
f2 (tan canvas tarp)	69.4

Table 7.1: SCRs for Five Targets in FR I run 05Image

Throughout this dissertation, we return to these five target types. In Figure 7.9, we show images of these targets captured during the HYDICE Forest and Desert Radiance



Figure 7.9: Targets f2, f4, f8, c6, vf124

Collects. Targets f2, f4, f8, and vf124 are the standard four targets that we employ in subsequent investigations. Target c6 is used in the sensitivity analysis of Section 7.3.

7.1.2.3 Target Size

It is intuitive that larger targets should be easier to detect. Accordingly, an image assessed using a larger target will produce higher utility than one assessed for a smaller target. In order to appreciate the role of target size in the utility metric, we present a target detection scenario consisting of four targets examined at ten implant fractions. Recall that the implant fraction is simply the ratio of the area of the target to the area of the pixel. Increasing the target implant linear dimension causes the output of the spectral matched filter to assume a larger mean, with the mean corresponding to the fractional implant. Figure 7.10 shows the target absent and target present distributions for each of the four targets at the ten fractions for the FR I Run 05 image. The target numbers 0, 1, 2, and 3 correspond to target types f8, vf124, f4, and f2 respectively. Note how the increased implant fraction causes the increased separability between target absent and present distributions. Note also that the target present distribution variance increases with increasing implant fraction. Target types that consistently show smaller target present variances produce better ROC curves, assessed up to the specified PFA, as shown in Figure 7.11. As the implant fraction increases, the target present

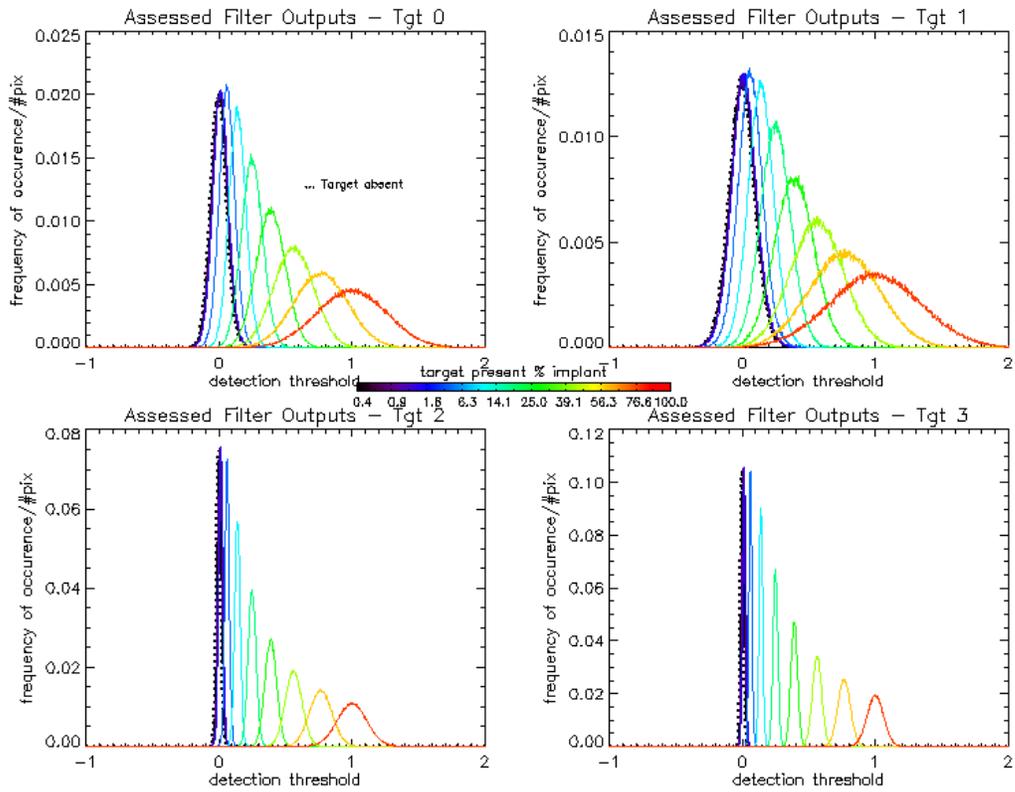


Figure 7.10: Target absent and present distributions for four targets and ten implant fractions in FR I Run 05 image

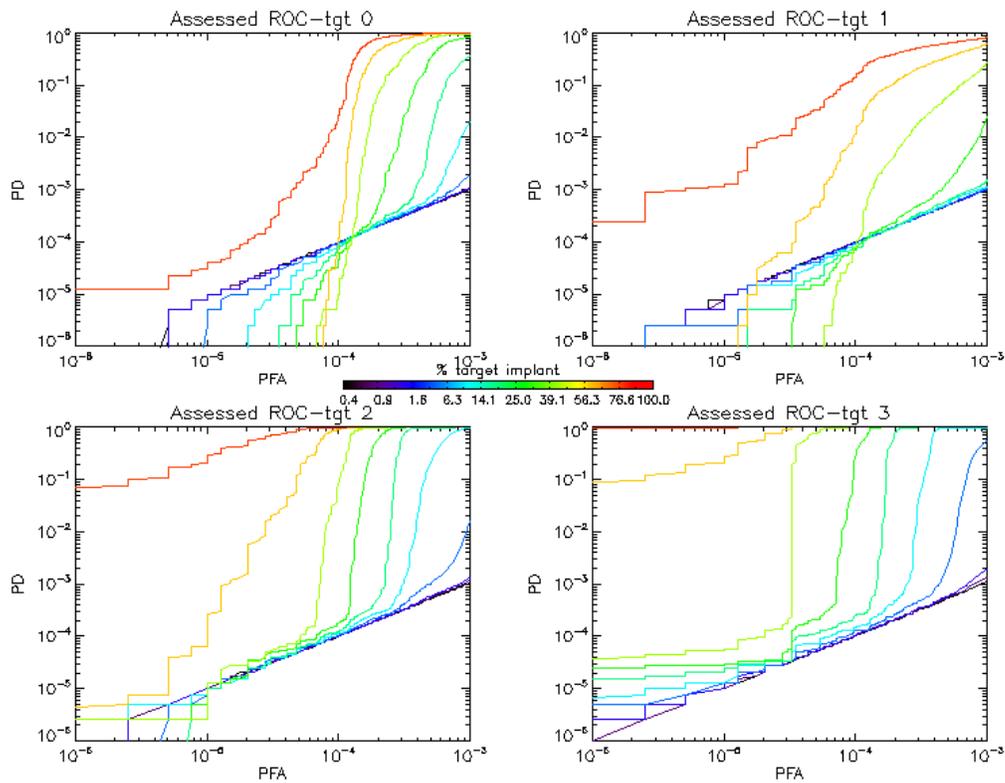


Figure 7.11: ROC curves for four targets and ten implant fractions in FR I Run 05 image

mixed pixels input to the filter begin to look more like the target, and will assume the statistical character of the random target distribution from which they were drawn. If that random target distribution has a large amount of variance associated with it, this will manifest itself in the linear filter output and create a more difficult detection situation. These ROC curves are plotted on a log-log scale and the PFA axis only extends to a little more than the specified PFA of 5×10^{-4} in order to facilitate comparison with the analogous utility metric. Note that the effect of the early false alarms impacts each target differently by pushing the start of significant PD over to larger PFAs. This effect is also seen in some cases as (targets 0-2) with increasing target fraction. Thus, increasing the target implant fraction will cause more false alarms in the target present case, but this is only evident at the very low PFAs. After an inflexion point, the increasing fraction ROC curves quickly build up the PD to higher levels. The blocky nature of the ROC curves at the low PFAs is due to individual false alarms in this region.

The normalized areas under these ROC curves may be plotted against the target implant linear dimension as a means of communicating the relationship between utility and target size. As seen in Figure 7.12, utility clearly increases with target size, but the shape of the curve describing this rise is driven by the type of target and the interaction of the target absent and target present distributions, as Figures 7.10 and 7.11 allowed us to appreciate. The

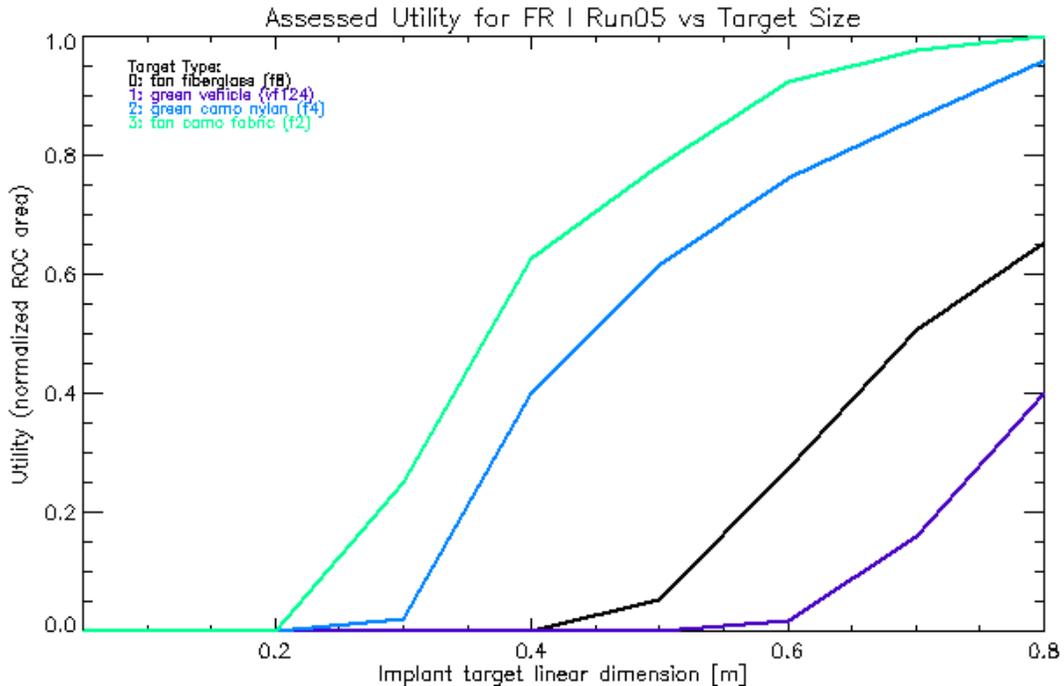


Figure 7.12: Utility plots for four targets and ten implant fractions in FR I Run 05 image

image assessed for the f2 and f8 targets shows a higher utility for smaller targets than for the vf124 and f4 targets. Note also that the relative ranking of the four targets in terms of utility roughly corresponds to the SCRs shown in Table 7.1, with f2 having the highest SCR and best utility over all target sizes.

7.1.3 Utility Trends

The previous section showed how utility responded to simple changes in the target detection parameters. In this section, we explore the more complicated nature of the relationship between the utility and the scene and image acquisition parameters. Although we will explore this in more detail in Section 7.3 as part of the sensitivity analysis, we present simple results here to highlight how utility responds to these parameters. The simplicity in this exposition stems from the fact that we choose to only look at four target types and ten target sizes as we compare two variations of the parameters of interest. This is obviously a very sparse sampling of a very large space, but it is presented here to confirm intuition with the image utility metric. We look at scene composition, sensor spatial resolution, spectral resolution, and signal-to-noise ratio. These last three parameters are the principle parameters in determining the design of a spectral imaging system. We wish to have a solid understanding of their effect on utility.

7.1.3.1 Scene Composition

Clearly, the content of the scene will play a large role in the amount of utility assigned to an image for a specific task. For the subpixel target detection task, where each pixel is treated independently, the spatial interrelationship between pixels is not a factor that determines the utility of the image. The main factor in this case is the spectral complexity of the image. We have no way of consistently characterizing this parameter, although there are measures of spectral complexity such as the intrinsic dimensionality. We believe that the complexity of the scene for the subpixel detection task is not absolute, but relative to the target that one seeks. Even here, the relation between target and image defies easy quantification, as the SCR shows from the earlier discussion in Section 7.1.2. But as long as we approach the problem of characterizing the scene complexity in a relative manner using the SCR, we can somewhat equitably (for the small sample set considered) compare images. We choose to

compare two images made by the same sensor of two different scenes. These are the HYDICE Forest Radiance I Run 05 (FR I) and Desert Radiance II Run 03 (DR II) scenes that have equivalent GRD, spectral resolution, and sensor noise characteristics. We examine the utility of each image in searching for a target of the same size. Thus, the factors that drive the assessed utility should be the scene composition and the accuracy of the atmospheric compensation. Figure 7.13 shows the FR I and DR II image utility versus target implant size plots where utility has been calculated using a specified PFA of 5×10^{-4} . We see that the DR II image clearly produces a higher utility for the f8 and vf124 targets than the FR I image, and f2 and f4 produce higher utility in DR until about 0.5m target size. We do not have a plausible explanation for this behavior, other than the desert scene composition provides the right spectral contrast to allow the spectral matched filter to better detect the targets than the grass/forest background. As a rough validation of this idea, consider again the SCRs associated with the two scenes for each of the four targets and see how the DR II image produces uniformly higher SCRs than the FR I image in Table 7.2. We also include the utility

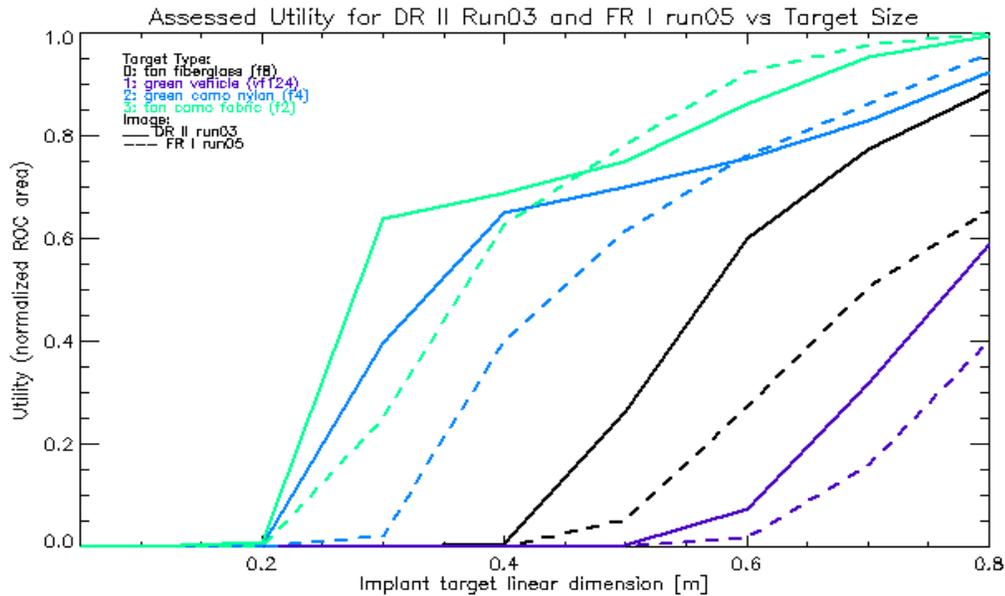


Figure 7.13: Utility versus target size curves for four targets and ten target sizes in FR I and DR II images

	Target f8	Target vf124	Target f4	Target f2
FR I SCR	14.7	9.2	47.5	69.4
FR I Utility	0.15	0.06	0.36	0.46
DR II SCR	22.4	11.7	54.2	73.7
DR II Utility	0.25	0.10	0.43	0.49

Table 7.2: SCRs for Four Targets in FR I and DR II Images

averaged over all target sizes for each target type to quantify the different utility of the two images based on the scene content. We conclude that the spectral characteristics of the DR II image make it more useful for detecting these four supixel targets than the FR I image.

7.1.3.2 GRD

For the same size target, a larger GRD will produce a situation in which the subpixel fraction is smaller. Thus, a larger GRD equates to more difficult detection and lower utility. We compare two different GRD images of the DIRSIG scene, because these simulated scenes have been built so that the SNR and sensor characteristics are exactly the same between the two images of the same scene. Figure 7.14 shows the resulting utility versus target size plots for the four targets. The two GRDs are 2 m and 4 m and the SNR is kept constant at 100 in both cases. We see that the 4 m GRD image has a lower utility at every target size when

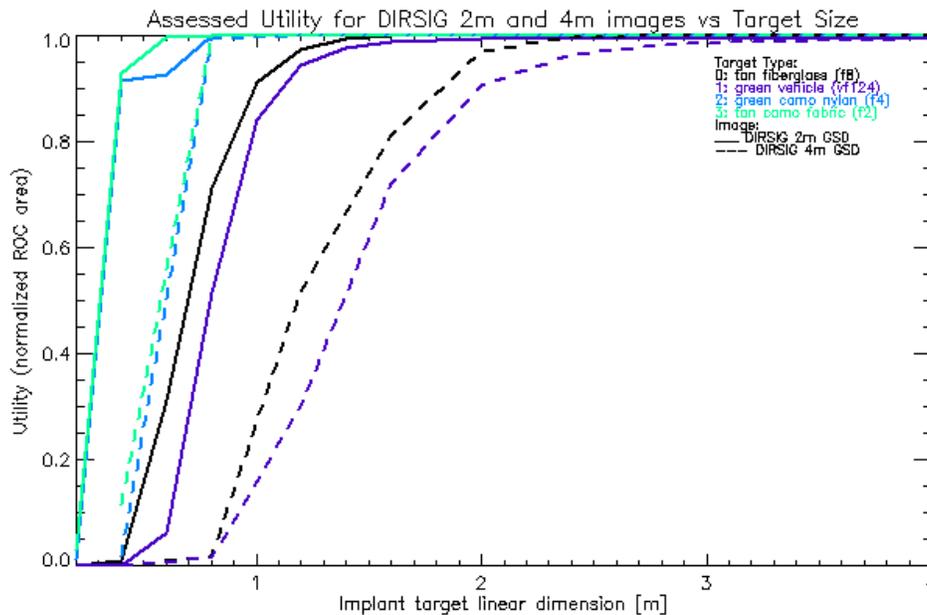


Figure 7.14: Utility versus target size curves for four targets and ten target sizes in 2 m and 4 m DIRSIG Megascene images

compared to the 2 m GRD image. This is because for the same size target, the fraction of implant is four times smaller for the 2 m GRD image.

Using target c6 at a constant 0.7 m linear dimension, we take a closer look at the effect of GRD using the same 2 m and 4 m DIRSIG images. The utility for the 2m image is 0.59, while that of the 4 m image is 0.01. By looking at the target absent and present distributions, we see that the larger GRD is a more difficult detection situation. Figure 7.15

shows that the target present mean is four times larger for the 2 m image, resulting in greater separability between target absent and present distributions. The right side of Figure 7.15 shows the pixel vectors corresponding to the earliest false alarms. There are 20 false alarms corresponding to the specified PFA for the 4 m image and 80 for the 2 m image (50 are shown) due to the different number of image pixels in each image. It appears that the same types of pixel vectors are causing false alarms in the two images. Thus, the difference in utilities is due to the greater distribution separation in the case of the 2 m GRD image.

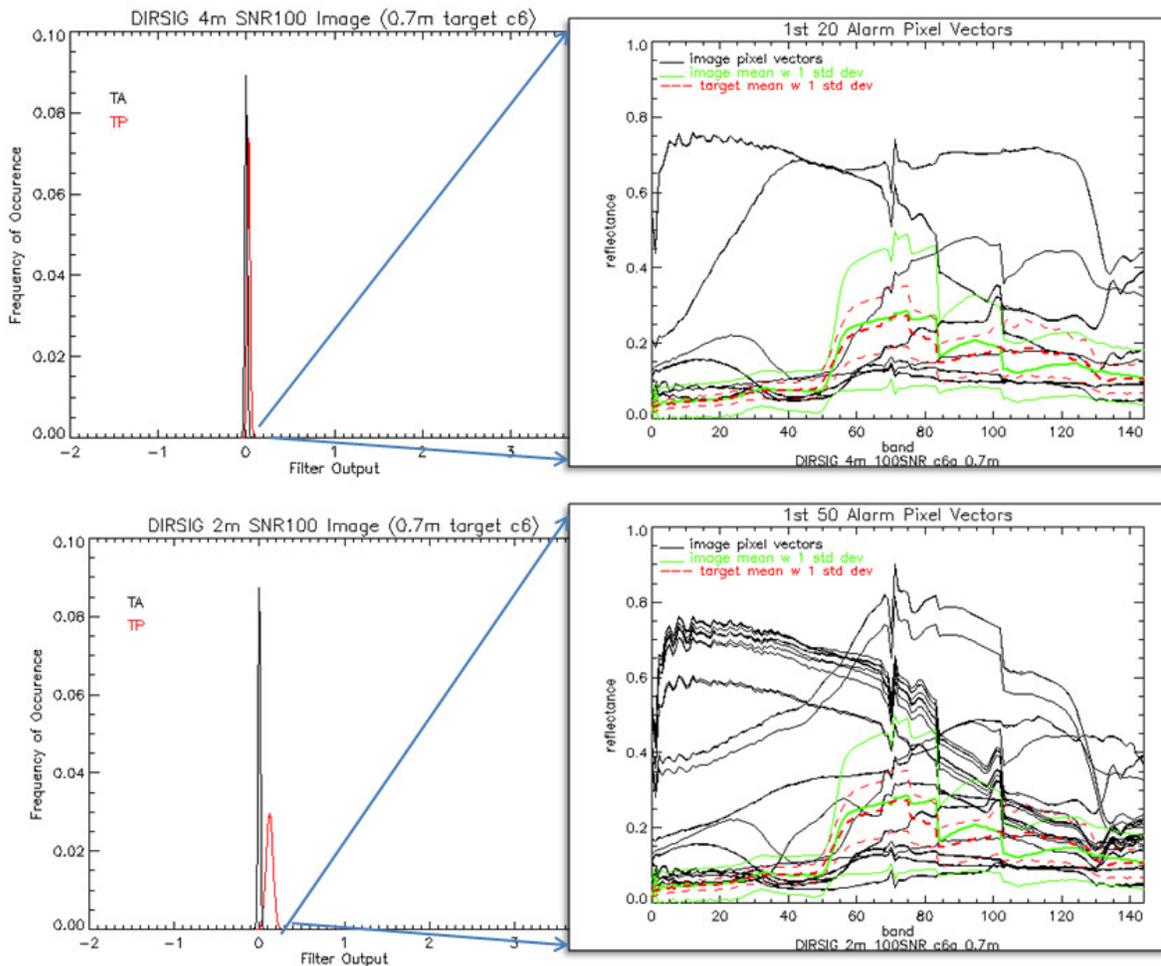


Figure 7.15: Target absent and target present distributions for DIRSIG 2 m and 4 m images with first false alarm spectra

7.1.3.3 Spectral Resolution

The ability to use the additional information captured by finer spectral sampling is an underlying motivation for using spectral imagery. Thus, we would expect utility to increase with finer spectral resolution. In order to validate this expectation, we look at two images of

the same scene, but at different spectral resolutions. The baseline image is the DIRSIG Meagascene Tile 1 image with 4 m GRD, 100 SNR, and 10 nm (145 channel) average spectral resolution. A wider spectral resolution is achieved by averaging contiguous bands in the original image and compensating to maintain a constant SNR with the wider spectral channels. The channel aggregation in this example uses four contiguous bands, resulting in an image with 37 bands (40 nm average spectral resolution). Thus, we divide by the square root of the number of channels aggregated (four) to give us a multiplicative factor of 0.5 for the SNR. Figure 7.16 shows the effect of the spectral resolution increasing from the baseline

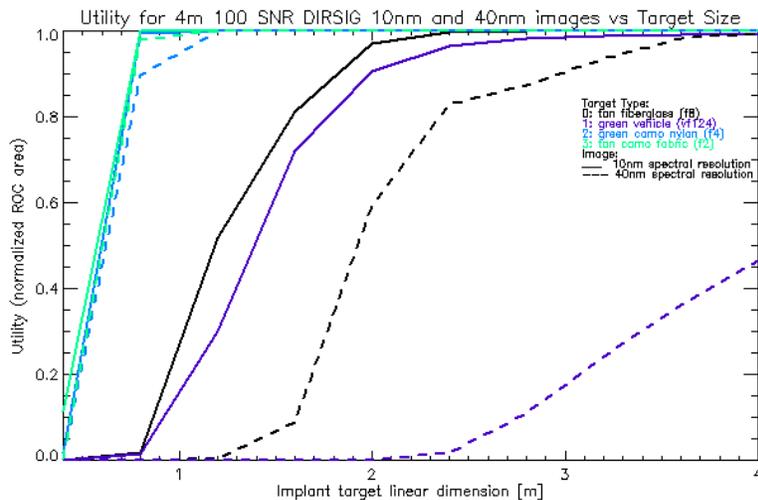


Figure 7.16: Utility versus target size for four targets in DIRSIG 4 m 100 SNR 37 channel and 145 channel images

10 nm to 40 nm for the four standard targets against the range of target sizes. We see that the utility is adversely impacted by the broader channels. The impact is more pronounced with the more difficult targets, in particular with the vf124 target.

In an effort to further understand the underlying mechanism that explains the relation of utility to spectral resolution, we explore a single target type and size in more detail. We use a 1 m target of c6 material in assessing the utility of the DIRSIG 4 m GRD 100 SNR image. At the top of Figure 7.17, we plot the mean image spectra of the two different spectral resolutions. Because of the noise compensation for the spectral channel aggregation (described in Section 6.7.4), we see that the resulting spectra are a factor of two smaller than the original. The loss of spectral character due to the channel aggregation of four adjacent channels is seen in both the image mean and the covariance matrix of the 40 nm spectral resolution image. The covariance matrices are plotted in the bottom part of Figure 7.17, and

Chapter 7. Results

whereas the means were plotted as a function of wavelength [nm], the covariance matrices are shown as a function of channels and do not include those channels affected by atmospheric absorption. The 40 nm spectral resolution (37 channel covariance) is smoother since we have averaged bands, but it is also four times smaller in magnitude than the 145 channel covariance. The result in the inverse covariance matrix is a much larger inverse for the 37 channel which forms a significantly larger weight filter vector. These larger weights,

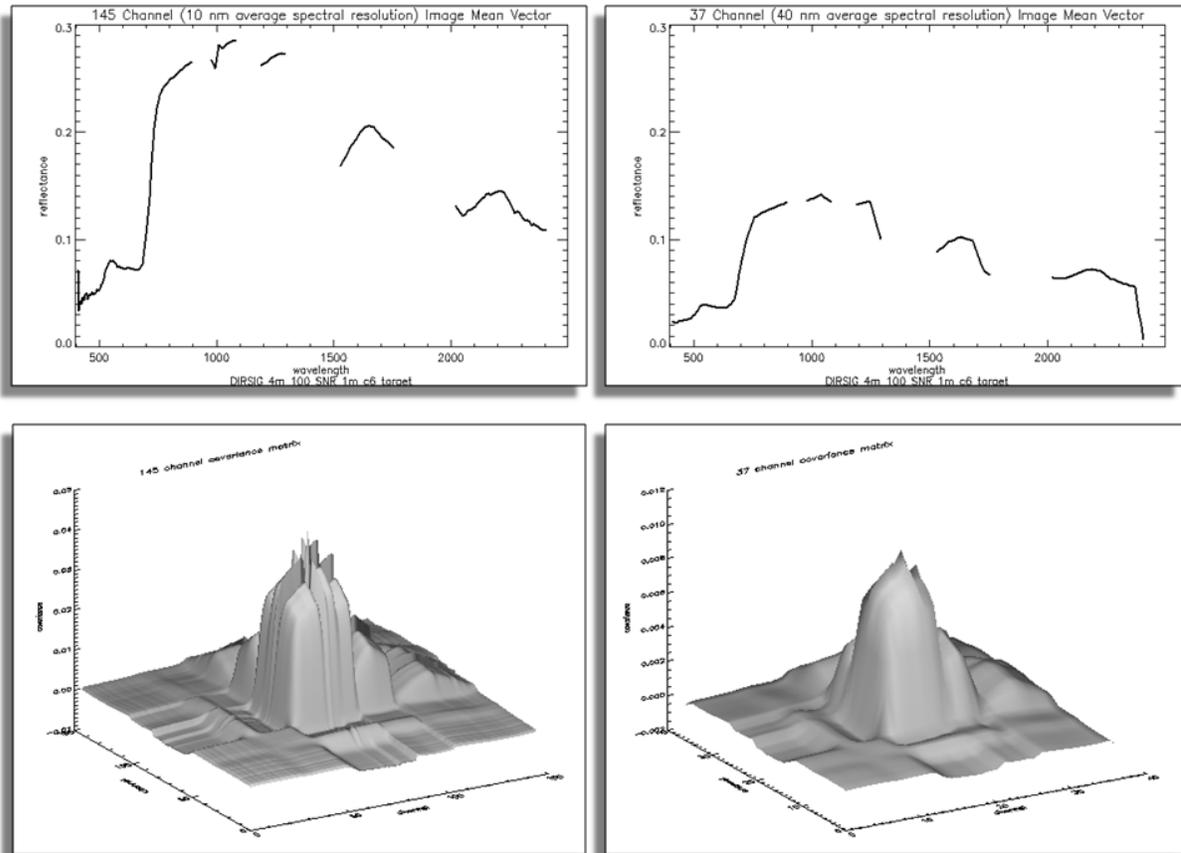


Figure 7.17: Mean spectra (top) and covariance matrices (bottom) of DIRSIG 4 m 100 SNR 145 channel (left) and 37 channel (right) images

when coupled with the large variability of the c6 target, lead to a wider target absent and present distributions, as Figure 7.18 illustrates. Figure 7.18 also shows the pixel vectors associated with the first twenty (corresponding to the number of pixels required to reach the specified PFA threshold in this image with 400,000 pixels). The spectrally “smoother” and “darker” (lower reflectance) character of the 40 nm pixels are clear in the right part of Figure 7.18. These larger weights appear to have the effect of creating wider TA and TP distributions that do not have as much separability as in the 145 channel case. While the false

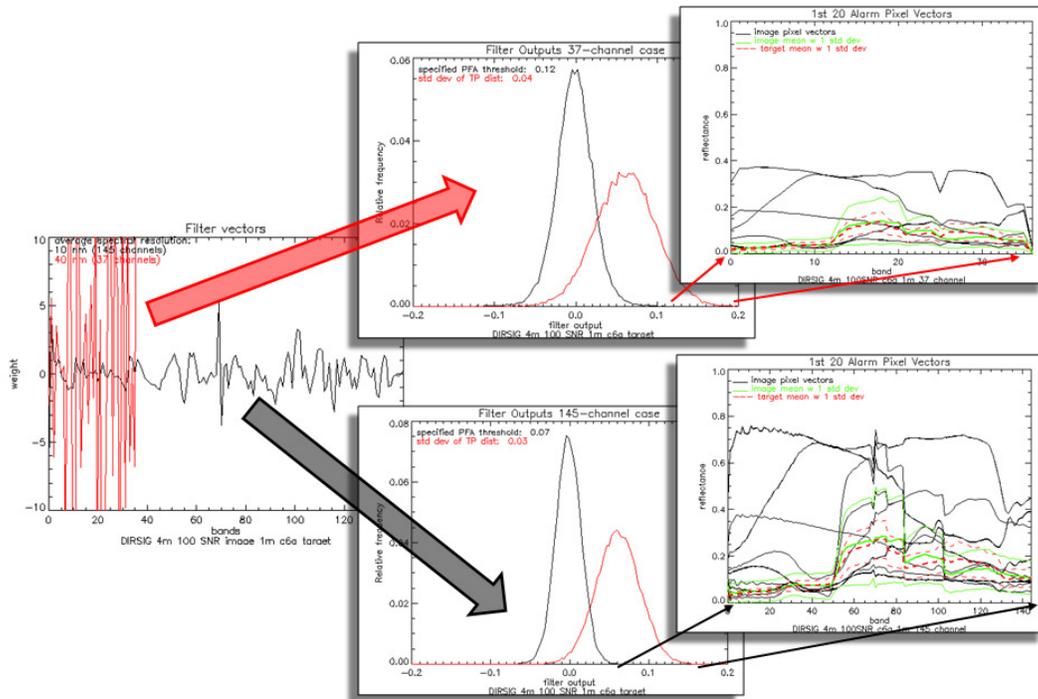


Figure 7.18: Filter vector, resulting filter output distributions, and spectra of first 20 false alarms for 1 m c6 target in DIRSIG 4 m 100 SNR 37 channel and 145 channel images

alarms are due to the same type of image spectra, the decreased utility comes about because of the wider distributions, which makes the threshold for the specified PFA larger than in the 10 nm case.

Figure 7.19 shows the spatial distribution of the first 20 false alarms for each of the

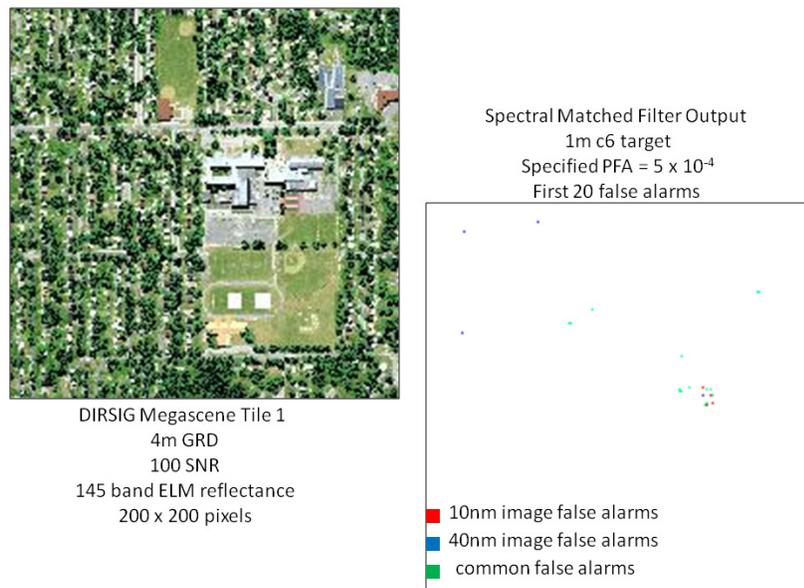


Figure 7.19: Spatial distribution of first twenty false alarms for 1 m c6 target in DIRSIG 4 m 100 SNR 37 channel and 145 channel images

spectral resolutions (red for 10 nm and blue for 40 nm). The false alarms common to both are indicated as green, and there are 15 of them, so that the similar appearance of the spectra in the plots of Figure 7.18 is confirmed. We have demonstrated with this simple example that aggregating spectral channels creates a spectral smoothing that takes away the spectral characteristic features and creates a more difficult detection situation and therefore lower utility.

7.1.3.4 Signal to Noise Ratio

The ability of DIRSIG to create images having different image acquisition characteristics is utilized to demonstrate the sensor noise characteristic impact on utility. The 4 m GRD image is used in two variants: 10 and 100 SNR cases. Intuitively, we expect that increased SNR will improve the spectral matched filter detector performance, since the target spectra being sought are more readily matched with the reference spectrum. In Figure 7.20, we see the effect of the decrease in SNR for the 4 m GRD DIRSIG image from the baseline of 100 to 10 for the four targets at the ten fractions. The effect of the lower SNR on utility is clearly deleterious, pushing the minimum useful target size out by 1.5 – 2 m for all four targets. In order to better understand this behavior, we select one target type and size and trace back to

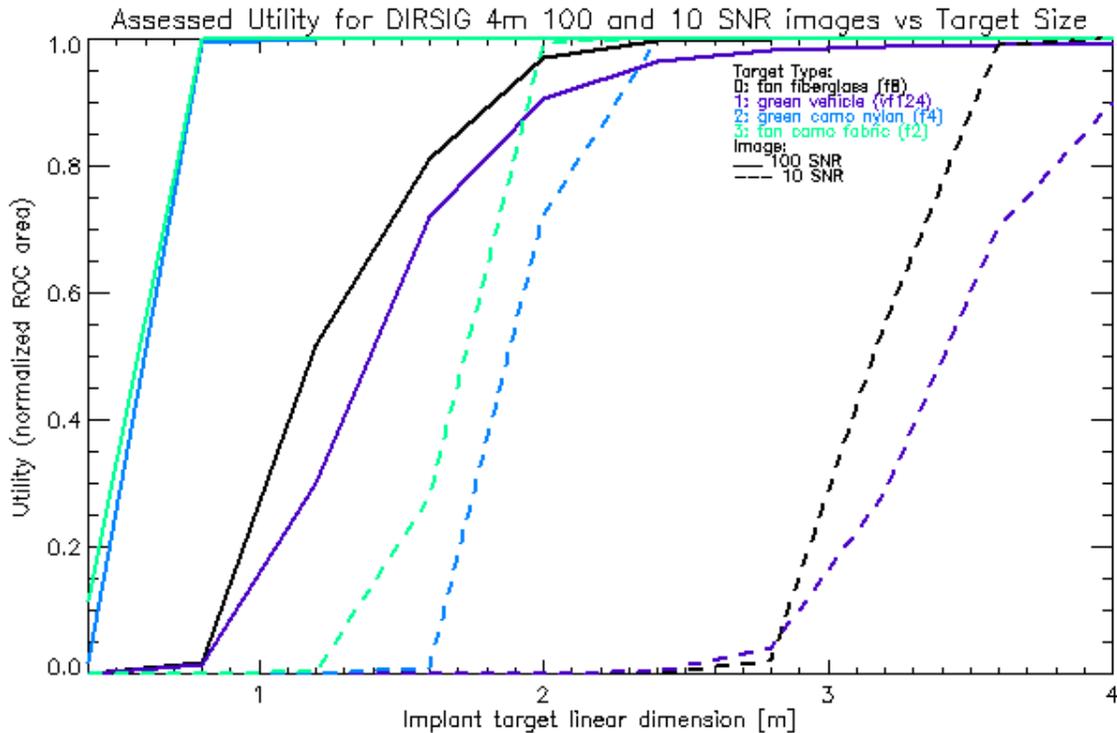


Figure 7.20: Utility versus target size for four targets in DIRSIG 4 m 100 and 10 SNR images

the filter-level characteristics to see how SNR impacts the utility. First, we look at the image mean vectors compared to each other and the target vector. Figure 7.21 shows the image mean vectors for the two SNR cases compared with the reference target mean vector. We see that the lower SNR manifests itself predominantly in the lower channels of the smaller SNR, but in general, the means look similar. The right part of Figure 7.21 shows the main diagonal of the inverse covariance matrix each of the two SNR images. The larger values associated with the higher SNR image are indicative of the smaller variances in the image covariance matrix. The smaller values in the inverse covariance of the 10 SNR image are due to higher variances in the image covariance. The mean of the 100 SNR inverse variances is roughly two orders of magnitude larger than the 10 SNR mean inverse variances. This reflects the order of magnitude degradation in SNR. The result of the application of the respective inverse covariance matrices to form the filter vector is shown at the bottom of Figure 7.21.

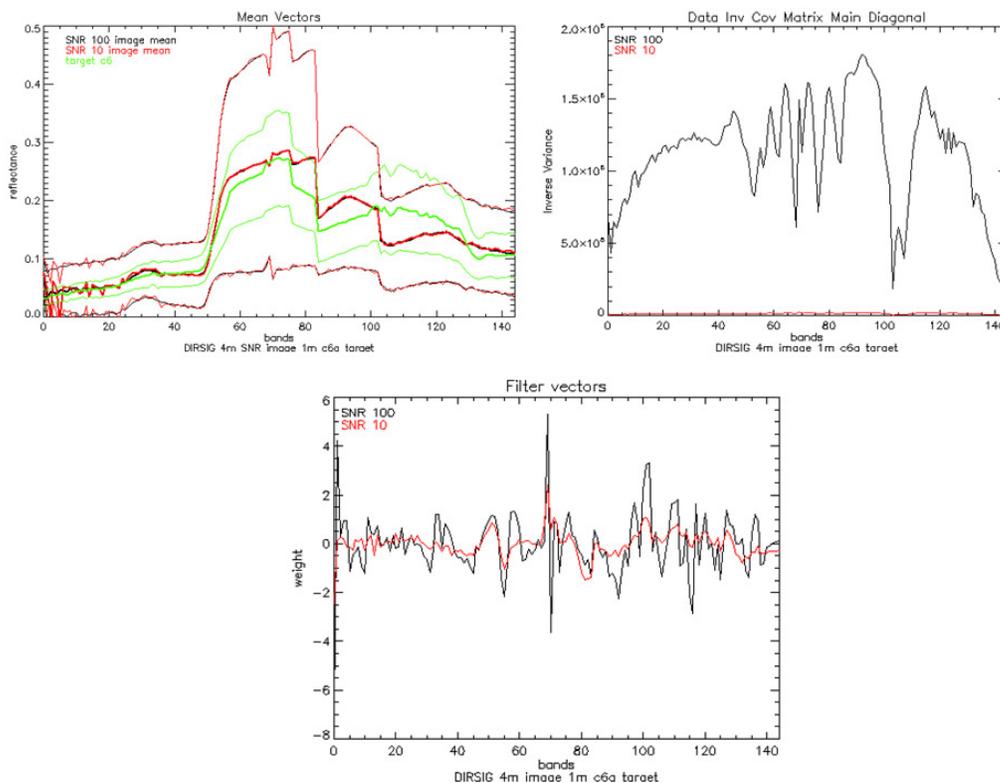


Figure 7.21: Mean spectra (top left), inverse variances (top right), and filter vectors (bottom) for 1 m c6 target in DIRSIG 4 m 100 and 10 SNR images

The somewhat smaller weights of the lower SNR case produce different results than those of the higher SNR case. Figure 7.22 shows the wider spread (two to three times larger than the variance of the 100 SNR image) of the target absent and present distributions

Chapter 7. Results

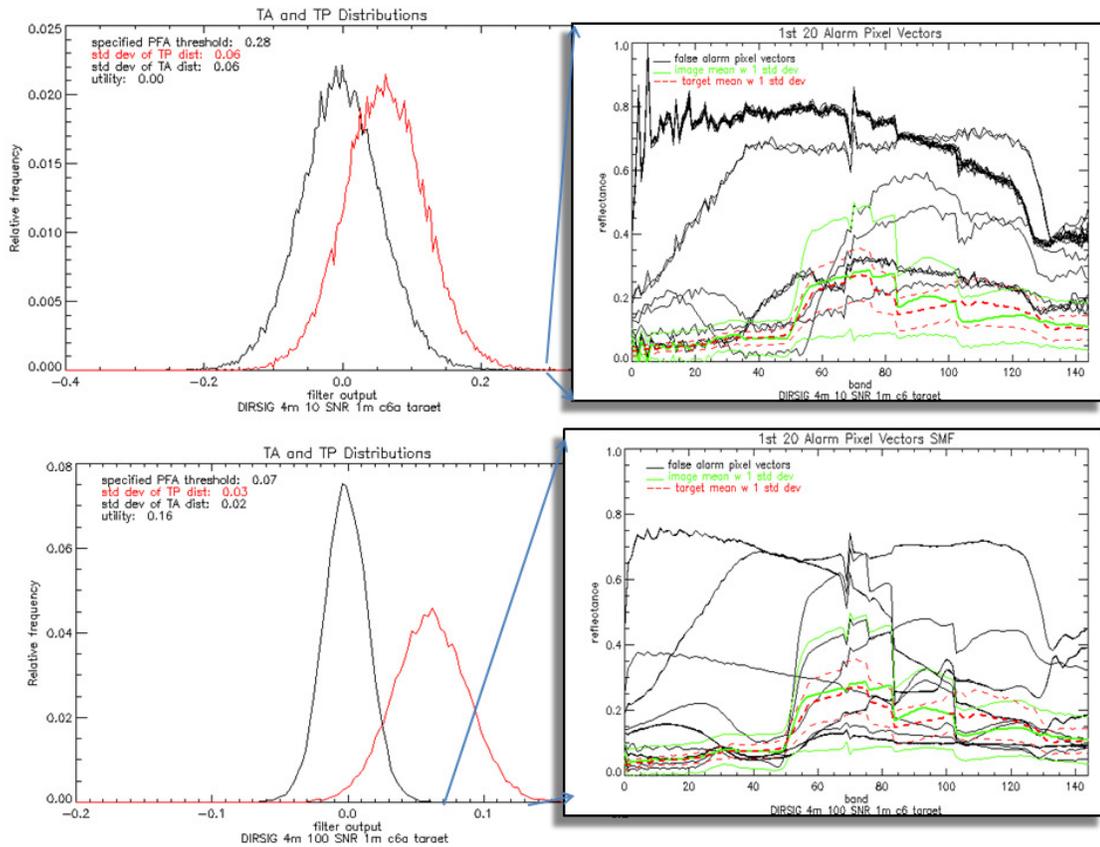


Figure 7.22: Filter output distributions (left) and first twenty false alarm spectra (right) for 1 m c6 target in DIRSIG 4 m 100 (bottom) and 10 (top) SNR images

associated with the 10 SNR image. The weights on the 10 SNR case filter vector are smaller yet they produce a wider filter output distribution. The utility of the 100 SNR image is 0.156 whereas that for the 10 SNR image is 0.004, and the target present distribution looks like a shifted version of the target absent distribution for the noisier image. This may be due to the fact that the noisier image spectra take a much larger fraction of implanted target to begin to look like the target, and at this relatively small target size (fraction), the target absent and present situations look almost identical. The right side of Figure 7.22 shows that there are different spectra involved in the early false alarms, depending on the SNR. Spectrally we can see this, and spatially we see that only two pixels are picked up in the first 20 false alarms in both images. Figure 7.23 shows the spatial distribution of the first 20 false alarms in each of the two SNR images. The high SNR (indicated as red) image strongly false alarms on the tennis courts, particularly on the border where there appear to be mixed pixels between the red and green court paints. The lower SNR (indicated as blue) does not have a discernable spatial pattern associated with its false alarms.



Figure 7.23: Spatial distribution of first twenty false alarms for 1m c6 target in DIRSIG 4m 10 and 100 SNR images

7.1.4 Versatility of the Utility Metric

In Section 7.1.2, we briefly investigated the behavior and characteristics of the utility metric with respect to target detection scenario parameters. We now examine perhaps the biggest strength of the target implant method of assessing utility – the ability to summarize a wide variety of target detection parameters to answer the specific informational needs of a particular image analyst. The starting point is the utility which is calculated for a single sample in an infinite space of possible utilities associated with the unique combination of image, target size, target type, specified PFA, and detection algorithm. If an analyst is interested in one specific combination (e. g. “What is the utility of the FR I run05 image for a 0.7 m c6 target at a specified PFA of 5×10^{-4} using a spectral matched filter?”) then the application of the utility metric methodology demonstrated thus far will answer his or her question. If, on the other hand, the analyst seeks a more general answer about the image (e. g. “What is the average utility of the FR I run 05 image for targets f2, f4, f8, and vf124

ranging from sizes of 0.05 to 0.8 m at a specified PFA of 5×10^{-4} using three different detection algorithms?") then we need to be able to summarize the 120 (4 target types x 10 implant fractions x 3 detection algorithms) individual utilities in a single metric. And if the analyst seeks a very specific range of information (e.g. "What is the utility of the FR I run 05 image for detecting the f4 target in sizes from 0.4 to 0.6 m?") then we need the ability to delve into many possible parameter combinations, retrieve the utilities of interest and summarize them to fulfill the specific information requirement. The method that we have outlined in the Section 6.3.3 for this sort of "utility mining" consists of applying windows to the data to answer the question posed by the analyst. In this section we demonstrate some results of applying these windows and summary "roll up" procedures to allow the analyst to answer questions about the target detection scenario parameters. We feel that this is very pertinent to the task of the analyst in exploiting information from the image, and that this tool to assess utility affords the flexibility to meet those needs. In the examples that follow, we look closely at the specified PFA, the target type, the target size, and the detection algorithm. Our investigation involves a single image and a range of four target types, ten implant fractions, three detection algorithms, and four specified PFA windows. Of course, this parameter space can be as large and robust as the analysis task requires. We offer a methodology for extracting information from the image in the form of a utility metric and hope that the examples will illustrate the great potential of this methodology.

7.1.4.1 Specified PFA

The ability to adjust the specified PFA range when calculating the normalized utility would be helpful to an analyst investigating how an image might behave in a more stressing detection situation (lower PFA or very restricted range of PFAs). This target detection parameter is slightly different than the others in its application because it is fundamental to the calculation of the utility metric. Thus, adjusting the specified PFA range and weighting scheme changes the utility calculated for each target detection situation. We first consider the four windows that could be applied to the ROC curve in calculating the utility and then show the effect of each of the windows on the four targets at ten implant fractions for the spectral matched filter.

The choices of window functions are a rectangular function from the lowest PFA to a specified PFA, a decreasing exponential, a Gaussian, or a delta function (that will pick out a

specific PFA, so that we have the utility metric analogous to PD at a specified PFA). Each of these windows has a different situation in which it might be desirable to apply it. The rectangular window function is the one used in this research, and it equally weights all PFAs between the low and specified PFA in an attempt to convey that the first false alarm is just as important in calculating utility as the false alarms associated with the specified PFA. The decreasing exponential is a means of placing the most emphasis on a particular PFA and then decreasing the significance of larger PFAs until the weights are negligible in the vicinity of the specified PFA. The Gaussian window places the most emphasis on the specified PFA, but provides equally decreasing weight to PFAs larger and smaller than the specified PFA. Finally, the delta function returns the PD associated with the PFA of interest, and it offers a readily understandable metric of performance, the detection at a constant false alarm probability. Figure 7.24 shows these four windows plotted on the PFA axis. In this example, the specified PFA is 5×10^{-4} , and the Gaussian is centered on 1×10^{-4} while the delta function sits at a PFA of 3×10^{-4} .

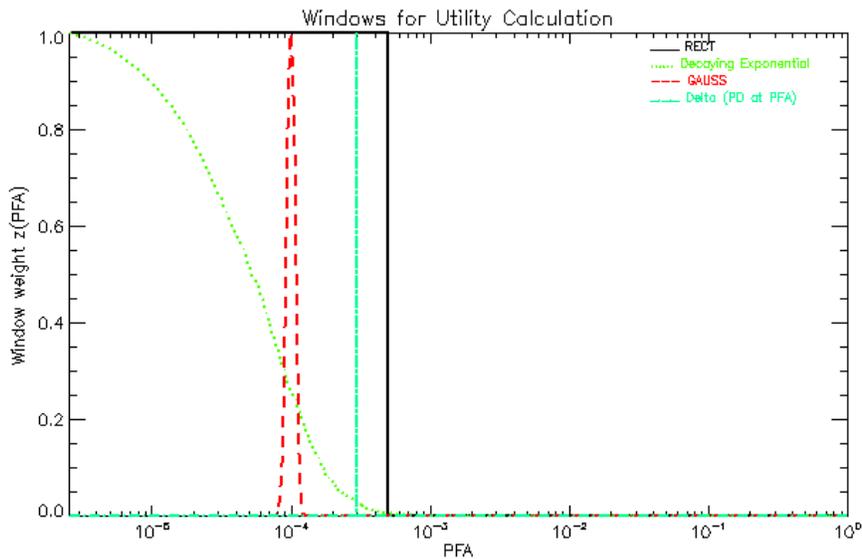


Figure 7.24: Window functions for utility calculation

The effect of varying $z(\text{PFA})$, the utility window function, on the assessed utility for a particular situation is shown in Figure 7.25. We show its effect on the overall summary rollup of the target detection parameter space (average over four target types and three detectors). In general, we see that the rectangular window applied to the specified PFA produces higher utilities than the Gaussian centered at 1×10^{-4} and the decaying exponential. This is expected since the Gaussian and exponential capture a smaller part of the ROC curve,

and the commensurate smaller area does not offset the small PD values that are being most heavily weighted. Thus, they appear smaller because they are emphasizing more difficult detection situations. The PD at PFA produced by the delta function window at 3×10^{-4} always produces the highest utility. This is because the PD at PFA is indeed sampling the PD curve and is not being normalized by a “perfect” detection area, as the other utilities were. While this measure is not exactly the same as a normalized utility, it is included because it is intuitive and simple to calculate. For the remaining target detection scenario parameters, we will maintain the rectangular window function for calculating utility as the standard for simplicity. This section has shown the versatility that may be applied to calculation of the utility metric by applying different windows centered at any PFA of interest.

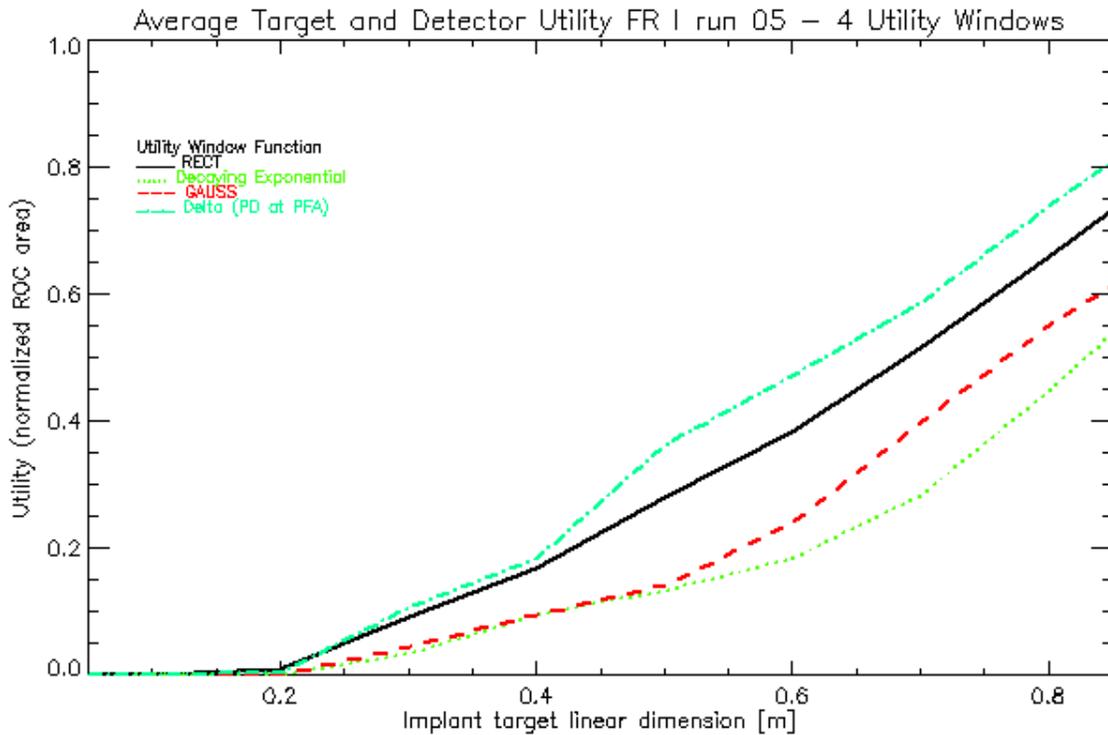


Figure 7.25: Effect of utility window function on summary utility

7.1.4.2 Target Types

The utility of an image greatly depends on the type of target for which utility is being assessed. If an analyst were to seek to summarize the utility of the image due to a range of targets, the simplest method would be to create a weighted summation of those utilities, and use the weights as a means of emphasizing the targets of interest for the utility metric. In the previous section, we demonstrated the average of all four targets as a summary. If we choose

to work with only the spectral matched filter detector, we demonstrate here that we can apply a weighting scheme so that we emphasize the targets based on their detectability. In Figure 7.26, we show two plots of SMF utility versus target size. One corresponds to a weighting scheme that places emphasis on targets 0 and 1 (using a weight vector $[0.5,0.5,0,0]$) and another corresponding to a scheme to place emphasis on targets 2 and 3 (weight vector $[0,0,0.5,0.5]$). The weighted average that emphasized targets 0 and 1 produces a lower utility than the one emphasizing targets 2 and 3. From an analysis perspective, the conclusion is that the FRI Run 05 image is less useful for finding targets 0 and 1 than for finding targets 2 and 3. While this example may seem simple, consider a library of hundreds of targets and the

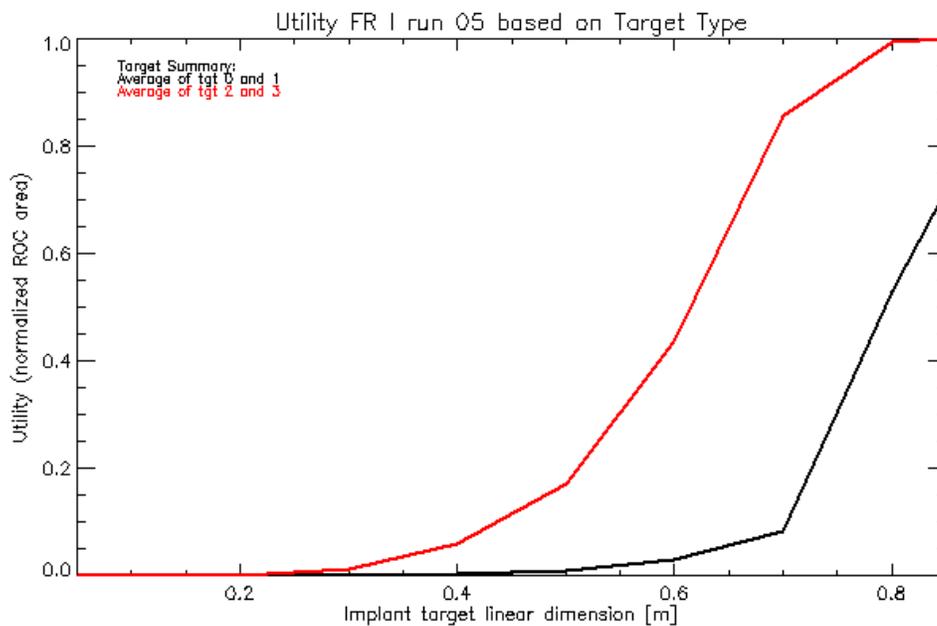


Figure 7.26: Two summary utilities for target type

helpfulness of being able to generate a utility summary for only those targets of interest in a given situation. Keep in mind that we are performing a weighted summation on the utilities once they have been calculated, and not the individual target spectra before the utilities have been calculated.

7.1.4.3 Target Sizes

The previous plots have presented utility as a function of target size, since this is an intuitive interpretation (bigger targets are easier to detect). The question now is, “what if I not only want to detect a certain size target, but a range of target sizes?” The utility of a specific size target is found by sampling the utility versus target size plot at the size of interest. There are

two fundamental approaches available. The first looks at the utility of a range of targets that could be found by a weighted summation, as with target type and detector. The weights on the different discrete samples of the utility at specific target sizes would determine which target sizes would be emphasized. Thus, all targets could be averaged, or only targets up to a specific size, or a range of targets of specific sizes. The window function that would be applied to the summation operation would determine the nature of the summary. Table 7.3 shows some of the possible iterations on this theme. Provided that the target sizes are evenly

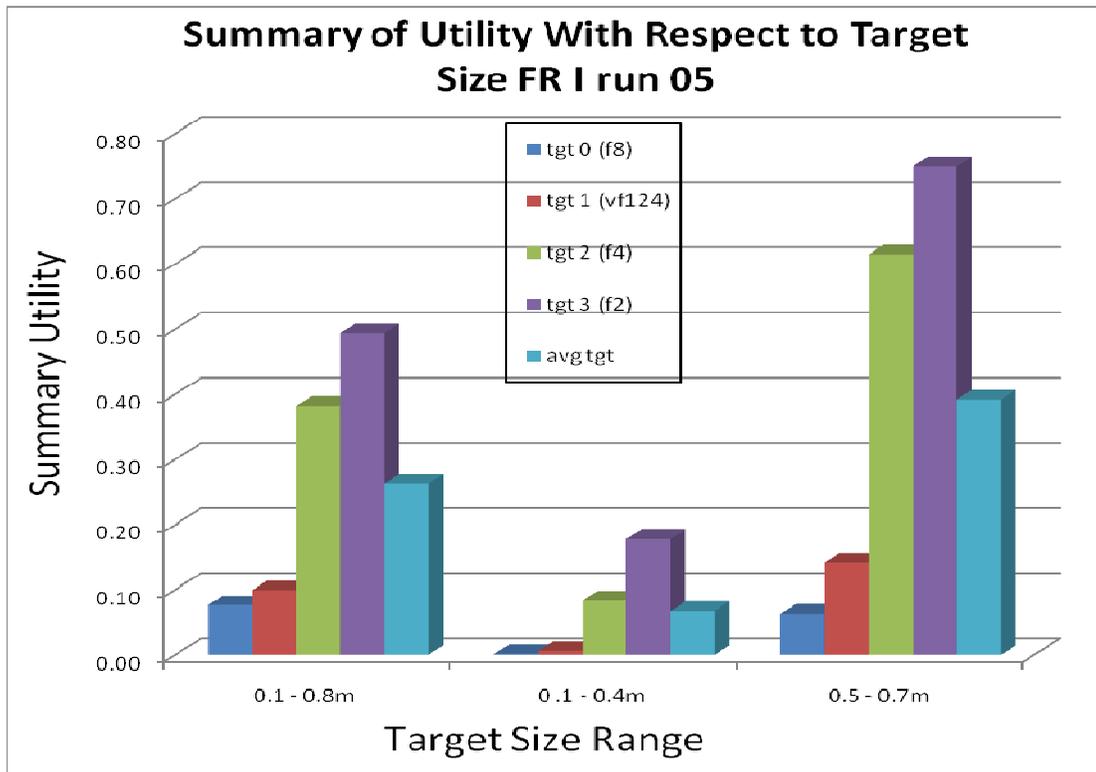


Table 7.3: Summary of Utility Using Target Size Ranges

spaced so that we do not inadvertently weight any particular size distribution, this approach should fairly characterize the utility. Note that these results are taken as an average of the three detector utilities. We see that the summary utility is higher for the 0.5 – 0.7 m range than for the 0.1 – 0.8 m range. This is due to the equal weight applied to all sizes in the larger range, whereas in the smaller range, just those values are emphasized. This means that on average, the utility is lower for a range of targets from small to large than for just large targets.

The alternative implementation would be to calculate the area under the utility versus size plot and then normalized it relative to perfect detection up to that particular size. This

would produce a summary metric between 0 and 1, and would seem to capture information about the shape of the utility versus size plot. This would have the advantage of creating an even size scale due to resampling of the vector. It is not intuitive, however, as to what the area under the utility versus size plot conveys. While we performed some investigations into this normalized area under the utility versus area plot, we opted for applying the simpler weighted summation of utilities at discrete sizes as a way of characterizing utility with respect to target size.

7.1.4.4 Target Detection Algorithms

The utility of the image as assessed by different algorithms may be summarized with a weighted summation of the utilities resulting from each algorithm. We saw the effect of averaging the performance of all three detectors in the above examples. In Figure 7.27 we average across the four targets and compare the utilities for the images as assessed by each of the three detectors. This could be simply summarized by averaging across all three or a more elaborate weighting scheme that placed more weight on a particular detector based on confidence in that detector's robustness. This would be helpful for image analysis because of the ability to summarize a wide range of detector performance in a single metric. It is

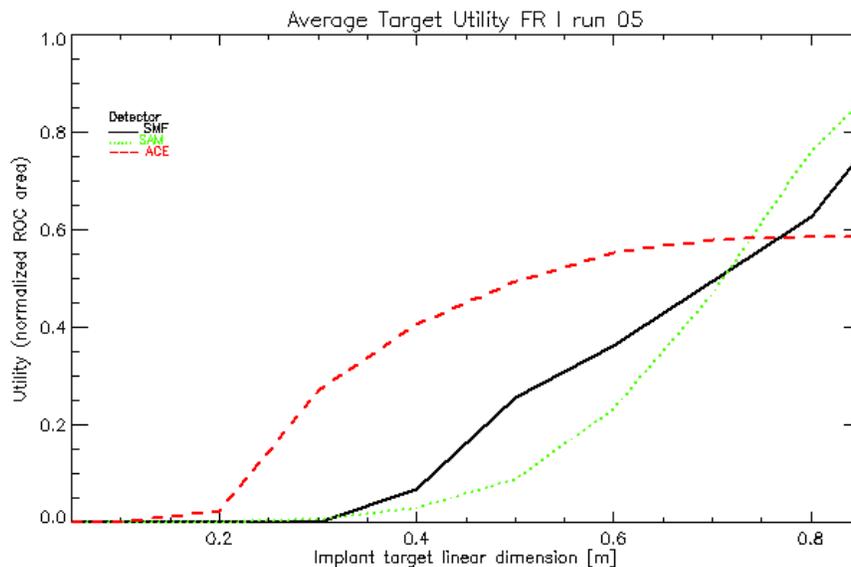


Figure 7.27: Detector type effect on utility

interesting to note the utility of each detector for different size targets. The ACE, which does well at suppressing background, clearly excels at the subpixel detection task when the target implant size is small. The SAM behaves oppositely, detecting fully resolved targets very

nicely, because it was not constructed with subpixel detection in mind. The SMF represents a performance somewhere between the SAM and ACE detectors. Note that the plots show the average utility over the four targets, and results may vary based on the specific target. Note also that all four of the targets we seek are actually in the scene, and their presence causes a depressing of the assessed utility.

7.1.5 Comparing Images Using the Utility Metric

Having established the fundamentals of utility metric behavior for individual target detection scenario parameters and for summarizing utility over ranges of parameters, we now apply these features to actually comparing spectral image utility. We draw on a collection of eight real and nine DIRSIG images. The real images represent six different sensors, including multispectral imagers. The underlying assumption is that we have a reflectance image for which we know the atmospheric absorption channels and the spectral response functions, since we need to match the reference target library spectrum to the image in order to build the detection algorithm. We begin by comparing images made by different sensors, then images of different scenes made by the same sensors, and finally images of the same scene made under different image acquisition conditions. The purpose is to demonstrate the application of the utility metric to a variety of image types and scenes.

7.1.5.1 Different Sensors

In this section, we examine the utility of six real images (one from each sensor) in order to highlight different sensor impact on the resulting image. However, we must keep in mind that the different scene content and image acquisition conditions also contribute to the image utility. One point of consideration is how to equitably present the utilities of sensors with different spatial resolutions. We can either use the implant fraction or the target linear dimension. We choose the target linear dimension because it is a feature that is intuitive in considering a target detection problem. It facilitates comparison of different images by addressing the utility associated with the same size target. If we do not do this, then the 100% fractional implant for the HYDICE image deals with 0.8 m targets whereas the 100% fraction for ALI is a 30 m target. By deciding to work with a constant target size instead of fraction, a 0.8 m target in a HYDICE image represents a target implant fraction of 100%, whereas in the ALI image it represents a fractional implant of 0.07%. This more accurately reflects the

reality that it is harder to find the same size object in a larger GRD image. The one problem that arises is how to address the utility associated with target sizes that become greater than the GRD of the image under consideration. How, for example, do we address the 30 m target size for the HYDICE image? Because the detectors that we use in this research treat each pixel vector independently and make no assumptions about spatial correlations, we cannot use the advantage of a multipixel target. We could attempt to compensate somehow by rewarding smaller GRD images when being evaluated for very large targets, but we feel that any such technique would be an arbitrary departure from the stated intent of using the detector for the subpixel detection task. While this may seem to be unfairly penalizing smaller GRD images, it is a consistent approach. Thus, in the plots that we show in the succeeding sections, the utility will remain constant after the target linear dimension reaches the GRD of the image.

Figure 7.28 shows the plots of utility against target implant size for the images under consideration. These show the utilities averaged over the targets, but only for the spectral matched filter. This is done for the sake of simplicity and clarity of presentation. The top of the Figure shows the target size axis up to 30 m. The bottom right shows utility against target sizes up to 5 m and the bottom left shows utility against target sizes up to 1 m. Clearly, the effect of target size is significant in the conclusions that we make about the utility of a particular image. The target size is a defining characteristic of the target as is its spectral nature relative to the background. The HyMap and AVIRIS images are the only ones that attain a utility of one before the target size reaches their image GRD. This is indicative of the ease of finding these four targets in these two images. While the HYDICE image clearly dominates the utility at target sizes smaller than 1 m, it does not attain a utility of one when it reaches 100% target implant fraction. The vf124 and f8 targets prove so difficult for the SMF that they pull down the average utility. The MISI image exhibits a slightly higher utility than the Quickbird image, probably because of the additional spectral channels that assist with the detection. The AVIRIS image shows its strong utility by outperforming the six times smaller GRD images of the Quickbird and MISI by the 5 m target size. Again, the AVIRIS spectral information of the wider range may be the reason for this higher utility. The ALI image slowly increases its utility with target size greater than 15m and then attains a 0.7 utility at full pixel fill. Another way of facilitating this image comparison is with a table of average

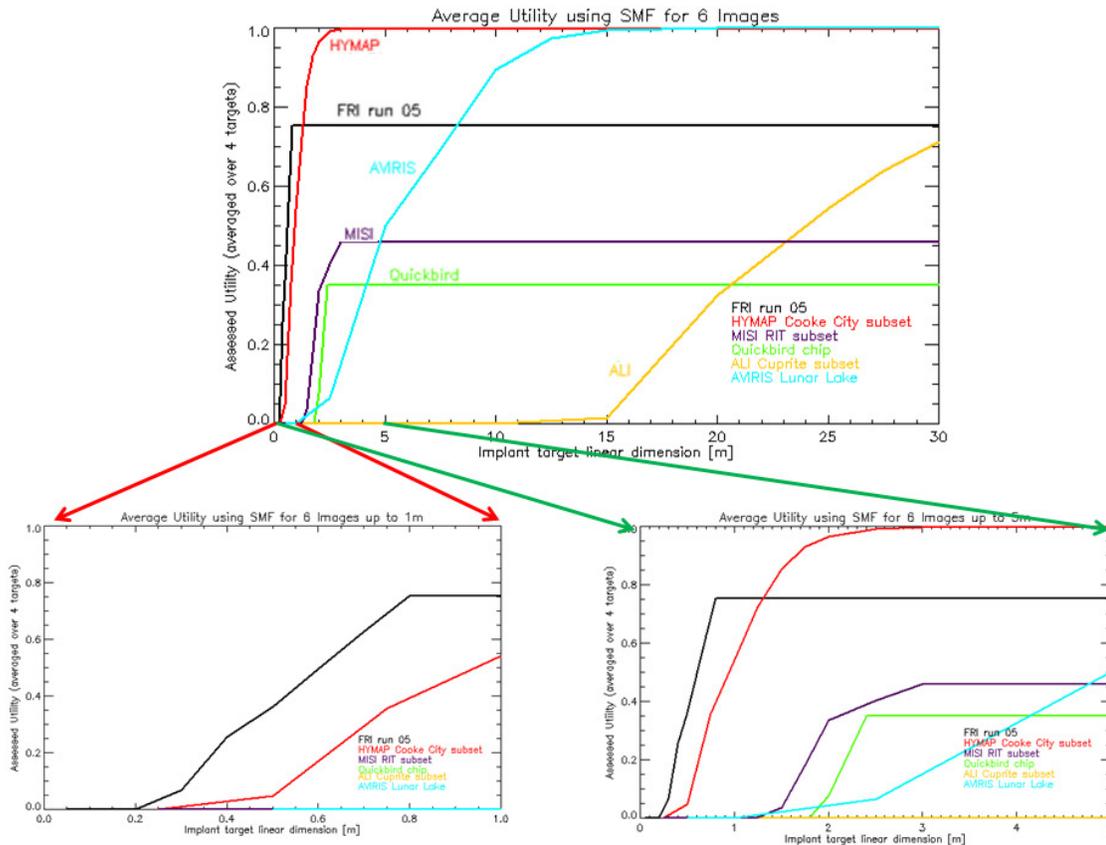


Figure 7.28: Utility of six spectral images plotted against target size

utility with the average taken over the range of target sizes from the smallest up to the upper limit. The values in Table 7.4 are the weighted average over a certain range of the sizes for the six images. The table clearly bears out what we saw from the plot of utility versus target

Image	Target Size Range		
	0 - 1m	0 - 5m	0 - 30 m
HYDICE FRI run 05	0.36	0.68	0.74
HYMAP Cooke City	0.22	0.83	0.97
MISI RIT	0.00	0.29	0.43
Quickbird Esparanza	0.00	0.21	0.33
ALI Cuprite	0.00	0.00	0.21
AVIRIS Lunar Lake	0.00	0.17	0.81

Table 7.4: Average Utility Summarized over Different Target Size Ranges for Six Images

size. We point out that the comparison needs to be put in perspective. This is for four targets only. The results could be very different if other targets were considered that reacted very strongly with a particular image (detection in that particular image was very difficult or very easy due to the spectral nature of the target and background). Also, the utility is only as good as the reflectance retrieval and the knowledge of the GRD for a particular image. If the retrieval is bad, then the utility will not be accurate, since the target we are injecting will look

very different from the image pixel vectors. The GRD incorrectly applied will cause the fraction for target implant to not be correctly matched to the true fraction. But while we must keep these in mind, the results confirm intuition for the most part. There is a tradeoff between GRD and spectral channels. The fact that the HyMap spectral channels are 15 nm as opposed to 10 nm and the GRD is 3 m does not seem to hurt this image from being the top ranked image. The large GRD of AVIRIS only hurts the ranking of this image when the target size is small, but for large targets, it clearly produces one of the better utilities.

7.1.5.2 Comparison of Different Scenes Using HYDICE Images

We would like to be able to compare different images captured by the same sensor to explore the role of scene composition on the utility. We showed this in Section 7.1.3.1 for HYDICE FR I run 05 and DR II run 03. We demonstrate it again here using HYDICE, except we do not know the GRD of the terrain and urban images. In order to deal with this lack of information, we will look at these utility plots in terms of the implant fraction. With knowledge of the GRD, the individual plots may be translated into having an x axis in terms of linear dimension. Without it, we use the fractional implant as the x-axis. We show the terrain and urban results here along with the FR I and DR II utilities in Figure 7.29 for completeness and to demonstrate the flexibility of the utility metric. Although the GRD and SNR may be different for these two images, we hope that this investigation gives us insight into the effects of scene composition. The plots are averaged over the four standard targets using the SMF. It is interesting to note that the urban image should give such high utility, something we would not expect due to the more complicated nature of the spectral character of the image. Part of the reason may be that the scene does not contain targets for which we are looking (like FR and DR), which we have shown to degrade the utility. Noting how the shape of the urban and terrain images is similar and that of the FR and DR is similar, it is possible that the GRD of the urban and terrain images is actually larger than that of the FR and DR. This is a feature that is hidden when using implant fraction as the x-axis.

We compare the two AVIRIS images, which show different spectral content. The utility of these images are plotted in Figure 7.30. These are comparable in their utility profiles relative to the target size. Perhaps Lunar Lake is the slightly easier detection situation because of the lack of man-made clutter with which to interfere with the target signature.

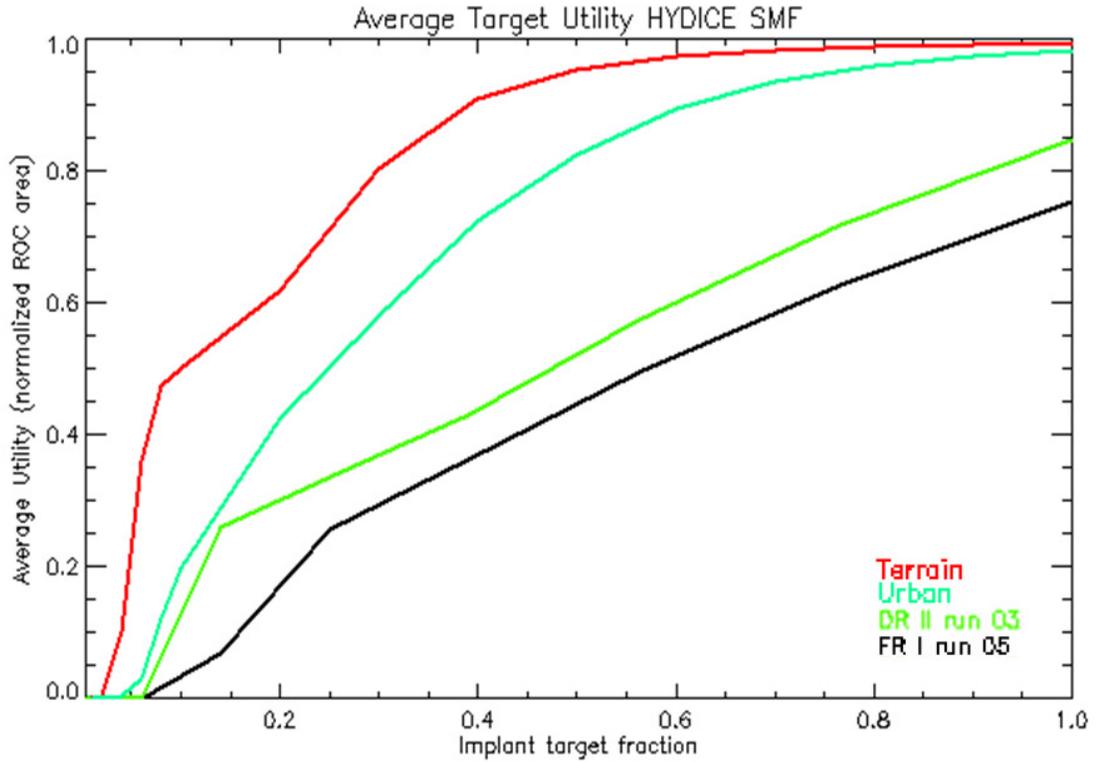


Figure 7.29: Four HYDICE image utilities plotted versus target implant fraction

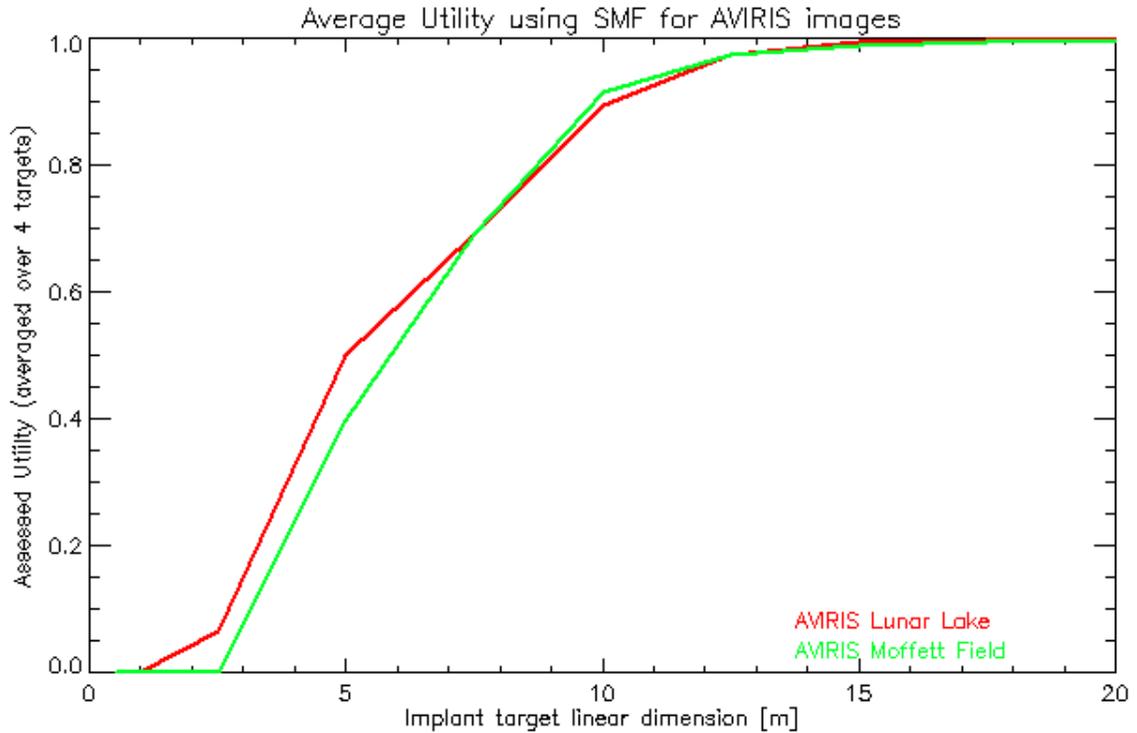


Figure 7.30: Utility versus target size for two AVIRIS images

7.1.5.3 Comparison of same scene using DIRSIG and HYDICE Images

We would like to investigate the utility associated with images of the same scene but with different sensor image acquisition parameters. We first use the three HYDICE images from different altitudes in FR I as an attempt to look at images of the same scene but using different GRD images. Unfortunately, this comparison is only valid to a certain point, since the SNRs of the images are different and each image has slightly different image acquisition parameters, such as the atmosphere and collection geometry. The scene content is roughly the same but not precisely so. We manually selected regions of the Run 07 (GRD 1.6 m) and Run 09 (GRD 3.2 m) to be comparable to the scene in Run 05. The Run 07 subset had 96,640 pixels and run 09 had 36,312 pixels. Figure 7.31 shows the results of this comparison. We see again the advantage of the smaller GRD image in detecting smaller targets. For targets larger than 1 m, the run 07 and run 09 both attain higher utility than run 05. The reason that

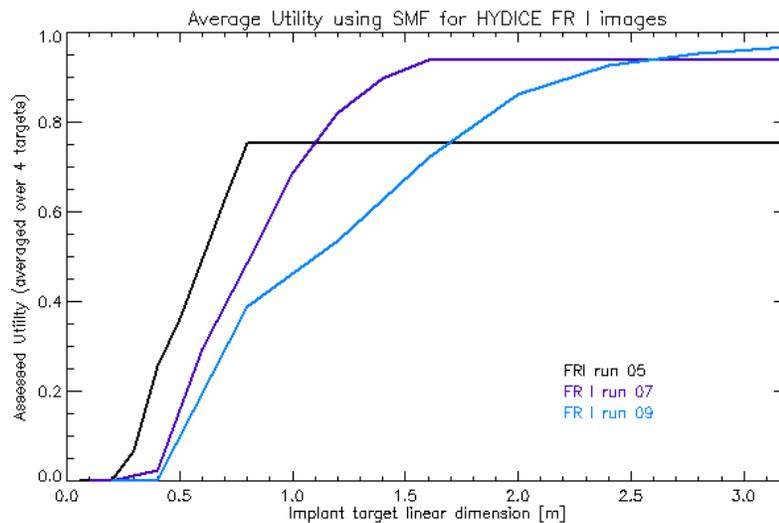


Figure 7.31: Utility of HYDICE FR I Runs 05, 07, and 09

the smaller GRD image may not have as high utility is due to the greater number of target pixels present in the scene at the smaller GRD. This number decreases with larger GRD and therefore does not interfere as much with the detector performance. The other factor is that the run 05 image is not being compensated for the multipixel advantage that it has in detection for targets larger than 0.8 m.

DIRSIG provides us with a very useful tool to carefully control the scene content and the image acquisition parameters. We use a set of nine images generated as part of an image

analyst test in 2006 to illustrate the effect of the same scene but different sensor parameters (Kerekes, et. al., 2006). Figure 7.32 shows the utility for these nine images. The resulting utility plots against target size confirm intuition about how the SNR and GSD will affect detection performance and therefore utility in these images. The effect of increasing SNR pushes the utility curves to the left (smaller targets yield higher utilities). Decreasing GRD likewise shifts the curves to the left. As expected, the 8 m GRD 10 SNR image yields the lowest utility, while the 2 m 1000 SNR yields the highest utility.

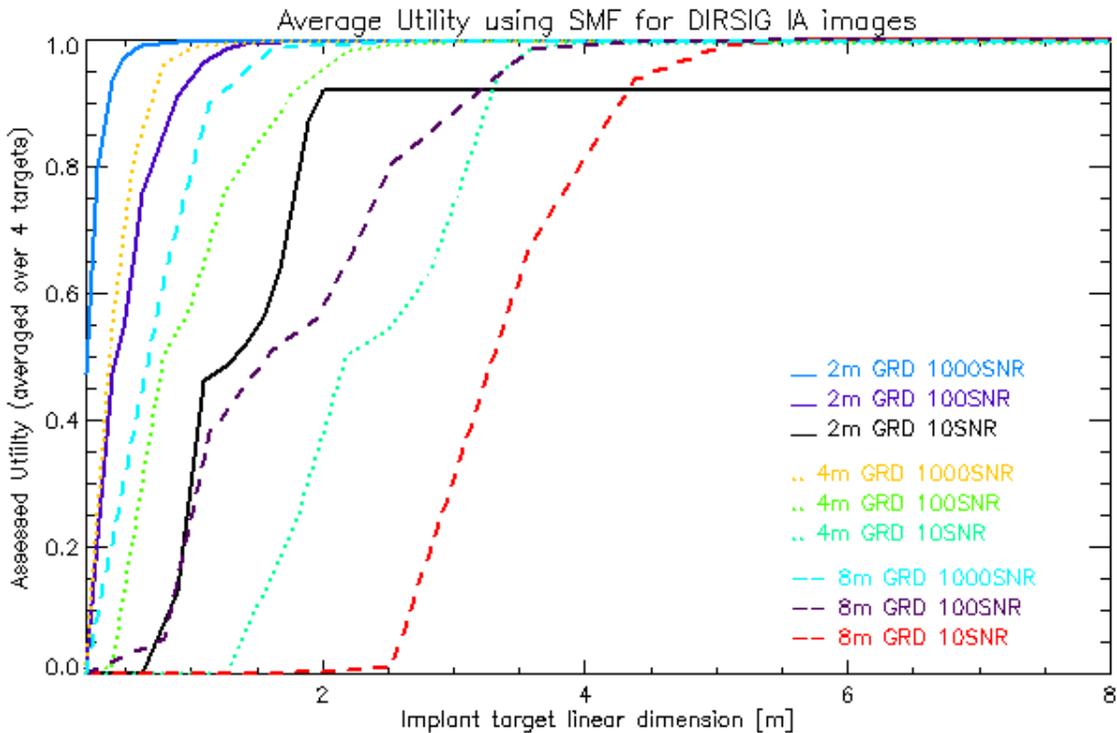


Figure 7.32: Utility plots for nine images from the DIRSIG Megascene

7.1.6 Summary of Utility Metric Results

In section 7.1, we have seen that the utility metric affords a great deal of flexibility in the result based on the specific needs of the image analysis question. In this section, we briefly recap the major characteristics of the utility metric and its advantages and limitations.

The utility metric formed by using the target implant method offers a self-contained way to investigate relevant aspects of the subpixel target detection task for any image. The resulting utility represents the overall detectability of a subpixel target spectrum in every pixel of the image. This allows us to assign a measure of target detectability, which we call

utility, to an arbitrary image. We do not need to have any target pixels in the image or ground truth about the image. The method only requires knowledge of the image spectral channels so that the reference target spectrum can be appropriately resampled to match the target, and that the image be in reflectance space. The utility metric may be adjusted to meet the specific operating requirements of an analyst, focusing on one PFA or a range of PFAs. It may be summarized across several target types, sizes, and detection algorithms. If the GRD of different images is available, then the images may be compared for their utility in finding a particular size target.

We confirmed that the utility metric responds with higher utilities for images having smaller GRD, higher SNR, better spectral resolution, and targets that are spectrally distinct from their backgrounds. We also saw that the utility is a metric that involves several factors, and defies easy explanation. The impact of scene composition is not clear, and we saw that the presence of a few target pixels of the target being sought will drive the utility down significantly. This is a limitation that needs to be considered, and may be best addressed by examining many different combinations and employing summary metrics and using the utility metric in a relative, not absolute, sense.

7.2 Image-derived Utility Prediction Results

The assessed utility for a linear target detector can be predicted using statistical models of the filter output. This is because the linear transformation of the detector acts on the multivariate statistics of the original image to create the statistics describing the scalar filter output. This is the idea underlying the FASSP model. The difference between FASSP and our image-derived approach is that we actually estimate the statistics from the image for which we are predicting the utility rather than use a notional description of a scene using class statistics. This section examines several aspects of our approach to predicting the utility. First, we closely look at the predicted utility for the standard four targets in three real images to understand how the predicted utility behaves. Because it is directly comparable with the assessed utility, our results show the assessed utility alongside the predicted utilities. We next discuss the specific nature of each of our four prediction models by quantifying the degree of fit to the assessed utility at the probability vector level. We next address the issue of evaluating the accuracy of the prediction relative to the prediction. We then look at the

sensitivity of the predictions to various parameters of the prediction models. Next, the tradeoff between time savings and the accuracy of prediction is examined in an attempt to lend some operational relevance to this prediction methodology. In order to appreciate the larger context of how our image-derived prediction method fits in with the larger body of spectral image quality methods, we use several real images to compare results between methods to determine if there are any commonalities in the diverse methods. Finally, we summarize the major characteristics, advantages, and limitations of the image-derived prediction method.

7.2.1 Behavior of Predicted Utility for Three Images

We are interested in investigating how the utility prediction behaves with respect to our assessed utility for three images. We select the three images to represent the range of utilities seen in Section 7.1. The first image is the HYDICE FR I Run 05 because it is a baseline that is well characterized. The next image is the HyMap image, which is a very high utility image for the four targets being investigated. The third image is the Quickbird image because its utility represents the low end of the range. We seek to gain a general sense for how prediction can be different for a different image, although we are not trying to compare image utility in an absolute sense by looking at the same-sized target. We are interested in comparing the assessed and predicted utilities for the same image, not between images. We use the same standard diagnostics as with assessing the utility, so that we can have an equitable comparison. The 4 targets (f8, vf124, f4, and f2) and range of target sizes (this range is really driven by the constant fractional spread from 10% up to 100%). We also maintain the same specified PFA of 5×10^{-4} for all of the images considered.

7.2.1.1 Predicted Filter Output Distributions

We have four statistical models with which to model the filter output distributions. We can use the target absent (TA) and target present (TP) distributions as a very simple explanation, but only a qualitative level. Figure 7.33 illustrates this point using the f8 target in the FR I image. It shows the assessed distributions in the upper left, and then presents the distributions associated with the normal, sum of Gaussian classes, t-distribution, and composite models introduced in Section 6.4.2. We can see that the different prediction models do a relatively good job at matching the assessed distributions in a gross sense. Upon closer examination,

Chapter 7. Results

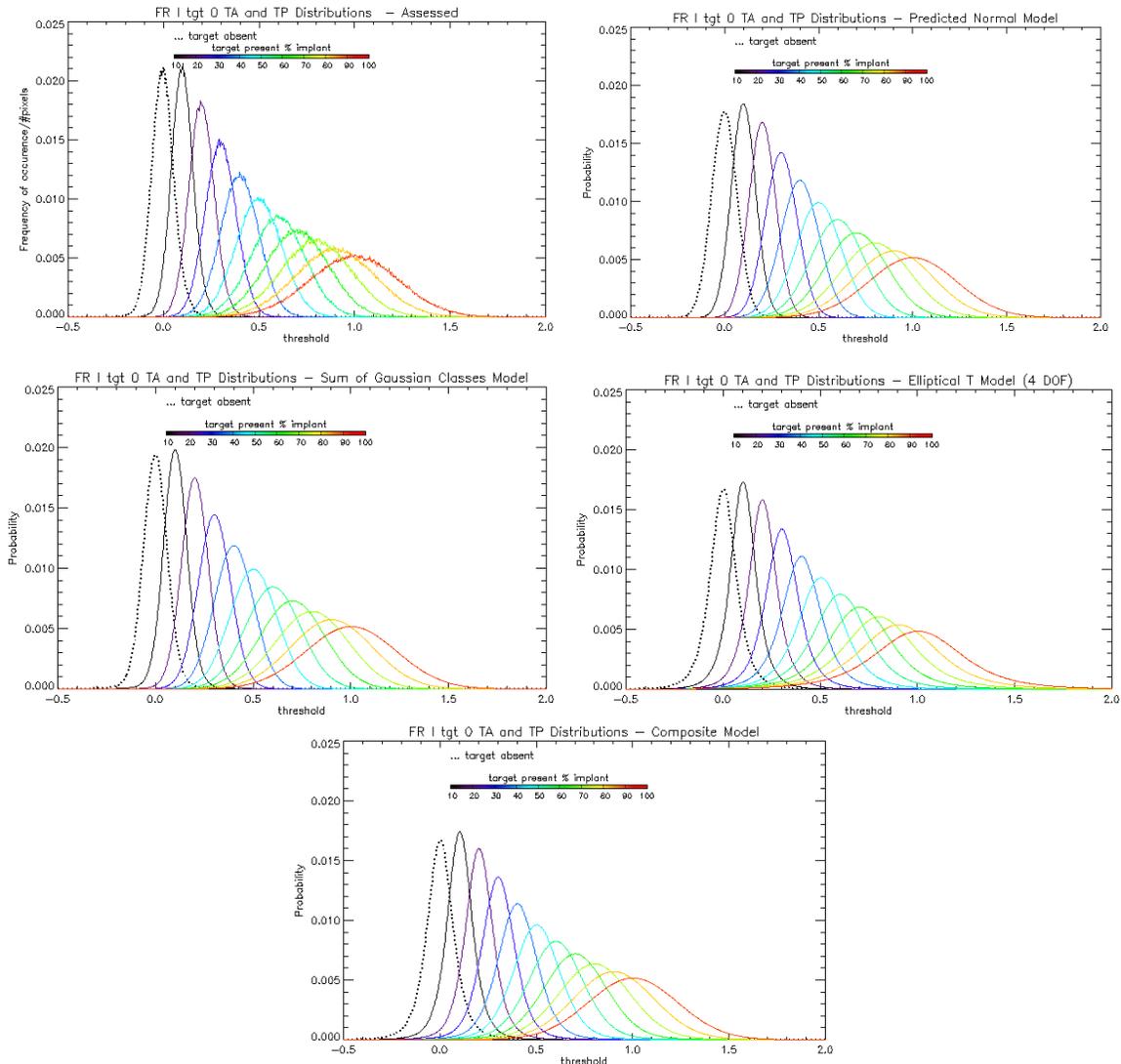


Figure 7.33: Assessed and predicted filter output distributions for target f_8 in FR I Run 05

we see that the magnitudes of the probability density functions predicted by the models and attained in the relative frequency by the assessed distributions are not quite the same, with the assessed having slightly larger values. Ultimately, the distributions are not very informative in telling us how good a job of predicting that we have done because they do not focus on the important part of the distribution, the outer tails. We could focus on distribution tails, akin to how we examined the false alarms in section 7.1, but instead we will look at the family of ROC curves generated by these distributions to help answer the question of the goodness of our predictions.

7.2.1.2 Predicted ROC Curves

The ROC curves result from plotting the PFA and PD vectors. Figure 7.34 shows the family of ROC curves generated for the 10 implant fractions of target f8 in the FR I Run 05 image. Note how the assessed ROC curves are pushed to the right towards larger PFAs before producing significant detections. The predicted ROC curves have a difficult time in matching this behavior. This is because of the prevalence of the early false alarms that occur on the far right edges of the target absent distribution. Accurately characterizing behavior in this region, the distribution tail, is the challenge to any prediction method that we may develop. The ROC curves have been plotted on a semi-log scale to correspond to the range of lowest to

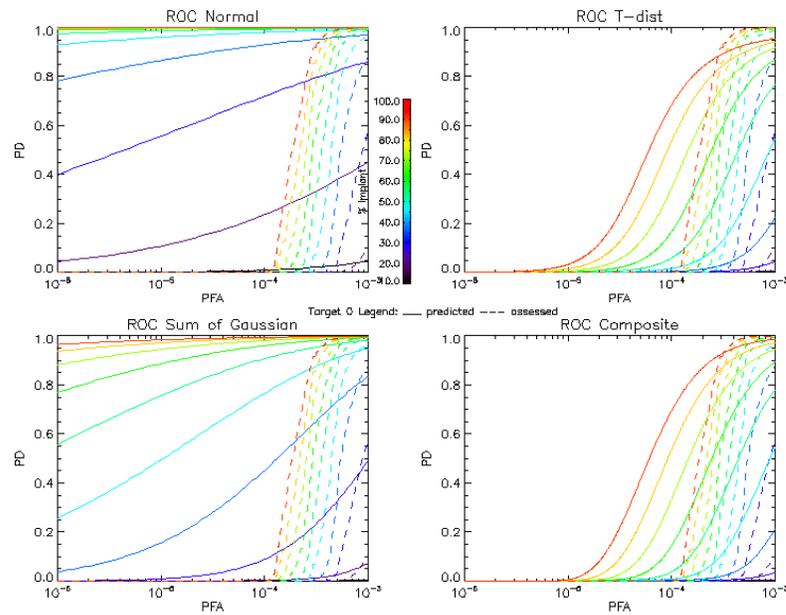


Figure 7.34: Assessed and predicted ROC curves for target f8 in FR I Run 05

specified PFA, so we are looking at the probabilities resulting from the far right tails of the distributions. If we look at these ROC curves, we may make a few observations about the accuracy of the predicted ROC curves relative to the assessed ROC curves. The normal model clearly overestimates the ROC curves for all fractions. The sum of Gaussian model overestimates them, but not as badly as the normal model. This is because the true image statistics are better represented by a linear mixture of the class statistics rather than the global image statistics. The T-distribution model using the global image parameters and the optimal DOF does a much better job at matching the ROC curves for very low implant fractions, but still displays a large divergence from the assessed ROC curve at large implant fractions. The

composite model most closely matches the assessed ROC curve. The different appearance of the 100% fractional implant ROC curve is because the composite model is all normal PD at 100% fractional implant. The PFA vector used in generating the composite ROC curve is from the T-distribution, which explains why it does not look like the normal model 100% ROC curve. The ROC curves illustrate the nature of the prediction accuracy relative to the assessed utility than the distributions.

7.2.1.3 Predicted Utility

We now move up from ROC curves to the utility plotted as a function of target implant size. In the following paragraphs, we explore the utilities for all three images and all four targets. When we plot the utilities resulting from a specified PFA of 5×10^{-4} for the ROC curves of Figure 7.34 and plot these against the target implant size, we obtain the plots of utility versus target size in Figure 7.35. For all targets, the normal and sum of Gaussian absolute error in utilities are the largest at all target implant sizes. The T-distribution and composite models are the closest to matching the assessed utility over all fractions. In target 3 (f4), the normal model utility is 1.0 for all fractions, and therefore not visible on the plot. These plots give us a sense for how well different prediction methods “track” the assessed utility as a function of target size. Note that each target has its unique character not only in the shape of the plot but also in the ability of the prediction to follow the assessed utility.

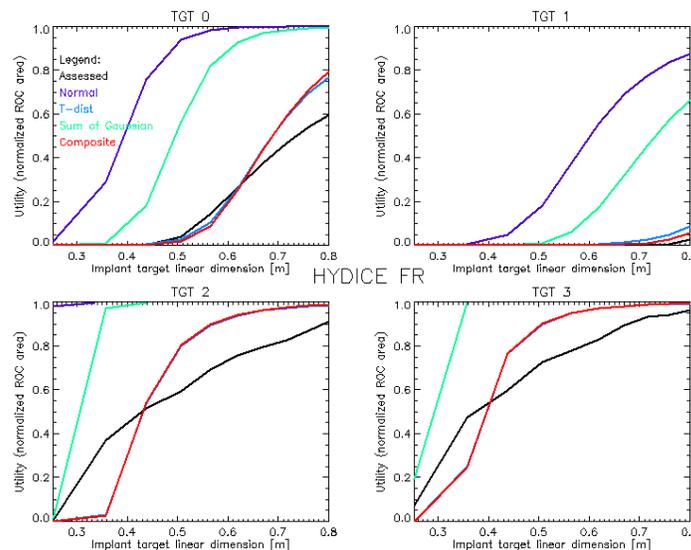


Figure 7.35: Assessed and predicted utility versus target size in FR I Run 05 for four targets at a specified PFA of 5×10^{-4}

Figure 7.36 shows the predicted and assessed utilities for the HyMap image for each of the same four targets over the range of 10 – 100% fractional target implant (translated to

equivalent linear target dimension). We see the same trend in the normal and sum of Gaussian models overestimating the utility as with the FR image. The T-distribution and

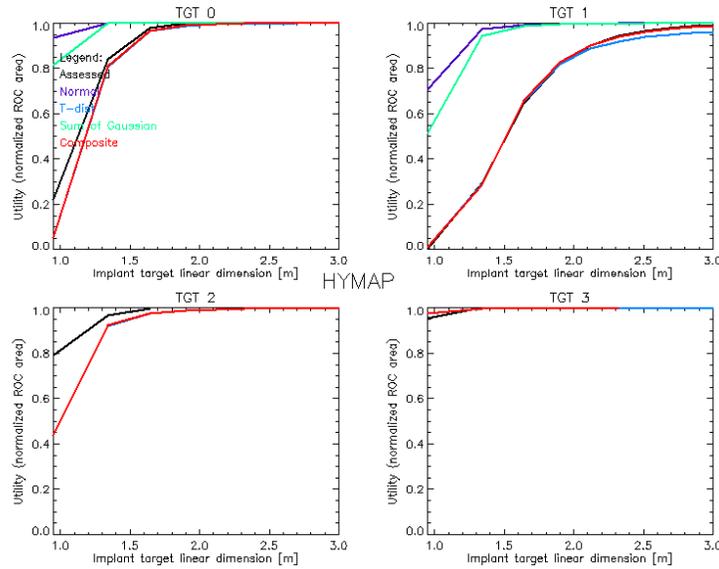


Figure 7.36: Assessed and predicted utilities versus target implant size in HyMap image for four targets at a specified PFA of 5×10^{-4}

composite model do the best job at matching the assessed utility. Note that the utilities in these plots are much higher than those in the FR image.

The Figure 7.37 shows the utility of the Quickbird subset for the same four targets plotted against target implant fraction from 10 to 100%. The utility in this image is much lower than the HyMap image. The image proves to be very low utility for Target 0 (f8) and 1 (vf124) in particular.

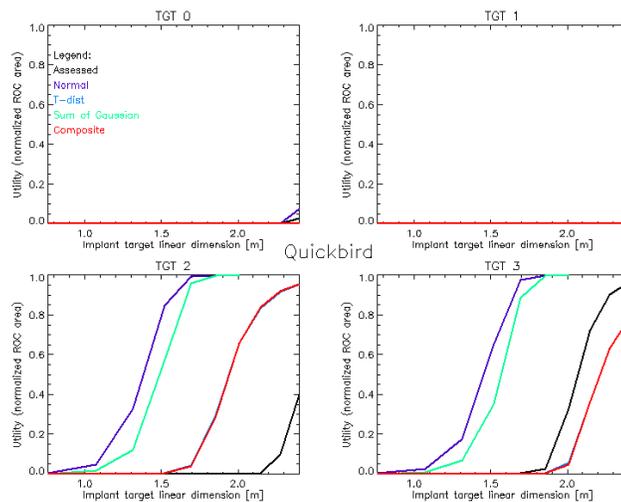


Figure 7.37: Assessed and predicted utilities versus target implant size in Quickbird image for four targets at a specified PFA of 5×10^{-4}

7.2.2 Exceedance Metric

We have seen that different prediction models will produce different target absent and target present distributions, resulting in different ROC curves and thus different utilities for each fractional implant. We have observed that some of these models produce lower utility error than others. But we have not been able to understand why this is the case. In this section, we strive to understand why certain models do a better job than others at matching the assessed utility. In so doing, we employ the exceedance metric of equation 6.17 described in Section 6.6.1 as a measure of the goodness-of-fit between the predicted and assessed probability vectors used to generate the ROC curve.

7.2.2.1 PFA Goodness of Fit

We begin with the PFA vectors, and show how well the different models match the assessed PFA vector for a specific image and target combination. We do not depict the composite model for the PFA vector because it is an amalgam of the T-distribution and normal models for the PD vector only, as its PFA vector remains the T-distribution model. Figure 7.38 shows the PFA vectors for the FR I image for the four targets. Note how the normal and sum

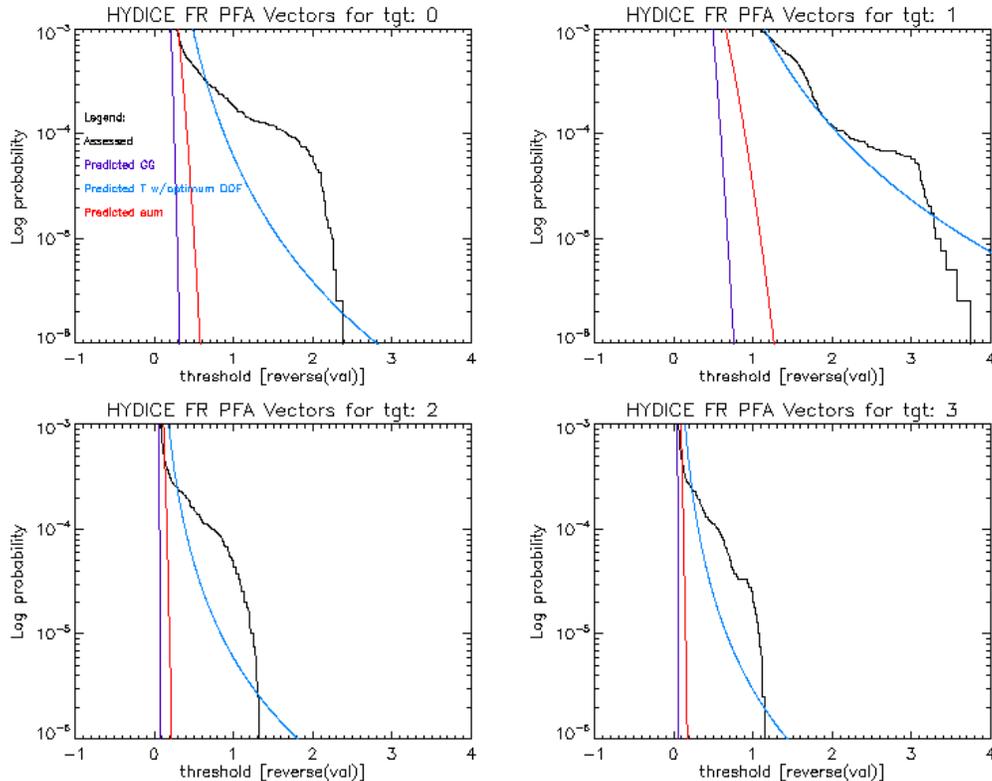


Figure 7.38: Assessed and predicted PFA vectors in FR I run 05 image for four targets

of Gaussian models cannot match the “heaviness” of the assessed PFA vector tail. The T-distribution, using the DOF parameter to minimize the exceedance metric with the assessed PFA vector clearly appears to be “heavy” at the low PFAs. This is key to being able to match the behavior of the PFA vector in the assessed case due to the early false alarms. Table 7.5 summarizes the exceedance metric for the four targets and three statistical models. It is clear that the T-distribution best captures the behavior of the assessed PFA vector over the PFA range of 5.6×10^{-6} to 5×10^{-4} , which we have taken to be our standard operating range. The number in parenthesis after the T-distribution model indicates the degree of freedom which creates the smallest (best fit) exceedance metric.

Image: FR I	Goodness-of-fit in the PFA range			
	Target 0	Target 1	Target 2	Target 3
Normal	1.51	2.15	0.83	0.69
T w/ best DOF ()	0.56 (4)	0.28 (4)	0.30 (3)	0.27 (3)
sum of Gaussian	1.33	1.79	0.73	0.61

Table 7.5: Exceedance Metric for Predicted PFA Vectors in FR I Run 05 Image for Four Targets

Figure 7.39 shows the PFA vectors for the HyMap image using the four targets. The T-distribution achieves a very close fit to the assessed PFA vectors. Note how the PFA

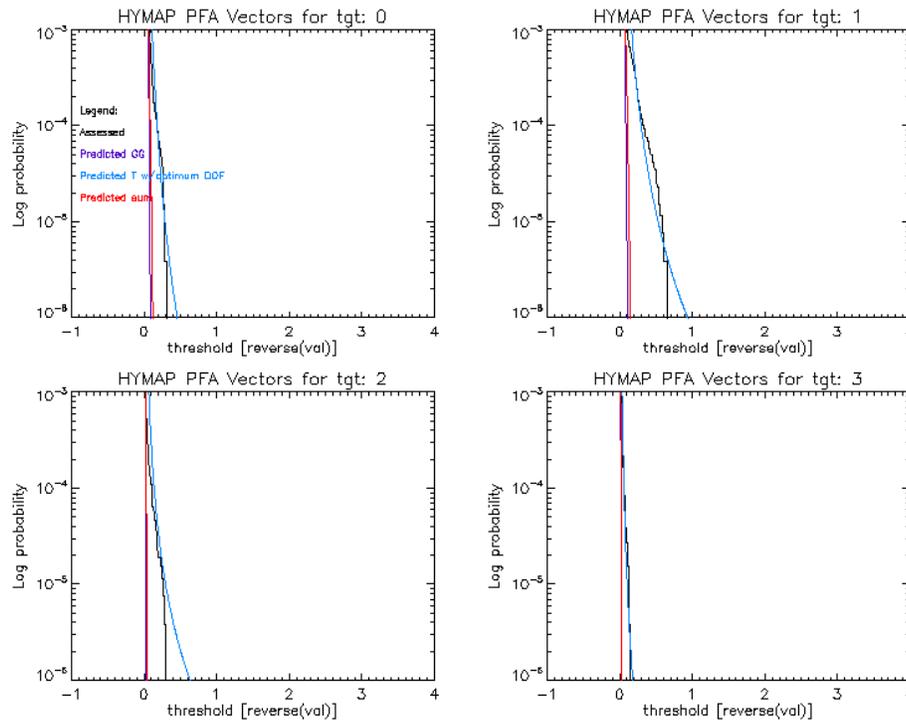


Figure 7.39: Assessed and predicted PFA vectors in HyMap image for four targets

vectors are steeper in this image for all targets than in the FR image. This is a result of the tighter, more Gaussian target absent distributions in this image, which create better separability between target present and absent distributions and thus easier detections.

Table 7.6 shows how much smaller the exceedance metrics for all of the prediction models are for the HyMap image than the FR I image. Clearly, the T-distribution with optimal DOF parameter achieves the smallest exceedance.

Image: HyMap	Goodness-of-fit in the PFA range			
	Target 0	Target 1	Target 2	Target 3
Normal	0.20	0.41	0.14	0.07
T w/ best DOF (ν)	0.07 (5)	0.19 (4)	0.03 (3)	0.01 (4)
sum of Gaussian	0.19	0.40	0.14	0.06

Table 7.6: Exceedance Metric for Predicted PFA Vectors in HyMap Image for Four Targets

The PFA vectors associated with the Quickbird image exhibit very heavy tails, as shown in Figure 7.40. The PFA vectors for target 1 extend beyond the upper range of the filter output values, so the exceedance metric for that target is not reliable. The assessed PFA

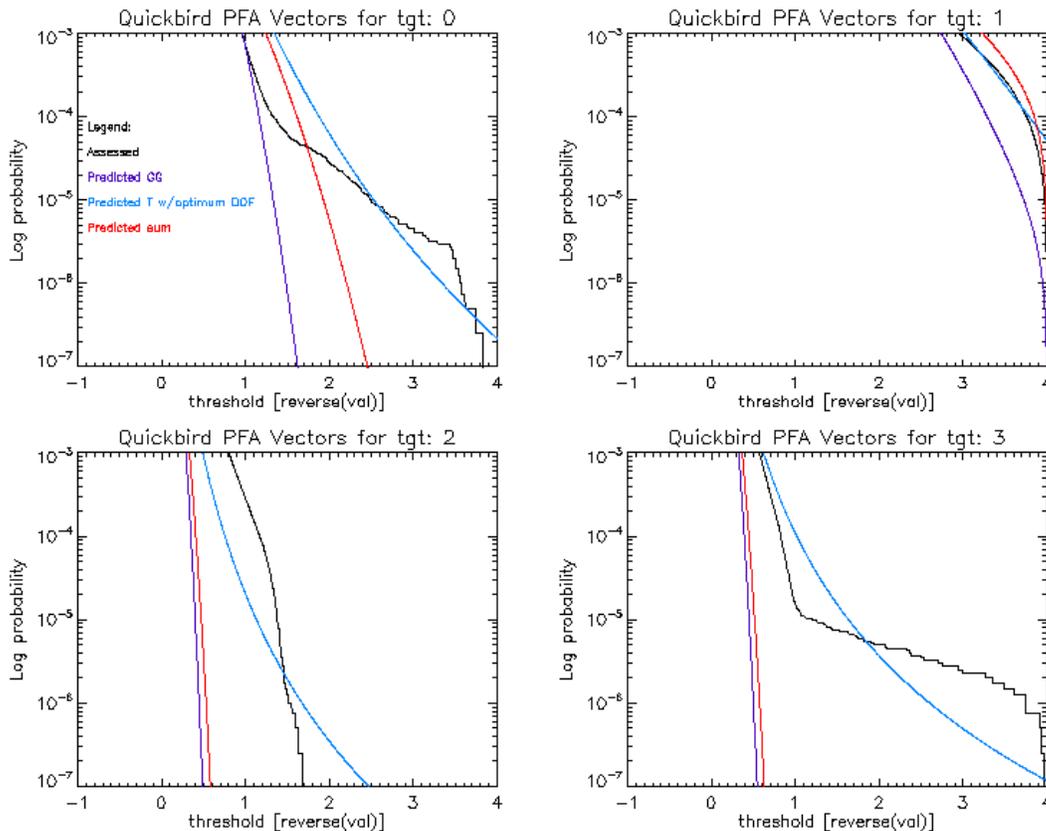


Figure 7.40: Assessed and predicted PFA vectors in Quickbird image for four targets

vectors for this image exhibit very different shapes than the previous two images. The exceedance metric shows this disparity in assessed and predicted PFA vectors in Table 7.7. The T-distribution with optimally-chosen DOF parameters achieves the best fit to the

Image: Quickbird	Goodness-of-fit in the PFA range			
	Target 0	Target 1	Target 2	Target 3
Normal	1.13	3.23	0.21	0.36
T w/ best DOF ()	1.06 (9)	3.11 (28)	0.19 (6)	0.34 (5)
sum of Gaussian	1.65	3.41	0.26	0.41

Table 7.7: Exceedance Metric for Predicted PFA Vectors in Quickbird Image for Four Targets

assessed PFA vectors in the range of interest. The ability to match the heavier tails of the assessed PFA is the most salient characteristic. The normal and sum of Gaussian models cannot achieve this effect.

7.2.2.2 PD Goodness of Fit

In the previous section, we evaluated the prediction models based on their ability to match the assessed PFA vectors over a designated PFA range. We extend that idea to the PD vectors associated with each implant fraction. Here, we choose a fixed PD range to calculate the mean absolute error in the exceedance metric from 0.1 to 1.0 in equal linearly spaced increments, since this is the range in which PD is of most interest. The PD vectors for each of the prediction models along with the assessed PD vectors are shown in Figure 7.41 for target

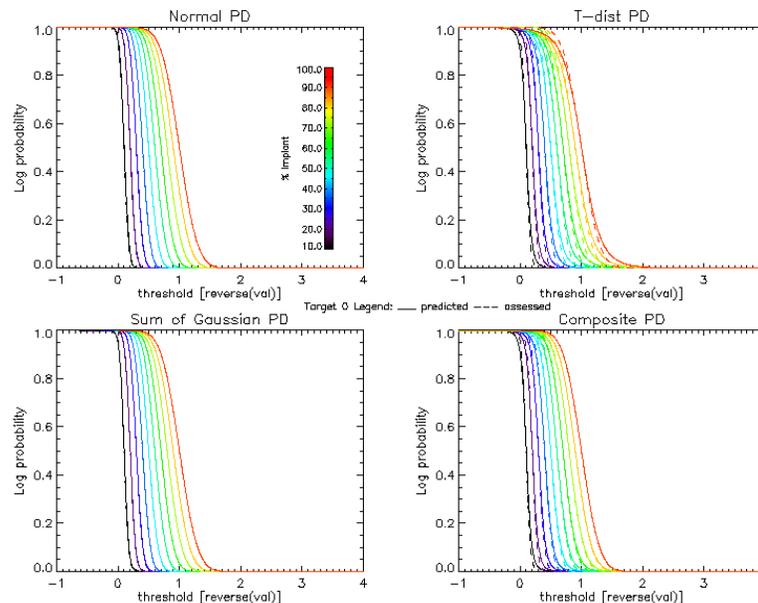


Figure 7.41: Assessed and predicted PD vectors in FR I Run 05 image for one target

0 (f8) in the FR I run 05 image. Again we see the characteristic feature of the T-distribution is the heavy tails. Note that as the implant fraction increases, the PD vectors look more Gaussian. This is because the target that is being mixed with the image statistics is from a normal distributional model.

When the exceedance metric at each target size (implant fraction) is plotted against the target size, the resulting exceedance error is shown in Figure 7.42. The trend here is that

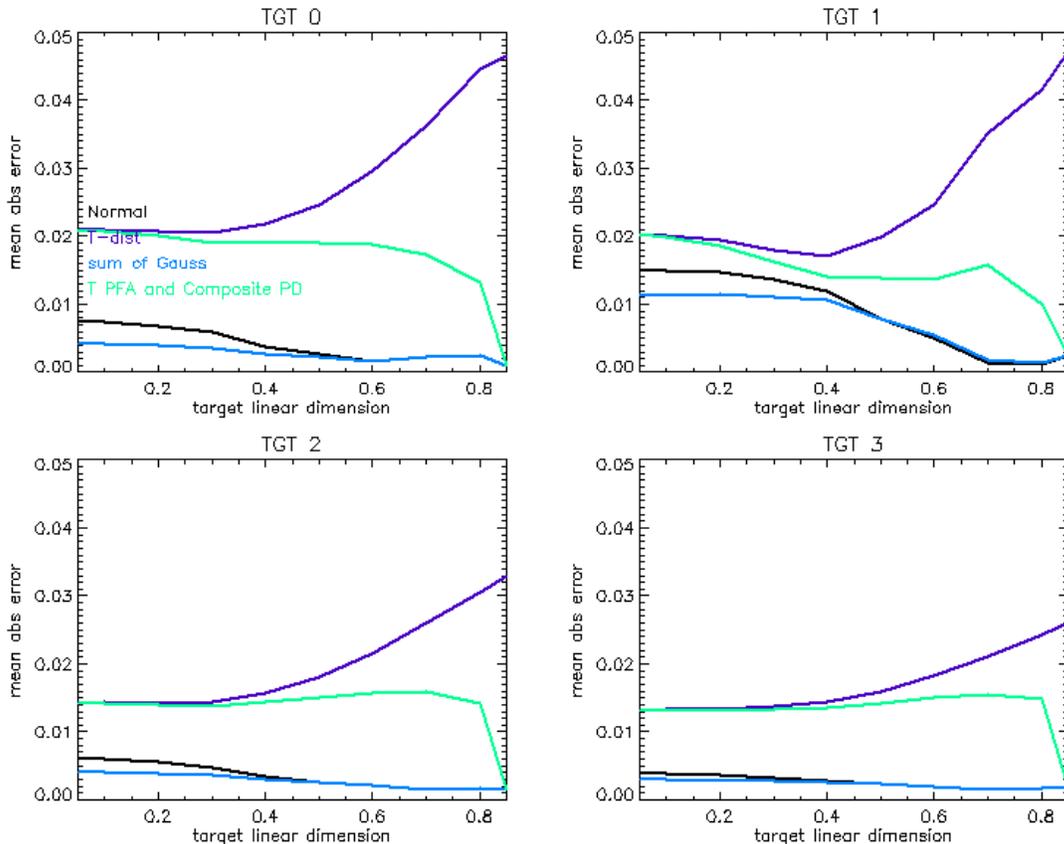


Figure 7.42: Predicted exceedance errors as a function of target size in FR I Run 05 image for four targets

the composite model shows a decreasing trend because it is incorporating more normal distribution PD as the fraction increases, and at the final fraction it is completely normal, as is the assessed PD vector. The T-distribution shows an increasing trend because it displays heavier tails with increasing fraction, while the assessed PD vectors are becoming more normal. Part of the poor performance of the T-distribution exceedance trend is due to the fact that the same T-distribution DOF is used for all fractional implants. So while, this may be the best fit for the PFA and very low fraction PDs, it ceases to be the best DOF as fraction

increases. The composite error mirrors the behavior of T distribution at very low fractions and that of the normal at high fractions.

Figure 7.43 shows the PD exceedance error plots for the HyMap image and Figure 7.44 shows the exceedance errors associated with the Quickbird image. The HyMap exceedance errors show similar trends as the FR I exceedance errors. The Quickbird error decreases for all prediction methods with increased fraction.

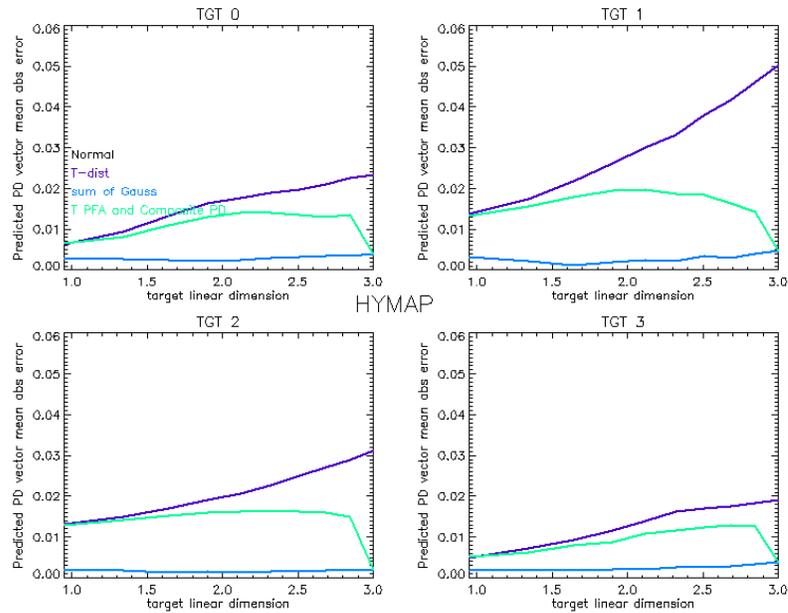


Figure 7.43: Predicted exceedance errors as a function of target size in HyMap image for four targets

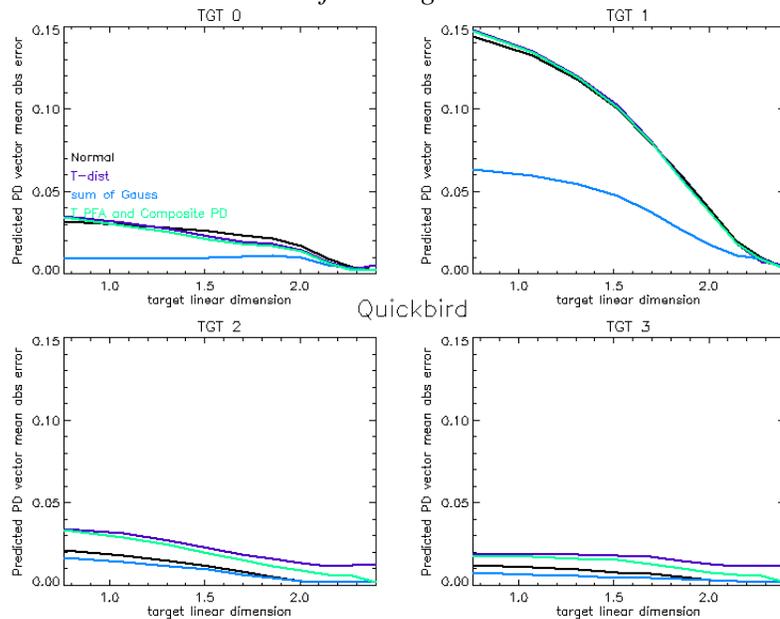


Figure 7.44: Predicted exceedance errors as a function of target size in Quickbird Image for four targets

7.2.2.3 Selection of Optimal T DOF using PFA Goodness-of-Fit

We select the optimum degree of freedom (DOF) for the T distribution by considering a range of DOF values and finding the DOF which achieves the smallest exceedance metric between the assessed PFA vectors and those generated for the range of DOF values. We investigated methods for estimation of the DOF parameter without needing to actually apply the filter to the image in order to get an assessed PFA vector, but none of the methods yielded DOFs that were optimal. The method of selecting the DOF that we use is not totally automated since it depends on the assessed utility; however, it offers a straightforward approach that can be applied to get the best possible prediction of the assessed utility. Figure 7.45 shows the PFA exceedance metric between the assessed and T-distribution PFA vectors plotted against the DOF used in the T-distribution. We use a small range of DOFs in this example because of empirical investigation into the typical optimal values. Each target has a

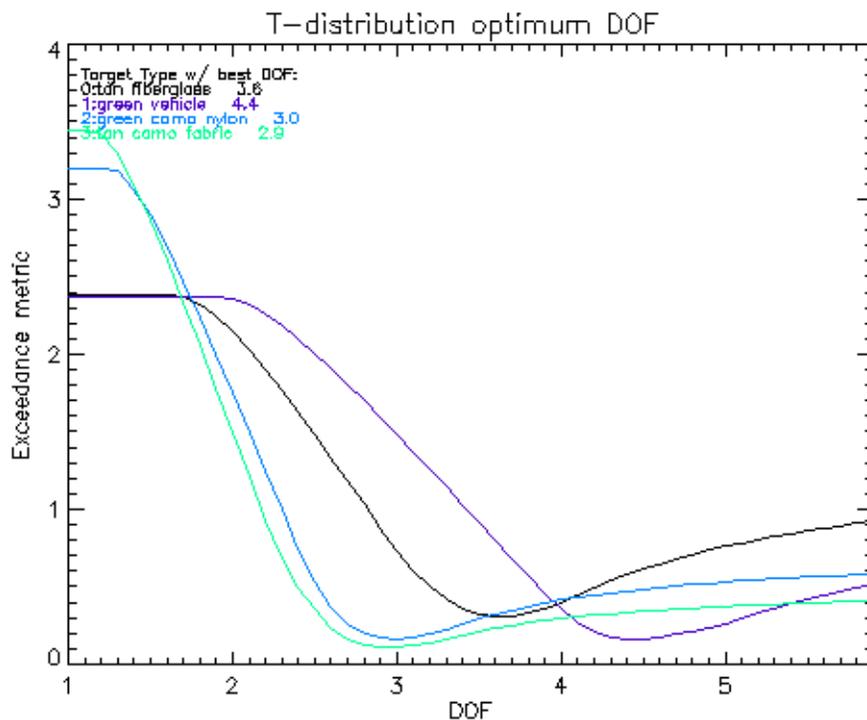


Figure 7.45: Optimal DOF selection for T-distribution PFA vector in FR I run 05 Image for four targets

unique minimum associated with it, though it seems that the “easier” targets have a lower optimum DOF, whereas “harder” targets have a slightly larger DOF. This is interesting, since a larger DOF makes the T-distribution behave in a more Gaussian manner.

In Figure 7.46, we show the PFA exceedance metric plotted against a wider range of

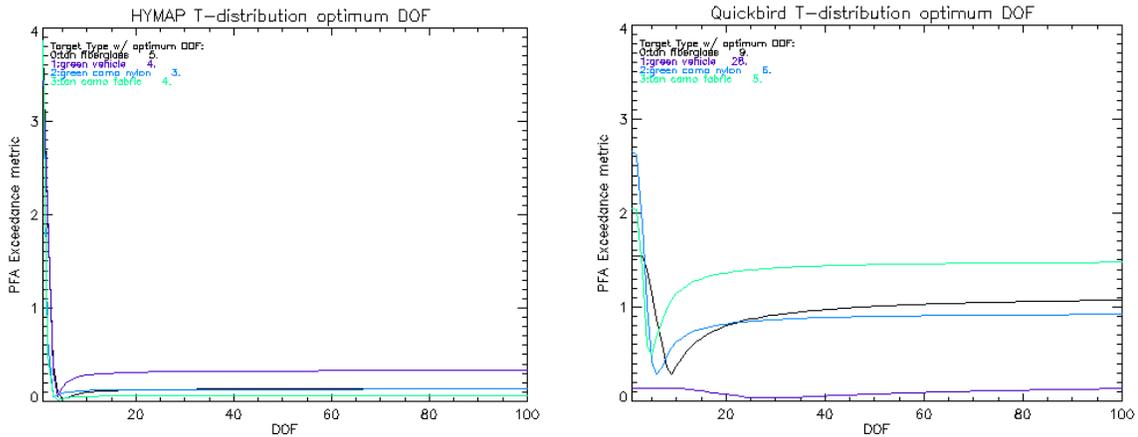


Figure 7.46: Optimal DOF selection for T-distribution PFA vector in HyMap and Quickbird images for four targets

DOFs for the HyMap and Quickbird images. Note that the vf124 (green vehicle) target on the right side of the figure has the largest DOF.

7.2.3 Characterizing the Accuracy of Predictions

In this section, we seek to show the absolute error between the predicted and assessed utilities as a function of target implant size, quantify the amount of error between each prediction method utility and the assessed utility, and explore the correlation between the PD goodness of fit and the utility.

7.2.3.1 Absolute Error Between Predicted and Assessed Utility

The absolute error between the image-derived predicted utility and the target-implant assessed utility is our primary indicator of the accuracy of the prediction method relative to the assessment. This metric is calculated at each target implant fraction. Rather than summarize this absolute error over all target implant fractions with a mean squared error, we present it as a plot of utility error versus target size (implant fraction) in order to show how the error behaves. Figures 7.47 – 7.49 show the absolute error for each of the three images. Each plots the utility error associated with the four utility prediction methods against target size for each of the four targets. In these plots, keep in mind that the utility error is zero when the assessed and predicted utilities take on the same value. Decreasing utility errors at the larger target sizes are generally due to the prediction having already attained a constant value of one while the assessed utility is still increasing. The HyMap utility errors all decrease

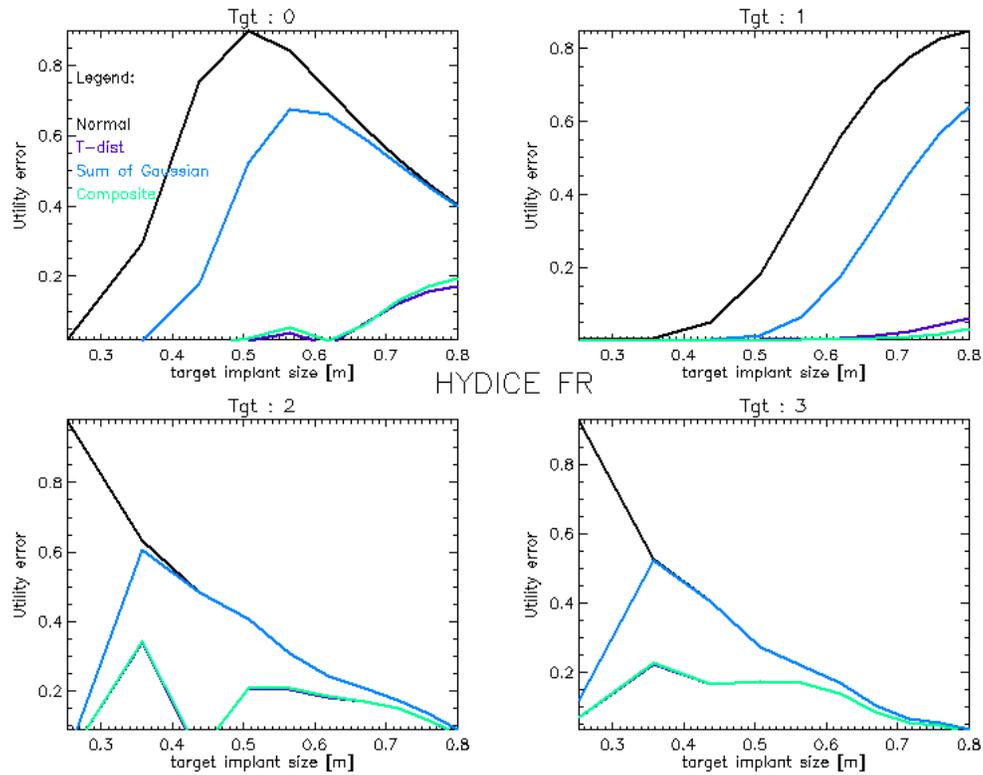


Figure 7.47: Utility error in FR I Run 05 image for four targets

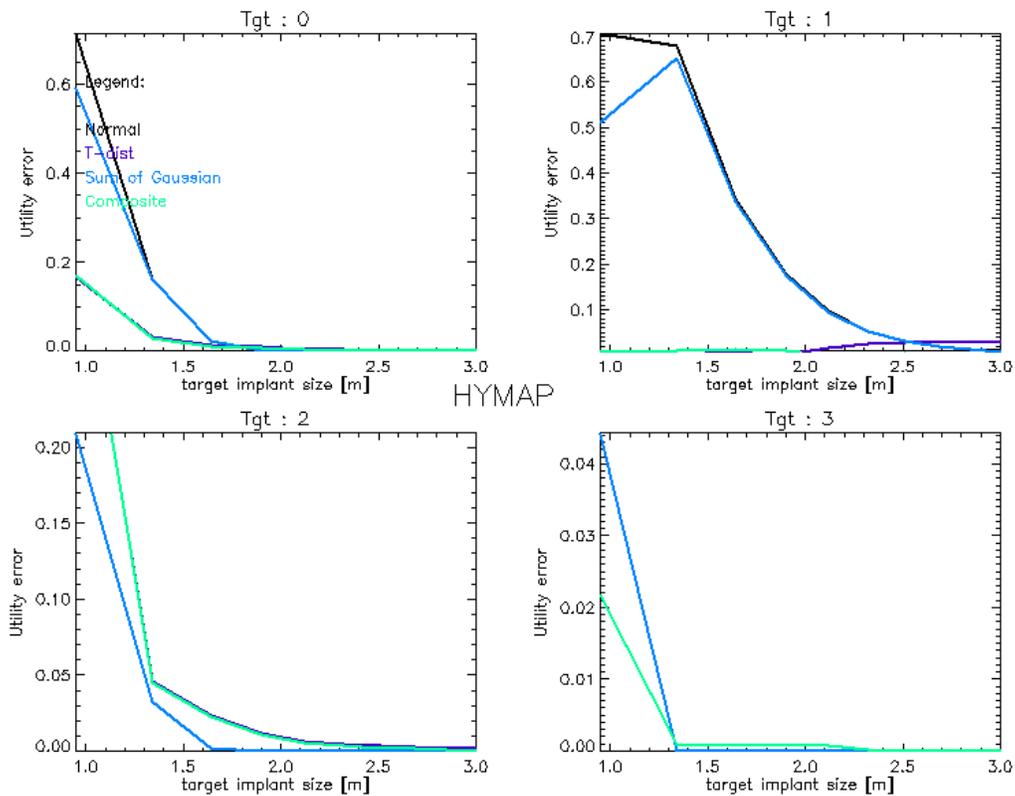


Figure 7.48: Utility error in HyMap image for four targets

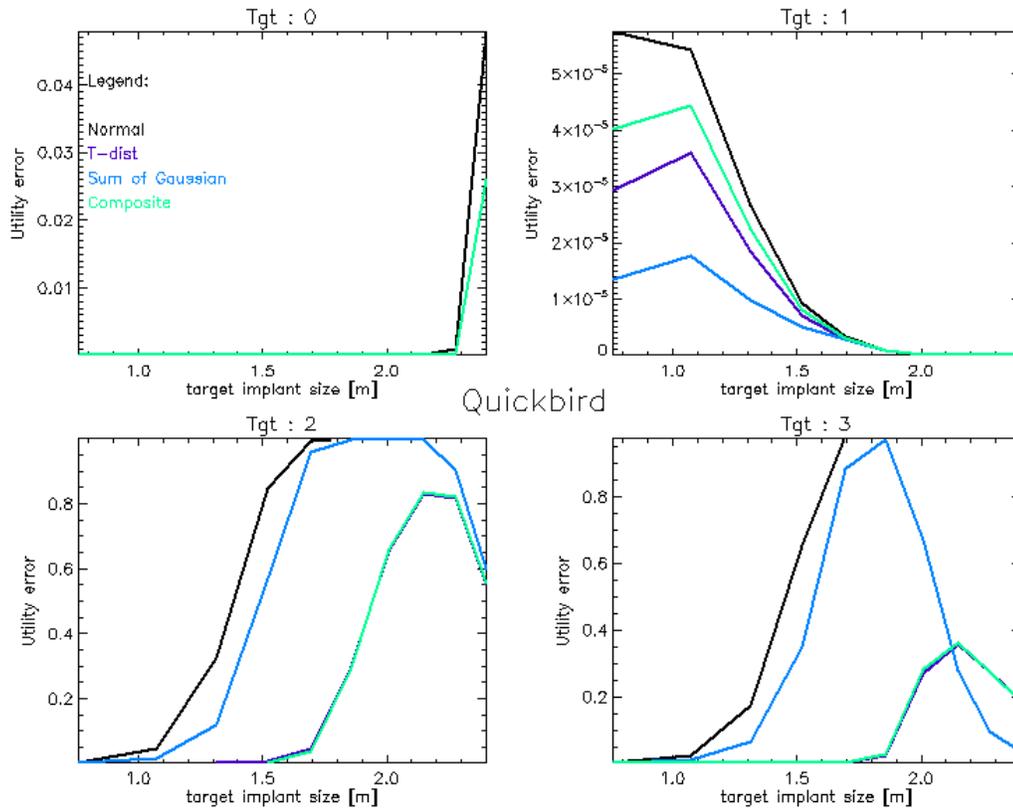


Figure 7.49: Utility error in Quickbird image for four targets

because of the very high utilities that are attained at relatively small target sizes by predicted and assessed methods. In general, it is a consistent trend that the normal model produces the greatest utility error followed by the sum of Gaussian model. The T-distribution and composite models both produce the smallest utility errors with a few small exceptions.

7.2.3.2 Metric for Quantifying Utility Error over 4 targets

In the previous section, we described the accuracy of the predictions using a plot of absolute utility error against target size for each of the four targets. In this section, we seek to summarize the accuracy of each method by characterizing how closely the predicted and assessed utilities come to matching the 1:1 line in a scatter plot of assessed and predicted utilities. This means of describing the accuracy came from an initial attempt to look for the degree of correlation between the two utilities. However, the correlation coefficient did not capture the accuracy of the predictions, or how closely the prediction values came to matching the assessed values. Figures 7.50 – 7.52 show the scatter plots of assessed and predicted utility for each of the three images. The assessed and predicted utility pairs have

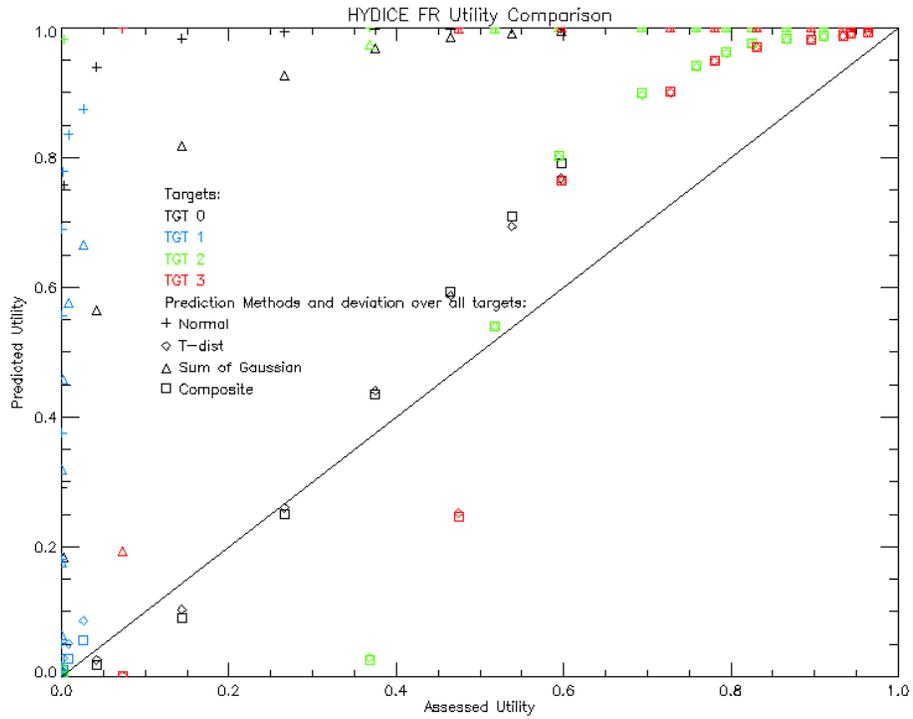


Figure 7.50: Utility scatter plot for FR I run 05 image for four targets and four prediction models

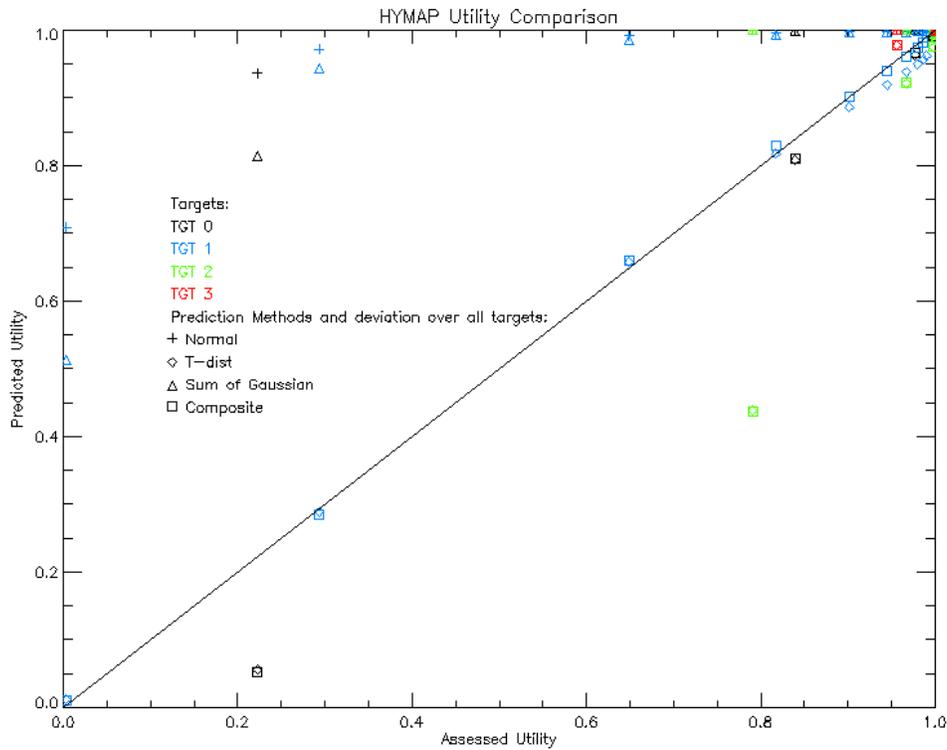


Figure 7.51: Utility scatter plot for HyMap image for four targets and four prediction models

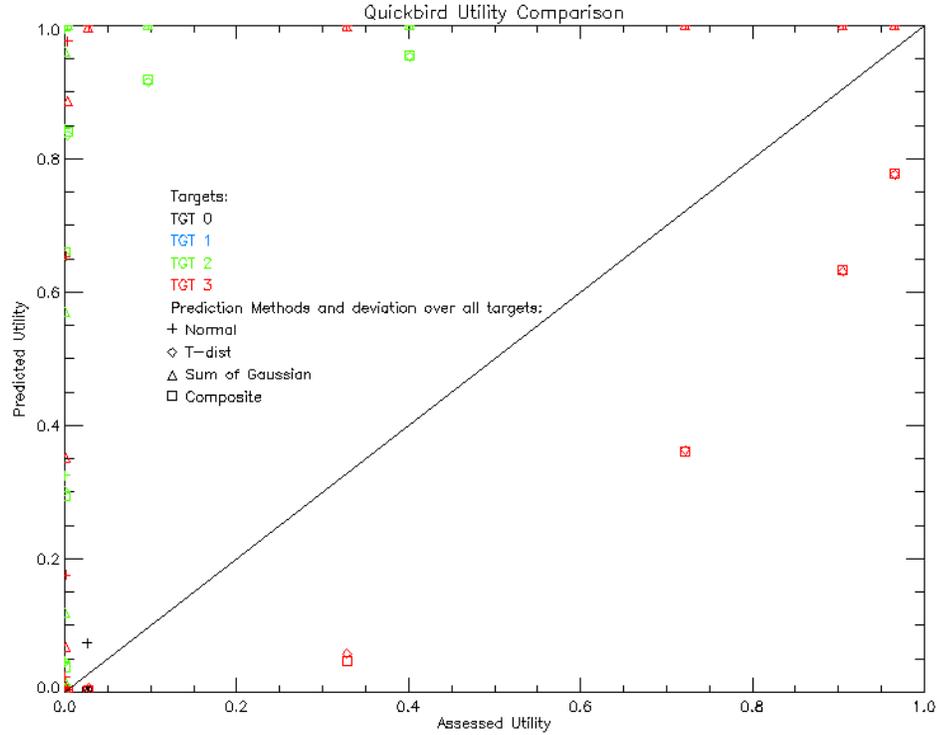


Figure 7.52: Utility scatter plot for Quickbird image for four targets and four prediction models

been color coded to correspond to each of the four targets and annotated with the appropriate symbol to show the prediction method. We see that the HyMap image has many utilities clustered in the upper right part of the plot because of the high predicted and assessed utilities for that image, whereas the Quickbird image has more clustering at the lower left part of the plot.

We seek to quantify the accuracy by summarizing the distance of all points for each method from the 1:1 line. This average distance is a measure of the agreement between the prediction and assessed utilities. Table 7.8 provides the average distance of the four targets and ten implant fractions for each prediction method from the assessed utility points.

Avg Distance	FR I run 05	HyMap	Quickbird
Normal	0.29	0.06	0.39
Sum of Gaussian	0.19	0.05	0.41
T-distribution	0.06	0.02	0.40
Composite	0.06	0.01	0.40

Table 7.8: Average Distance of Predicted from Assessed Utility for Four Prediction Methods

The summary metric shows the effect of many points concentrated at large values or small values, contributing to the very small values for HyMap. The results confirm the relative ordering of prediction accuracy that we saw in the earlier plots of utility error. The relatively large values for Quickbird reflect the large utility errors seen in Figure 7.49, in which targets 2 and 3 were difficult for every method to predict.

We investigated the relationship between the PD exceedance metrics of Section 7.2.2.2 and the utility error, but found no trends in correlation between these quantities. This is not too surprising, since the PD vector and the PFA vector are taken into consideration for the utility error, whereas the PD exceedance metric is only concerned with the PD vector.

7.2.4 Sensitivity of Predicted Utility to Prediction Parameters

In this section we examine the sensitivity of the utility prediction accuracy to variations in parameters required by the utility prediction models. We look at the degree of freedom parameter employed by the T-distribution model (and by default the composite model), the number of classes used by the sum of Gaussian classes, and the number of samples used to make the estimate of the image mean and covariance for the normal model. As an extension to this last line of inquiry, we look at the effect on the accuracy of the selection of the optimal DOF parameter and the predicted T utility when the number of samples used to estimate the image mean and covariance is reduced.

In this sensitivity analysis, we use a single target (f4) at a single implant fraction, 20%, with a specified PFA of 5×10^{-4} for calculation of the predicted utility. We vary the prediction parameter over a reasonable range (determined empirically) and record the utility error for each parameter setting. We then plot the utility error versus the range of the parameter investigated and calculate the slope about the midpoint of the parameter range and use this slope as a metric of sensitivity.

7.2.4.1 T-Distribution Degree of Freedom Parameter

We know that the DOF of the T-distribution is an important parameter in controlling the extent (heaviness) of the tail of the probability distribution. A large DOF will result in a more normal appearance. We showed results in Section 7.2.2.3 of a semi-automated method of selecting the DOF that achieved the minimal exceedance metric for the PFA vectors for each of our four targets and three images. Here, we want to see the impact of DOF choice on the

utility error between the T-distribution model predicted utility and the assessed utility. We would like to confirm that our method of selecting the DOF produces the lowest utility error. On the left side of Figure 7.53, we show the utility error plotted against the DOF parameter range for the FR I and the HyMap images. The degree of freedom parameter was allowed to cover a range of integers 1 to 10 in steps of 1. The predicted T-distribution utility error was

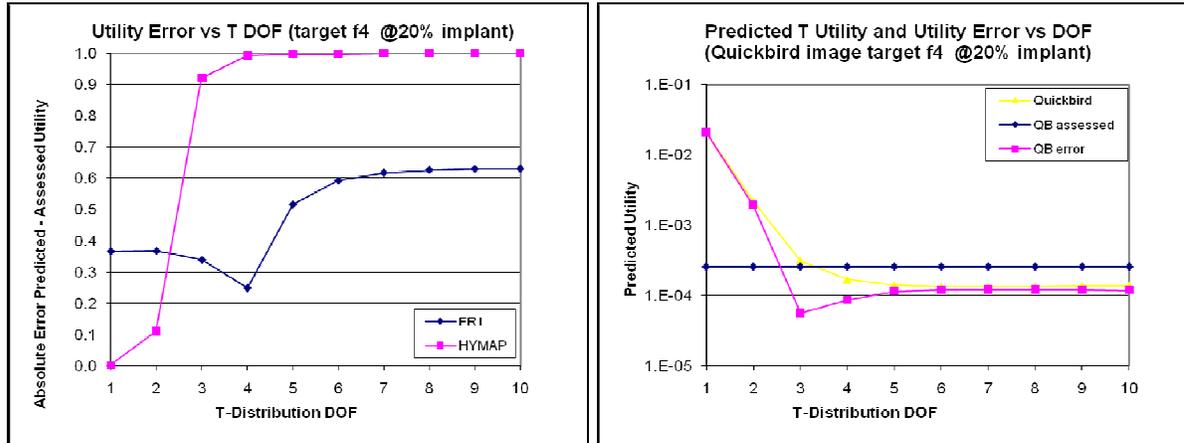


Figure 7.53: Utility error for T-distribution plotted against DOF parameter for FR I Run 05 and HyMap (left) and Quickbird (right) images.

plotted at each DOF. We see that the minimum utility error occurs at a DOF of 4 for FR I and 1 for HyMap. In Section 7.2.2.3, using the PFA vector exceedance metric as the cost function to be minimized, we obtained optimal DOFs of 3 for both the FR I and HyMap images for target f4. The reason for the disparity is that the utility error incorporates the effect of the 20% f4 target and the resulting PD vector that is then plotted against the PFA vector, whereas our estimation method only used the PFA vector. Thus, while our optimal DOF based on only using the target absent image to make the estimate is close, it is not the DOF that will produce the minimal utility error for all target implant fractions. For this target, the difference in DOF would produce a significant change in the utility error, particularly in the HyMap image, leading us to conclude that the DOF parameter plays a very important role in determining the accuracy of the utility prediction. We plot the Quickbird utility error on the right side of Figure 7.53 because the utility for this image/target/fraction combination is much smaller than for the HyMap or HYDICE. This plot shows the predicted T-distribution utility, assessed utility, and the absolute error of these utilities, with the utility axis on a log-scale. We see that the optimal DOF for minimizing the utility error is 3. Our estimate of the

optimal DOF based on the PFA vector exceedance metric was 5. Again, we see that the utility error is sensitive to the DOF parameter selection in this image.

7.2.4.2 Number of Classes used in Sum of Gaussian Model

This section seeks to uncover the sensitivity of the prediction accuracy to the number of classes used in the sum of Gaussian model. The idea behind the sum of Gaussian model is that a linear combination of spectral class statistics can better represent the complicated nature of a spectral image than the image-wide mean and covariance used in the normal model. The class statistics could be obtained in a variety of ways. Certainly a supervised classification would yield the best results, but ground truth is not always available, so we employ a simple unsupervised classification scheme to obtain a classification of the image. Specifically, we use k-means classification for this purpose, replicating the steps that an image analyst might employ. We estimated the number of classes to use as an input to the unsupervised classification. For each of the three images, we used 4, 8, 12, and 16 classes to start the k-means classification. Figure 7.54 shows the impact on the utility error between the sum of Gaussian prediction and the assessed utility for the f4 target at 20% fractional implant plotted against the number of classes. We see that the utility error is relatively insensitive to the number of classes used to form the model. Only the FR I image shows a slight decrease in utility error as the number of classes increases. The sum of Gaussian model describes a linear combination of filter outputs that could conceivably better match the non-Gaussian character of the assessed filter outputs. But while the mixture of normal class statistics may

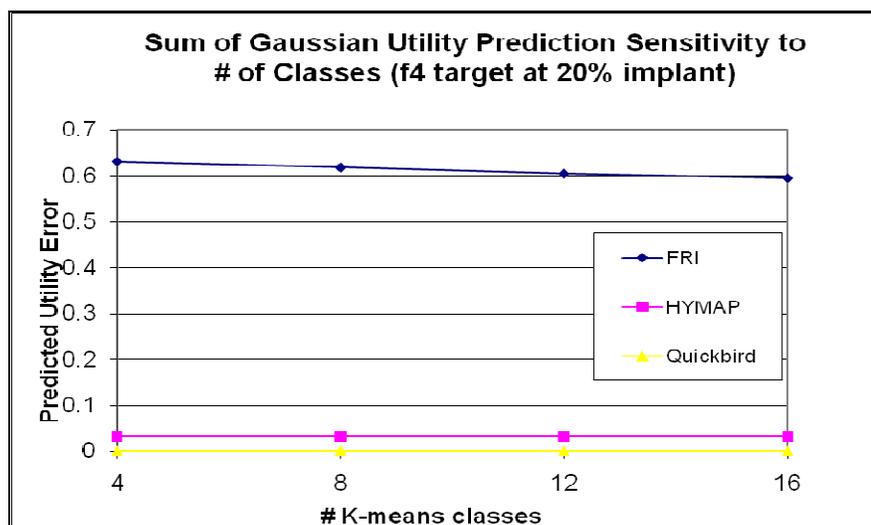


Figure 7.54: Utility error plotted against number of classes for FR I, HyMap, and Quickbird images

be a little better than using overall image statistics, it still does not address the key to our definition of utility, which lies in the tails of the filter output distributions. The normal (and by extension sum of Gaussian classes) will always do poorly in attempting to match the behavior of real spectral image filter outputs in the low PFA regions of interest.

7.2.4.3 Sample Support for Model Parameter Estimates

A key step in both assessing the utility with the target implant method and predicting the utility with the image-derived method is the estimation of the image covariance and mean. The assessment requires it because these parameters are needed to construct the spectral matched filter. If another detector such as SAM were used, then this requirement could be obviated. However, we have seen that SMF is a good detector and due to the fact that it is linear, allows us to directly compare the assessed and predicted utilities, since we cannot predict the utility for nonlinear filters using the image-derived method. The prediction methods all require calculation of statistical parameters not only to effect the transformation of multivariate pixel vectors to scalar filter outputs, but also to form the filter operator. A natural question is to examine the effect of using less than full sample support to form these statistical estimates. The primary motivation is to save time by using utility prediction instead of assessing image utility. We would like to specifically look at how utility accuracy is degraded as the sample support is decreases in the normal model, using the image mean and covariance estimates.

In Figure 7.55, we plot the utility error against the number of samples (as a percentage of the total number of image pixels) used to form the estimates of the image mean and covariance. The sampling of the image is done randomly using a uniform distribution. We see that the utility error associated with the normal model for the f4 target at 20% fractional implant is insensitive to the reduced sample support. We have used a lower limit of 0.1% of total image pixels as the smallest sample support in this plot, corresponding to 394 pixels for FR I, 262 pixels for HyMap, and 4,000 for Quickbird. The trend of no change in utility error continues until the sample support is on the order of the number of spectral channels, which for FR I is 0.0368 %, corresponding to 145 samples. After that point, the statistical estimate is not reliable and neither is the utility prediction. Although the smaller sample support is helpful in reducing processing time associated with forming the estimates,

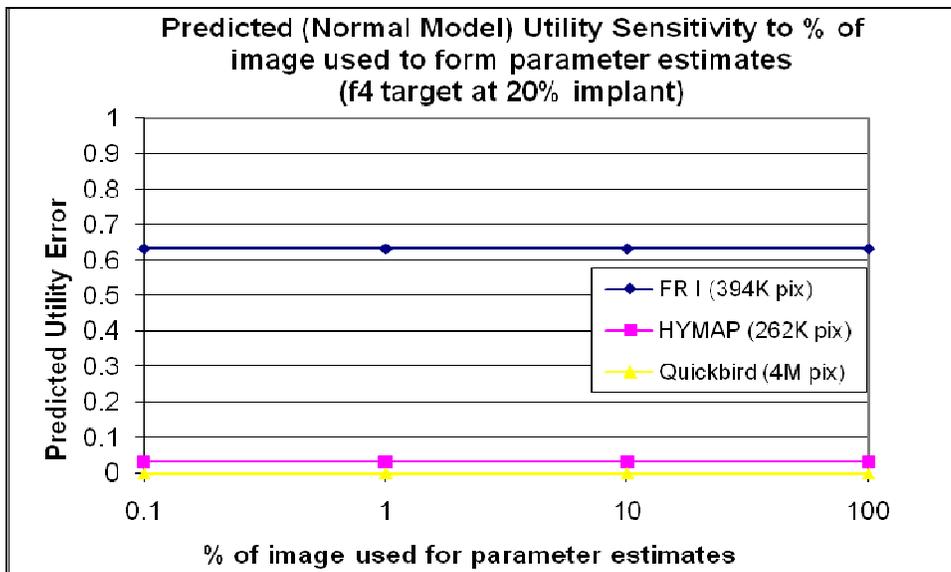


Figure 7.55: Utility error plotted as a function of sample support for normal model parameter estimates in three images

we should not draw conclusions based on Figure 7.55. The normal model always produces the largest utility errors, so degraded estimates of the image statistics do not impact a normal model for the filter output that already is known to not fit the tails of the distribution that we are interested in for characterizing the utility. In the next section, we explore the more meaningful question of the impact of reduced sample support on the T-distribution prediction model.

7.2.5 Prediction Accuracy and Time Tradeoff

This section deals with the competing goals of making the predictions faster and at the same time more accurate. In the previous section, we saw what happens to the normal distribution as the sample support is decreased. Here, we focus on the impact on accuracy of the better prediction methods (T-distribution and composite model) as we degrade the sample support used to estimate the global image mean and covariance and, by extension the optimal DOF parameter. We note the times required to complete the assessments and predictions in Table 7.9 along with the utility error for different sample support sizes for each of our three images. The prediction methods' most time consuming operations are the estimation of parameters and the selection of the optimal T-distribution DOF parameter. Without these requirements, the utility prediction is much faster than the assessment. But, since the prediction models all rely on estimated parameters, they need to perform these operations. Also, when estimating

the optimal T-distribution DOF, an assessed target absent distribution needs to be generated in order to estimate the optimal DOF that minimizes the exceedance metric for the PFA vectors. We see how much accuracy is sacrificed by cutting down sample support used to make the parameter estimate and DOF selection.

Note that the utility error actually decreases with smaller sample support for all images in both the composite and T-distribution cases. This is because the DOF estimated using the PFA vectors and image-wide mean and covariance are causing the predicted utility to increase as the sample support is decreased. This is true for all images. The effect of this is that the predicted utility comes closer to the assessed utility at 10% sample support and then greatly overestimates it for smaller sample support. The DOF increases so that the prediction is effectively like the normal distribution at this point. This effect is partially due to the fact that we are only looking at one target implant fraction and the absolute error metric employed. Although the utility error is smaller when forming the parameter estimates with 10% of the image pixels, we should not conclude that this will always be the case.

FR I (394,20 pixels and 145 channels)					
% pixels	T-distribution error	composite error	estimated DOF	Time to assess [sec]	time to predict [sec]
0.1	0.63	0.63	100	30.1	12.8
1	0.63	0.63	22	31.1	13.2
10	0.25	0.26	3	32	15.3
100	0.34	0.34	3	48	33.9
HYMAP (262,144 pixels and 96 channels)					
% pixels	T-distribution error	composite error	estimated DOF	Time to assess [sec]	time to predict [sec]
0.1	0.03	0.03	100	13.8	10.7
1	0.03	0.03	9	14.3	10.9
10	0.03	0.03	4	19.5	13.3
100	0.05	0.05	3	30.2	20.9
Quickbird (4,000,000 pixels and 4 channels)					
% pixels	T-distribution error	composite error	estimated DOF	Time to assess [sec]	time to predict [sec]
0.1	2.3E-05	3.8E-05	100	20.3	11.9
1	2.3E-05	3.9E-05	90	20.9	12.3
10	7.3E-05	1.0E-04	13	35	13.6
100	8.3E-05	1.1E-04	9	72.5	39.7

Table 7.9: Utility Errors and Computation Times With Decreased Sample Support for Three Images and 20% f4 Target

The other feature to note in Table 7.9 is that the times (in seconds) required for processing decrease up to a certain point for the both predicted and assessed utility operations and stabilize. The reason for this is that the assessed utility operation consumes most of the processing time in the application of the filter to the image pixels twice (once for target absent and once for target present). Likewise, the predicted utility has operations that are much longer in duration than the estimation of statistical parameters. The predicted times require the calculation of the PFA vector, which requires the running of the target absent image. As sample support is reduced, the time required to estimate the covariance and mean become less significant in both methods. Clearly, predicting the T-distribution is faster than assessing the utility, cutting 30 – 40% off of the time required to perform all operations.

Figure 7.56 illustrates the tradeoff between prediction accuracy and prediction time for the Forest Radiance image. On the left axis, the prediction error is plotted as a function of sample support (in percent of total image pixels used to estimate model parameters) for the

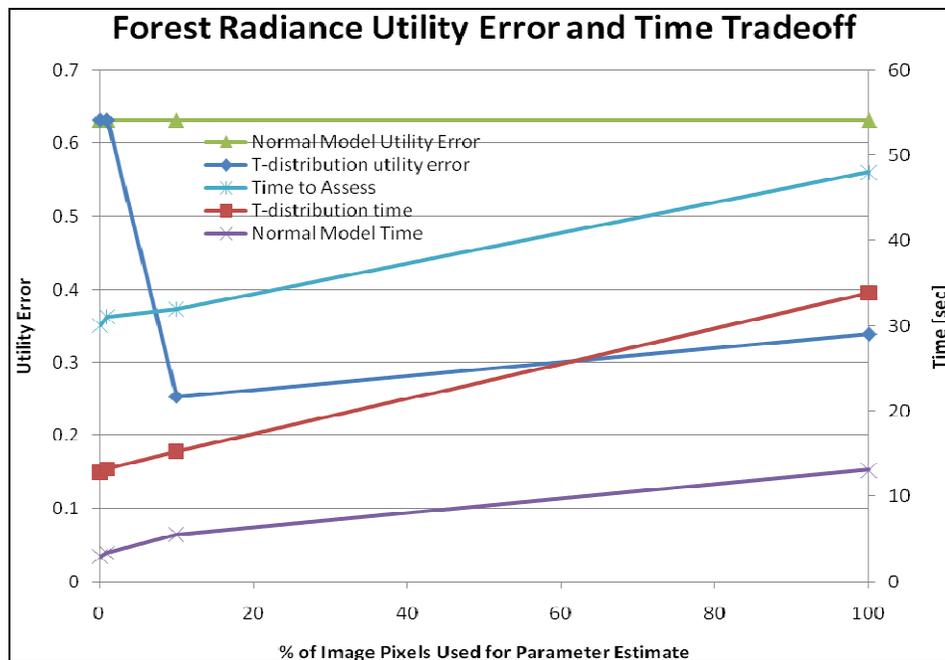


Figure 7.56: Tradeoff between prediction accuracy and time

normal and T-distribution models. The normal model clearly has the larger error and shows insensitivity to the sample support. The T-distribution prediction error displays a generally increasing trend as the sample support is reduced, but is not linear, for the reasons discussed above. The time to make the prediction is plotted on the right vertical axis, and we see that the time to assess is the largest, followed by the time to predict using the T-distribution

distribution and normal models. This demonstrates the fact that better prediction accuracy, as expressed using the utility error measure, is only attainable with the T-distribution model that requires more time to create better parameter estimates than the normal model.

7.2.6 Comparison with Other Spectral Image Quality Methods

The objective of this section is to compare other spectral image quality approaches found in the literature to our target implant utility assessment methodology and image-derived utility prediction methods. We do not presuppose that any one particular method is “right.” The only frame of reference that we have is the assessed utility, but this is the closest that we have to the ground truth for the particular situation, since it results from actually conducting target detection on an image. As such, we do not try to overstate any of the results in this section in terms of absolute comparisons. Rather, we hope that these comparisons will show trends in the applicability of various attempts to predict the utility of images.

The methodology we follow in the next subsections is to apply each method to ten real spectral images and note the value that each method returns as well as the relative ranking of images. We also note any subtleties associated with implementing each method and how the specific variables required by each were obtained. We organize the spectral image quality metrics into the three groups introduced in Section 6.6.3 as those that are target specific, those that are target type and size specific, and those that are independent of the target. We compare our assessed and predicted utilities for the specific image and target detection scenario with the spectral image quality methods in each category. We only show the results of our best prediction method, the composite model, to simplify the discussion. The baseline target detection scenario for the assessed and predicted utilities is a full pixel vf124 target at a specified PFA of 1×10^{-3} . The only exception is when comparing to methods that are target size-specific, in which case we fix the size of the target at 0.8 m for all images to capture the effect of increasing GRD.

7.2.6.1 Target-type specific methods

This category includes the spectral similarity vector (SSV) and the spectral quality rating scale modified to incorporate the signal-to-clutter ratio (SQRS_SCR). We choose to work with target vf124 (green vehicle) at a 100% target implant fraction, since we cannot specify target size in the SSV and SQRS_SCR approaches. This allows equitable comparison

with our image-derived utility prediction methods, in which target implant fraction can be specified. The SSV in its simplest application requires the estimation of the image mean vector with which to compare the target mean vector. The threshold for the SQRS_SCR metric is selected empirically from the target absent filter output using the same specified PFA as our utility assessment (1×10^{-3}). The bar chart in Figure 7.57 shows the SSV, SQRS_SCR, assessed utility, and the four image-derived utility prediction methods using a specified PFA of 1×10^{-3} for the assessed and predicted utility methods. It appears that SSV consistently returns a higher value than SQRS_SCR. The assessed and predicted utilities do

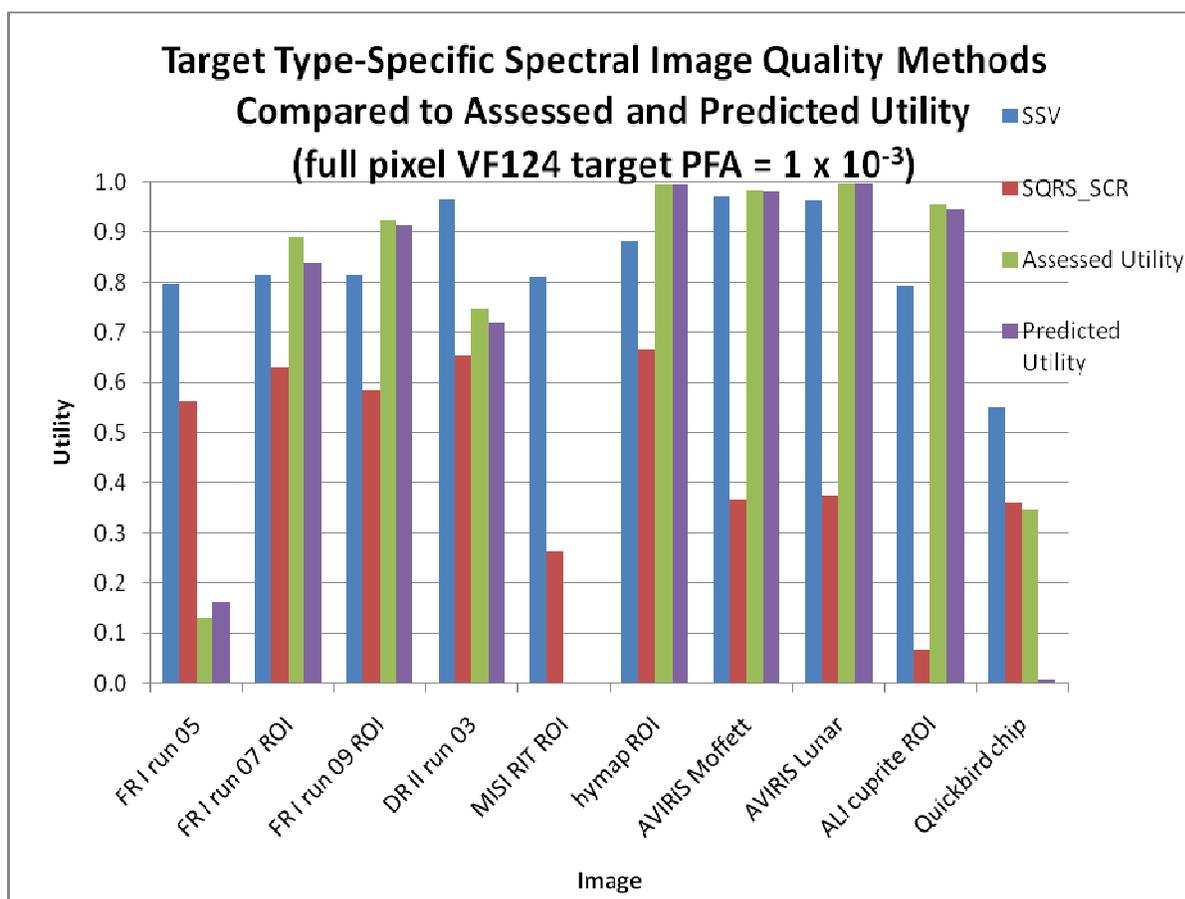


Figure 7.57: Comparison of target-specific spectral image quality methods for ten images at specified PFA 1×10^{-3}

not appear to be very correlated with the SSV and SQRS_SCR values. The assessed and predicted utilities return very low values for the MISI image, as this is a difficult target.

In order to investigate correlations between the methods, we look at the scatter plot between the assessed utility and the other prediction methods in Figure 7.58. If the linear correlation coefficient is used, then the predicted composite utility has the highest correlation

with the assessed utility for target vf124 using a specified PFA of 5×10^{-4} over the ten images considered. The SSV shows a surprisingly high correlation coefficient, while the SQRS_SCR shows a low correlation. We use the correlation coefficient as a measure since the two axes are produced by different methods and cannot be equitably compared as we did with the assessed and predicted utilities in Section 7.2.3.2 with the distance from the 1:1 line. We expected SQRS_SCR to have a better correlation since it is based on the specified PFA threshold value used in the assessed utility, and the SCR is in many ways similar to the SMF used in assessed utility.

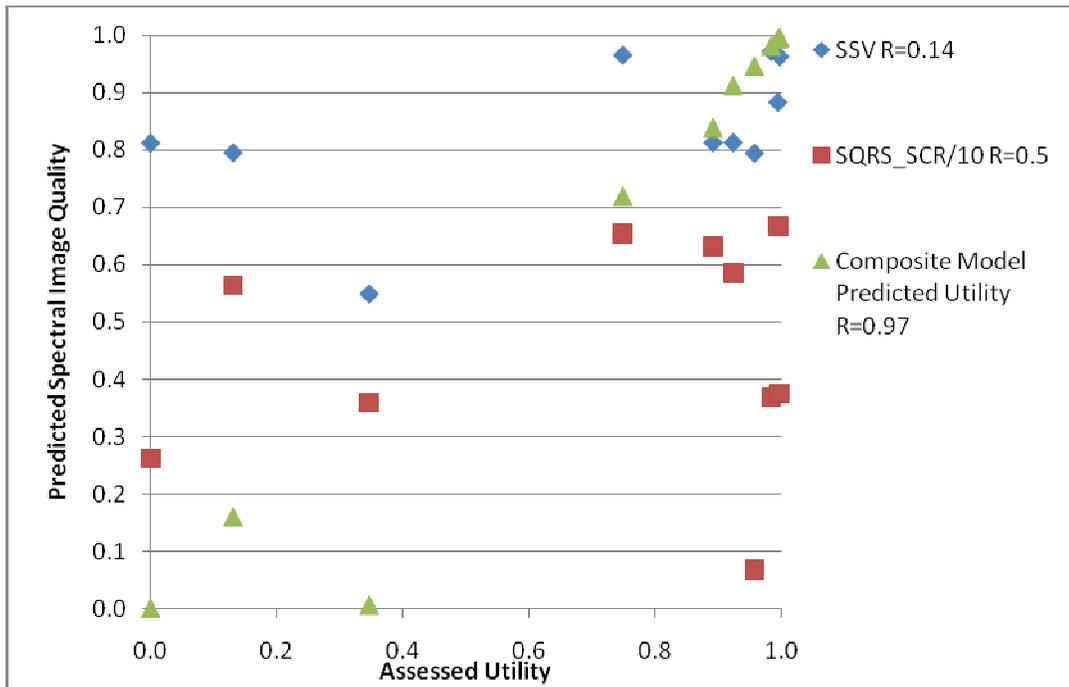


Figure 7.58: Scatter plot of assessed utility versus predicted utility for target-specific spectral quality methods in ten images

7.2.6.2 Target type/size independent methods

The spectral quality rating scale (SQRS) and spectral quality equation (SQE) methods are examined in this section. Fundamentally, these approaches are not very similar in derivation to our utility metrics, so we do not expect much correlation between them. Implementing the SQRS was straightforward, using standard sensor and image parameters. The implementation of the SQE included the same SNR as SQRS but derived the scene standard deviation averaged over all spectral channels. The units of standard deviation are calibrated HYDICE radiance units [$4/3 \mu\text{W}/\text{cm}^2\text{-sr-}\mu\text{m}$]. Table 7.10 shows the parameters used to derive the SQE

and SQRS values. Calibrated radiance units were not available for the MISI image. Figure 7.59 shows the SQRS and SQE methods compared with the assessed and predicted utilities calculated for a full pixel of 124 target at a specified PFA of 1×10^{-3} . This PFA corresponds to the PFA used in Shen (2003), and although SQE and SQRS do not allow specification of a

Reflectance Image	GSD [m]	notional sensor SNR (from specs)	average spectral resolution [nm]	scene standard deviation [HYDICE]
FR I run 05	0.8	100	10	179.1
FR I run 07 ROI	1.5	300	10	181.2
FR I run 09 ROI	3.0	300	10	170.6
DR II run 03	0.8	100	10	198.1
MISI RIT ROI	3.0	30	10	-
hymap ROI	3.0	800	15	246.3
AVIRIS Moffett	20.0	1000	10	146.3
AVIRIS Lunar	20.0	1000	10	142.0
ALI cuprite ROI	30.0	500	89	75.6
Quickbird chip	2.4	100	100	233.4

Table 7.10: Parameters Used in Calculation of SQRS and SQE Spectral Image Quality Metrics

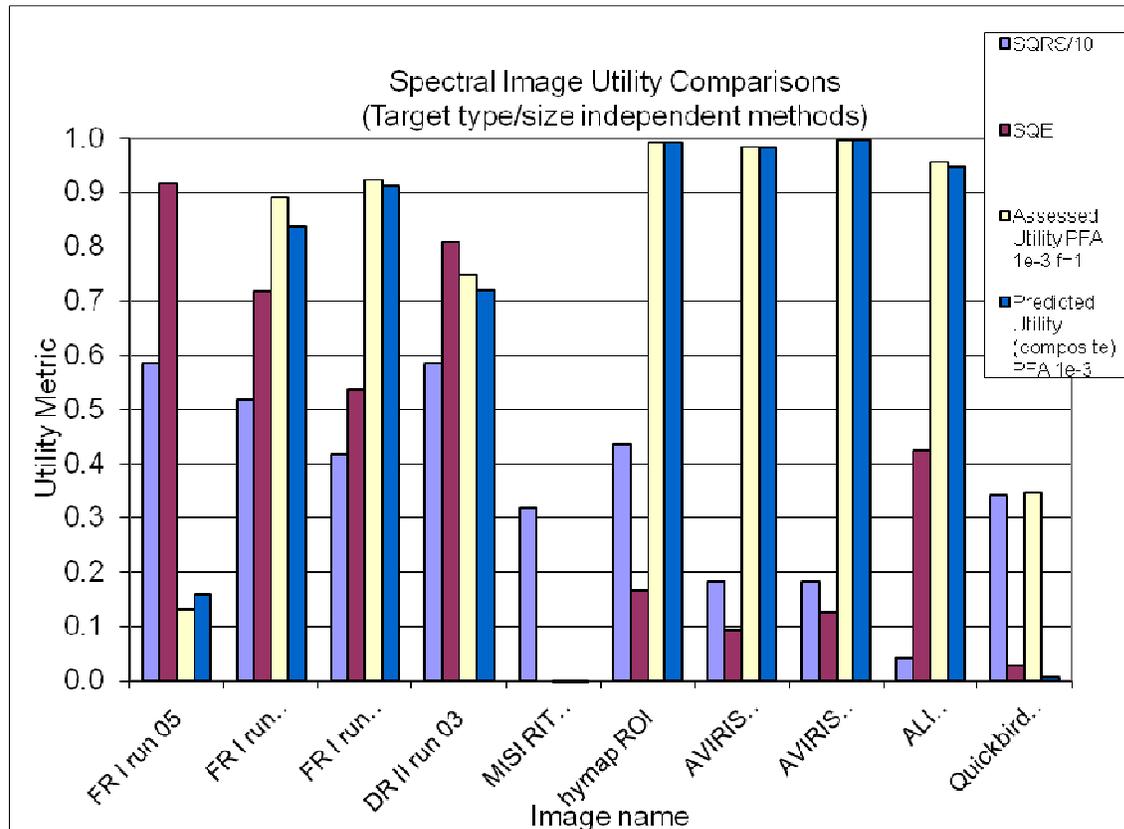


Figure 7.59: Comparison of target type/size independent methods with assessed and predicted utility for ten images

particular target type, we use the same vf124 target as in the comparisons with the other spectral image quality methods to maintain consistency. The results of this comparison confirm that there is no significant correlation apparent between these methods. We observe that with the exception of the FR I and MISI images, the SQE and SQRS metrics appear to be consistently lower than the assessed or predicted utilities.

7.2.6.3 Target type/size-specific techniques

Here, we investigate the general spectral utility metric (GSUM) spectral image quality method relative to our utility methods. The spatial confidence for the GSUM is calculated by assuming a 0.8 m target and translating this fixed target size into the number of cycles across the target depending on the sensor GRD for use in the Johnson criteria (equation 3.24). For spectral confidence, we use the PD at the specified PFA associated with the target vf124 obtained from the ROC curve generated by the target implant utility assessment method for a full pixel target. The rationale for this is that the spectral confidence should be based on spectral information that is independent of the size of the target. In calculating our assessed and predicted utilities, we consider a 0.8 m target, since this is the smallest GRD image we consider. Figure 7.60 shows the results of our comparison with GSUM for this case. We see that we have a very difficult detection situation for the larger GRD sensors, such as AVIRIS

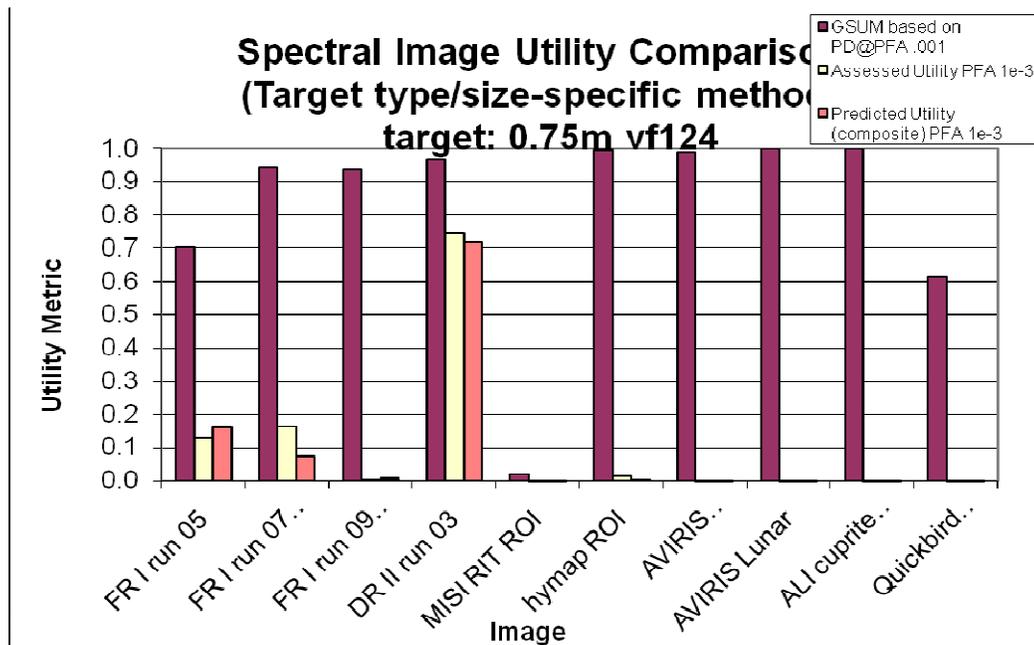


Figure 7.60: Comparison of GSUM with assessed and predicted utility methods for ten images

and ALI. Our utility is very low in these situations because of the very small target size. GSUM does not capture this difficult situation because it puts weight on the spectral confidence obtained using a full pixel target in deriving the metric. GSUM consistently overestimates the utility relative to the assessed utility. The spectral confidence for the GSUM is obtained by a spectral confidence associated with finding a full pixel target, thus, it will be high. GSUM combines spatial and spectral information deliberately using the semantic transform. In the assessed and predicted utility method, the spatial information is introduced by the mixing fraction that target size translates to for a particular image. The spectral information comes in based on the particular target being sought.

7.2.6.4 Observations

The comparison of the various spectral image quality methods with our spectral utility metric is instructive. We see the great flexibility that the target-implant approach affords in defining the specific detection situation. We can define the target type, size, and specified PFA. By estimating the image covariance to form the detection operator, we pull in information about the sensor parameters and the scene composition. Thus, without developing an equation, but rather approaching the problem in this empirical manner, we have a utility metric that can tell us quite a bit about an image relative to the specifics that we asked of it. The challenge is to consider how to really compare this method fairly to the other spectral image quality methods. The ideal for a utility metric would be broad applicability. We see that methods of predicting based on a small sample set will only provide limited applicability. This is not by any means implying that our method is better or more applicable than the other spectral image quality methods. It approaches the problem in a different manner, and it is only applicable to the narrow target detection task around which it was designed.

We compare the methods by their rank ordering of the images from highest to lowest in Table 7.11. We only present the composite prediction from our image-derived method, since this method typically provides the closest match to our assessed utility. AVIRIS and HyMap are ranked highly by all of the methods with the exception of SQRS and SQE. Quickbird is ranked consistently low by all methods. The rank ordering of the assessed and predicted utilities is very close to that assigned by GSUM. Note that we show the ranking of assessed and predicted utility for full pixel targets as well as 0.75 m targets. The difference in

Rank	Assessed Utility (full pixel target)	Predicted utility (full pixel target)	Assessed and Predicted (0.75m target)	SSV	SQE	SQRS_SCR	SQRS	GSUM
1	AVIRIS L.	AVIRIS L.	DR II	AVIRIS M.	FR run 05	HyMap	DR II	HyMap
2	HyMap	HyMap	FR run 07	DR II	DR II	DR II	FR run 05	AVIRIS L.
3	AVIRIS M.	AVIRIS M.	FR run 05	AVIRIS L.	FR run 07	FR run 07	FR run 07	ALI
4	ALI	ALI	HyMap	HyMap	FR run 09	FR run 09	HyMap	AVIRIS M.
5	FR run 09	FR run 09	FR run 09	FR run 09	ALI	FR run 05	FR run 09	DR II
6	FR run 07	FR run 07	AVIRIS L.	FR run 07	HyMap	AVIRIS L.	Quickbird	FR run 09
7	DR II	DR II	AVIRIS M.	MISI	AVIRIS L.	AVIRIS M.	MISI	FR run 07
8	Quickbird	FR run 05	ALI	FR run 05	AVIRIS M.	Quickbird	AVIRIS M.	FR run 05
9	FR run 05	Quickbird	Quickbird	ALI	Quickbird	MISI	AVIRIS L.	Quickbird
10	MISI	MISI	MISI	Quickbird	-	ALI	ALI	MISI

Table 7.11: Rank Ordering of Image Utilities by Spectral Quality Methods

these rankings is that the larger GRD of AVIRIS and ALI is taken into account with the fixed target size, resulting in lower utility. The range of utility scores assigned for the ten images is from 0 to 1 for the assessed and predicted utilities, 0.06 to 0.79 for SSV, 0.03 to 0.92 for SQE, 0.07 to 0.67 for SQRS_SCR, 0.04 to 0.59 for SQRS, and 0.02 to 1 for GSUM. It is difficult to make direct comparisons between the methods. Perhaps the most valuable insight gained in this comparison is to understand the flexibility afforded to the image analyst by our image-derived prediction methods. These may be fitted to the specific image and target detection situation of interest. None of the other techniques provide such flexibility in application.

7.2.7 Summary of Utility Prediction Results

In this chapter, we have described in detail how the image-derived utility prediction methods operate. We have seen that the accuracy of the utility prediction may be directly compared to the assessed utility since both methodologies follow the same binary hypothesis test methodology. The prediction is only possible for linear detectors at this point, so all of our comparisons were made using the spectral matched filter. The accuracy of prediction is predicated on the behavior of the PFA vector and PD vector that together form the ROC curve. We saw that the statistical model used to describe the filter output distribution was crucial to the accuracy of the prediction, with the normal and sum of Gaussian models typically overestimating performance because they could not accurately account for the light

tail of the PFA vector, which is caused by the first few false alarms. The T-distribution model accommodates this reality of spectral imagery, but only up to a certain point. By combining the T-distribution and normal distributions, we saw that the composite model accounted for the non-Gaussian nature of the image while simultaneously accounting for the target present situation, in which the implanted targets created a more normal image. We demonstrated our method for measuring the goodness of a prediction based on the exceedance metric in the PFA vector and applied this to the task of semi-automated estimation of the optimal DOF for the T-distribution model. We explored the sensitivity of the utility predictions to the DOF, number of spectral classes, and number of samples used in forming the prediction models. We saw that the utility error between predictions and assessed utilities was most sensitive to the T-distribution DOF parameter. The time needed to accomplish prediction as opposed to target-implant assessment is 30-40% less when using full image statistics, and could be even larger if using reduced sample support for parameter estimation.

We performed a comparison of our spectral image utility methods with other spectral image quality metrics. In general, no significant commonalities were discovered, although none were expected due to the very different nature of the approaches to utility.

7.3 Sensitivity Analysis

This section looks at each of the major image chain components and the parameters in them, endeavoring to understand the sensitivity of utility to them. The behavior of utility with respect to the parameter variation is explored in each of the subsequent subsections (scene, image acquisition, preprocessing, and target detection scenario), along with a subsection devoted to quantifying the sensitivity in a summary sense. We will only examine the assessed utility since it would be too confusing to also include predicted utility and we have already discussed the sensitivity of the predictions to their own prediction parameters. In order to make this analysis tractable, we consider a single target and implant fraction combination as we vary the parameter of interest. Thus, in most cases, we will use targets c6 (camo tent) and f4 (tan canvas), but only one at a time. We use the FR I run 05 image and the nine DIRSIG images in this analysis. We will only employ the spectral matched filter as the detection algorithm. The emphasis in Sections 7.3.1 – 7.3.4 is to demonstrate the behavior of utility as the parameters under investigation are varied. Section 7.3.5 attempts to quantify and rank the

sensitivity based on the FR image and the DIRSIG set of images.

7.3.1 Scene Parameters

The intent of this section is to explore what effect the composition of the scene might have on the utility metric. Intuitively, we believe that the spectral content of the scene will be the most important factor in the detectability of a particular target, as it determines how the target interacts with its surroundings. We would like to see the utility metric reflect this intuition. Specifically, we want to see the effect that a more complicated background will have on the utility metric. This includes the presence of very small numbers of pixels that are very like the target or are impacted by nonlinearities such as shadow or non-Lambertian surface effects. We look at two of these, the presence of target of interest in the scene and the scene spectral composition.

7.3.1.1 Target Presence

In order to investigate the effect of target pixels of the material of interest in the scene, we choose to work first with the FR I Run 05 image, since we have a target mask available as part of the canonical data set. We select target c6, the woodland camouflage tent material because there are 172 pure pixels of this material in the image. This corresponds to roughly 5×10^{-4} of the total image pixel count, which is our baseline specified PFA. Thus, we have the situation in which we know that the image contains the targets that we seek. We pose two questions in investigating the effect of target presence on utility. The first is, “How sensitive is utility when just the target of interest is present in the image?” In order to do this, we need a target free scene so that we can implant the target of interest in an increasing number of image pixels. Thus, we incrementally replace the first 172 pixels of the target free image with c6 pixels so that total number of pixels is maintained. The target free image has 387,149 pixels after the target mask is applied to the original image. The second question is, “How sensitive is utility when the target of interest and interfering target signatures are present in the image?” In order to implement this, we use the same pixel replacement scheme as in the target-free image, except with the original 394,240 pixel FR I image with targets resident in it. Figure 7.61 shows the results of assessing the utility of the original image and target free images for two targets: c6 and f4. We see that the utility for target c6 is very sensitive when there is no other target in the image as c6 pixels are added into the image. The utility assessed using target c6 is insensitive to addition of c6 pixels when there are interfering target pixels in the image – it is equally poorly in this situation. Utility for f4 is

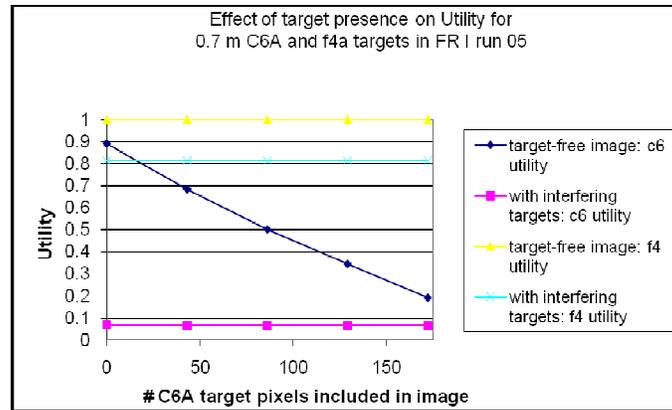


Figure 7.61: Utility versus number of c6 pixels in the FR I Run 05 image

insensitive when c6 pixels are added to the original and target-free images. The detection operator formed for f4 is not affected by presence of target c6 in the image. The lower utility for both targets when assessing the original FR I image is due to these interfering target pixels causing the covariance estimate used in the detector to be unable to suppress the target of interest. Clearly, utility is sensitive to the presence of the target being sought if it is the only target in the image. If there are a significant number of target-like pixels in the image, then utility will be significantly lower and insensitive to the presence of the target. If there are not other interfering signatures, then the utility will be highly sensitive to the presence of the target of interest. Keep in mind that we are using a spectral matched filter, which is not the optimal filter for suppressing interfering signatures, and results may be different for a different detector.

We repeat this investigation using a 1 m c6 target implanted by pixel replacement into the DIRSIG 4 m SNR 100 image. We incrementally replace original image pixels with 10 random c6 pixels generated using reference library statistics into the DIRSIG image, which had no c6 pixels in it originally. Figure 7.62 shows the effect on utility of replacing pixels in a target-free image.

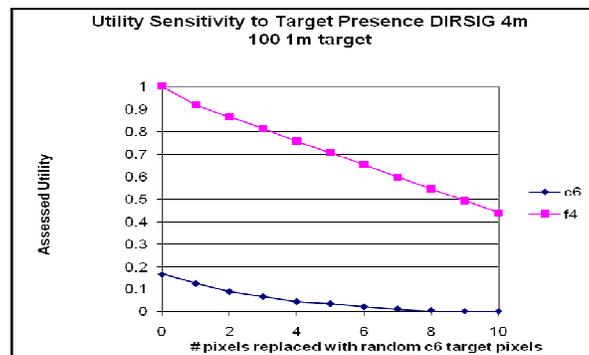


Figure 7.62: Utility for DIRSIG 4 m 100 SNR image plotted against number of c6 target pixels in the image

The utility assessed for both the c6 and f4 targets clearly decreases with more c6 target pixels introduced. The utility based on the f4 target is more sensitive since it falls to a lower utility than the utility based on c6. Thus, even though the two spectra are different, the presence of c6 significantly degrades the utility when the filter seeks target f4. The utility based on c6 is low even when no c6 pixels are resident in the image, as this is a more difficult detection situation. These two examples lead us to conclude that utility may be very sensitive to the presence of the target of interest in the scene.

7.3.1.2 Scene Composition

We investigate the effect of varying the scene constituent materials on the assessed utility. We first use the DIRSIG 4 m image to accomplish the demonstration. We take pixels from different spectral classes in the image and form a new “image” from this collection of pixel vectors. We then assess the utility of the composite image for different spectral class combinations using the c6 target. We progress from a simple detection scenario of the target against one background class to more complicated scenarios of the target against multiple classes. Using four spectral classes with 2000 pixels in each drawn from the original image, we formed all possible 2000 pixel combinations of those classes, leading to four single class possibilities, six two-class combinations, four three-class combinations, and one four-class combination. The utility associated with each class is plotted against the number of classes. Where multiple utilities are generated for the combinations in a class, the utilities are averaged. We use DIRSIG because we know exact ground truth that allows us to find unique and distinct classes. We use new asphalt, old asphalt, dead grass, and trees. Figure 7.63 shows the resulting plot of utility against number

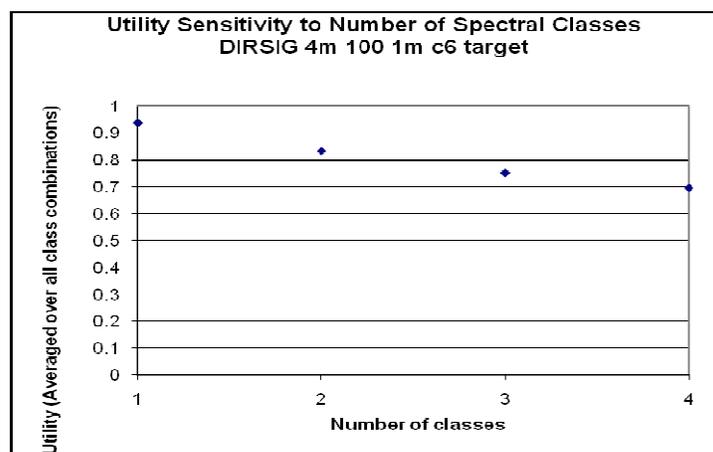


Figure 7.63: Utility for 1 m c6 target versus number of spectral classes in the DIRSIG 4 m 100 SNR image

of classes. We can see that the utility decreases with more classes, verifying the premise of a scene with more complex spectral character being a more difficult detection situation than one of homogeneous spectral composition.

In an attempt to reinforce the importance of the spectral composition of the scene and the fact that there will be some backgrounds with which the target spectrum interacts that make it challenging to detect the target against, we view the four single classes described above from a different perspective. We quantify the separation between the target and background class in multivariate statistical sense using the Mahalanobis distance between the target and each of the background classes and also add another class – healthy grass. We then plot the utility for each background class against this distance, as depicted in Figure 7.64. The two pairs of points

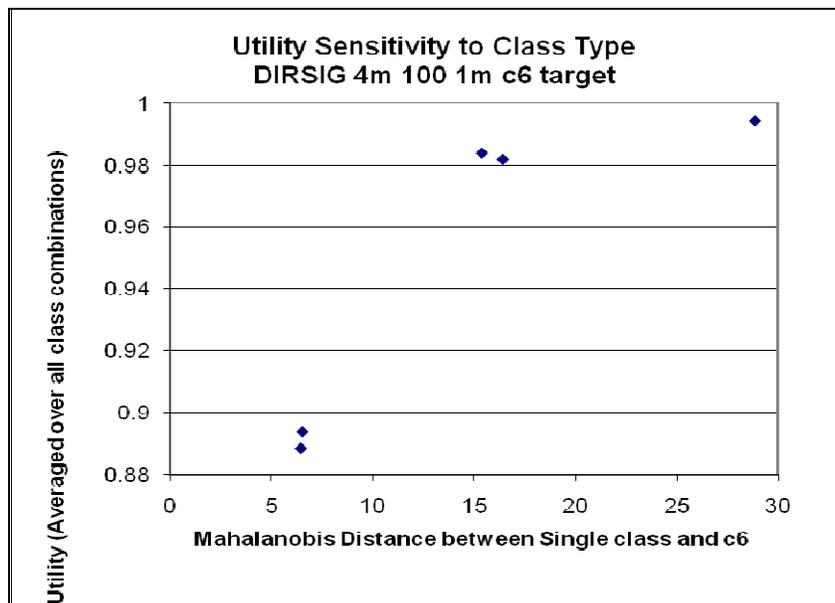


Figure 7.64: Utility plotted against target-background class statistical distance in DIRSIG image

correspond to the two asphalt and two grass classes. We note that we also ran this for target f4, but the utility was uniformly 1.0 for both of these experiments, making that particular target and image and fraction combination totally insensitive to the spectral composition.

An alternative and less rigorous approach is to use image subsections drawn directly from the image using regions of interest determined by the predominant material in the scene based on visual inspection of the image. This approach is applied to the FR I image and takes a grass and tree subset and then an equal proportion mix of the two and looks at the utility of each. This is plotted against the Mahalanobis distance from the target to each of the spectral classes. Figure 7.65 shows the utility for 0.7 m target c6. The grass background represents the lowest

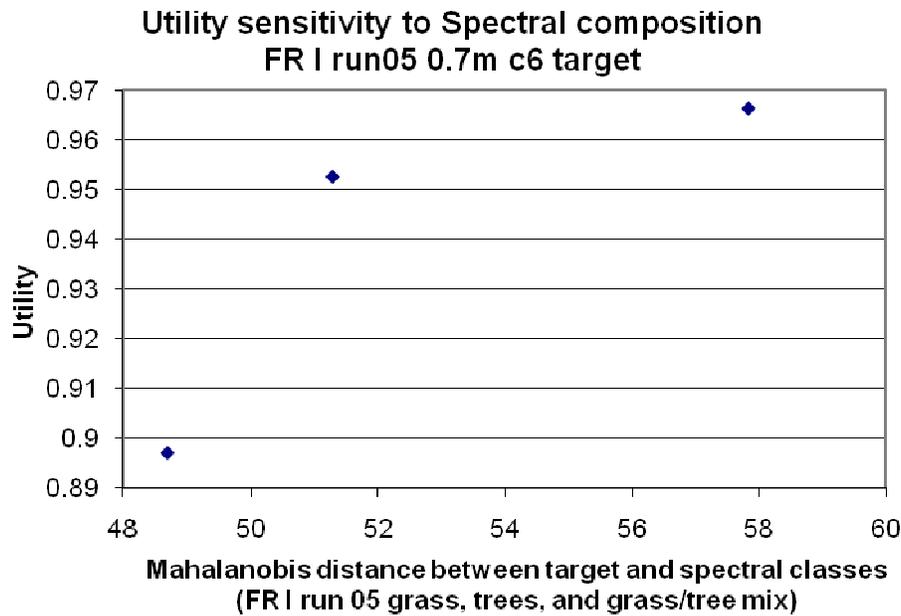


Figure 7.65: Utility plotted against target-background class statistical distance in FR I image

Mahalanobis distance and utility, the tree background represents the highest, and the mixture is in between. This shows that the number of spectral classes is not the ultimate determinant of utility, but rather the spectral character of the classes involved. There may be other methods of describing the utility sensitivity to scene spectral composition, but these are easily implemented. They confirm intuition that multiple spectral class images generally constitute more difficult detection situations and that the character of a specific spectral class and target interaction plays a large role in the assessed utility. The assessed utility appears to show moderate sensitivity to scene spectral composition.

7.3.2 Image Acquisition Parameters

In this section, we investigate the image chain parameters involved in image formation. First, we consider sensor design characteristics. We examine spatial resolution, sensor noise, and spectral resolution as the three dominant characteristics of the sensor. We next examine two image formation artifacts: misregistration and spectral shift. We do not look at radiometric calibration errors and do not attempt to pursue the many possible variants on atmospheric conditions and image acquisition geometry effects due to the effort involved in constructing experiments to investigate these effects. We rely on the DIRSIG images exclusively in this discussion, since we have total and independent control over the image formation parameters, whereas with real images, we have to accept those parameters as fixed.

7.3.2.1 GRD

We investigated the effects of GRD on the utility qualitatively in Section 7.1.3, where we compared the details of utility calculation for the 2 m and 4 m GRD image. We pursue that first look in more detail with Figure 7.66, by examining the trend in utility for 2, 4, and 8 m GRD

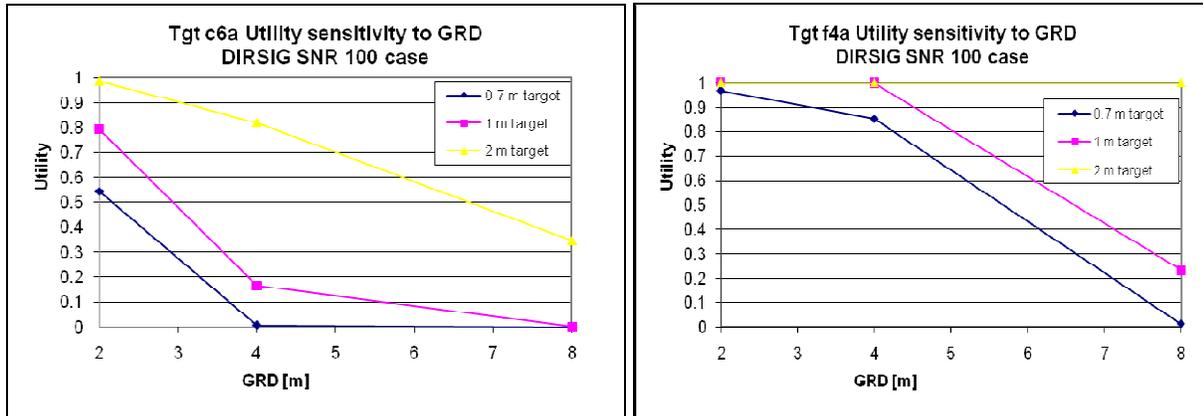


Figure 76: Utility sensitivity to GRD for DIRSIG images

images. We look at the effect of increasing image GRD on the utility assessed with the c6 (left) and f4 (right) targets. We use three different target sizes to emphasize the effect on the utility sensitivity of the operating point determined by the combination of target size, target type, and image. We will see this common theme of the operating point running throughout this investigation of utility sensitivity. While we illustrate operating point in the context of GRD here, it applies in every parameter that we examine, and makes the task of quantifying the utility sensitivity challenging. In cases where the detection operating point is too difficult or too easy, the effect on the utility of varying GRD is minimal, leading us to conclude that the utility is insensitive to that parameter. On the other hand, if the operating point is in a transition region of the utility versus parameter value plot, then we would conclude that the utility is more sensitive to that parameter. Note that the GRD interacts with the target size to control the separation between the target absent and present distributions. In the 4 m GRD images, the various target sizes investigated represent target implant fractions of 3%, 6%, and 25% for the 0.7, 1, and 2 m targets, respectively.

7.3.2.2 SNR

The investigation of sensor SNR entails using the three different images generated with DIRSIG to the specified SNR. We look at the same two targets as the previous section and examine

different detection operating points by keeping target size constant while examining three GRDs. Figure 7.67 shows the plots of utility against SNR. We see again that target c6 produces lower

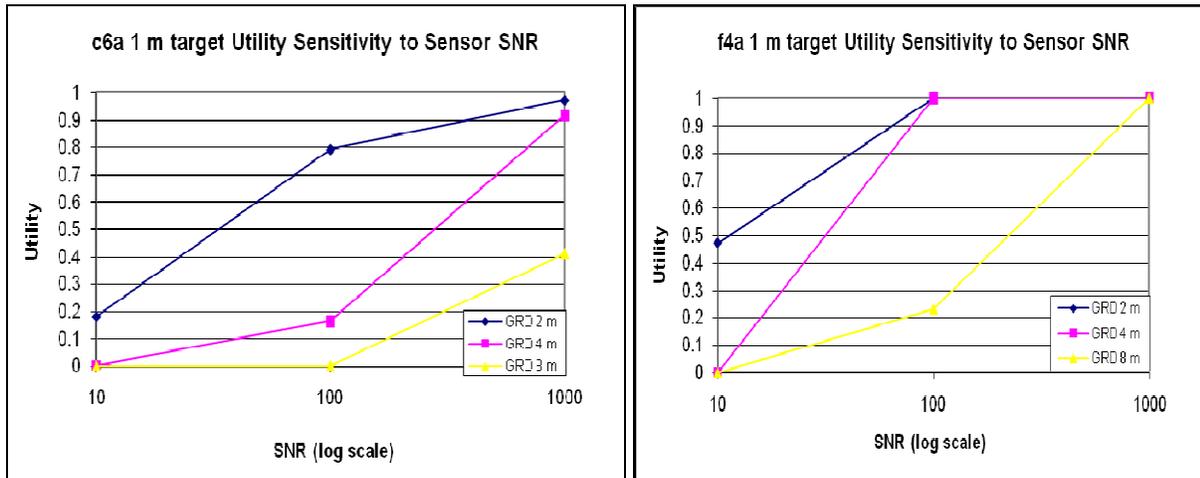


Figure 7.67: Utility plotted against SNR for DIRSIG images

utilities in these images than target f4. The increase in utility from 10 to 1000 SNR images is clearly not a linear one. The utility of the smallest GRD image exhibits more sensitivity at lower SNRs than the larger GRD images, whereas the larger GRD images appear to be more sensitive at the larger SNRs. We conclude that an easy detection situation implies that it will be more sensitive at low SNRs, whereas a difficult detection situation will be more sensitive at higher SNRs. The larger GRD is a more difficult situation because of the spectral mixing occurring in each pixel, and it is offset by higher SNR, whereas the smaller GRD has more readily identifiable spectral components and needs lower SNR for the utility to show significant change.

7.3.2.3 Spectral Resolution

The trend in spectral resolution is for utility to decrease as the channel bandpass is increased. This is due to the loss of spectral detail, as we saw in Section 7.1.3. We looked in detail at the spectra associated with two cases of spectral resolution. Here, we examine the utility associated with three spectral resolutions achieved by aggregating adjacent channels. We aggregate two, four, and eight channels to create average spectral resolutions of 20, 40, and 80 nm. We adjust each level of aggregation to maintain a constant SNR by dividing by the square root of the number of channels aggregated. Figure 7.68 shows the utility plotted against spectral resolution for three sizes of the c6 target and one size of the f4 target in the 4 m DIRSIG image. This plot shows that the utility of the 4m GRD image for the 0.7 m target is very low for all spectral

resolutions. The 1 m and 2 m targets respond most strongly to the doubling of spectral resolution from 10 to 20 nm. We also include the f4 target at 1 m size and see that it is insensitive up to 40

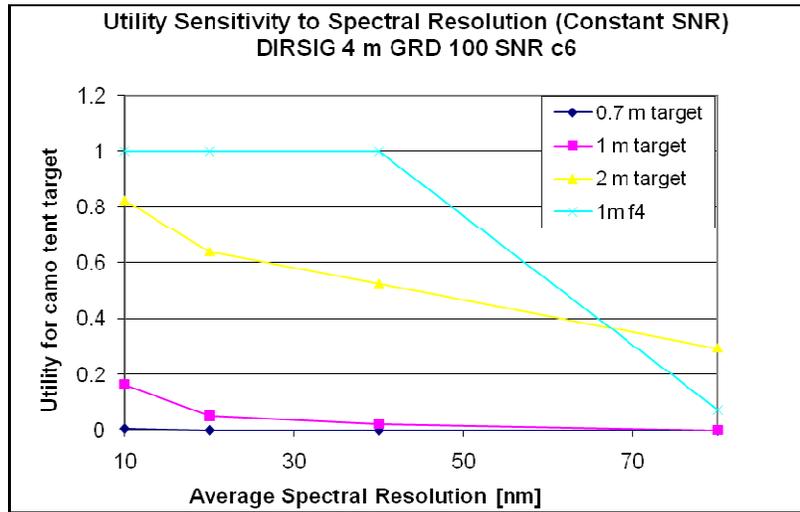


Figure 7.68: Utility plotted against spectral resolution for 4 m 100 SNR DIRSIG image

nm and then decreases significantly at 80 nm. By looking at 0.7 m, 1 m, and 2 m targets, we create difficult, medium, and (relatively) easy detection situations. It appears that the most utility sensitivity is in the transition from 10 nm to 20 nm spectral channel bandwidths. Beyond that, the loss of spectral resolution does not seem to make a significant difference.

7.3.2.4 Misregistration

In this section, we investigate a phenomenon that could be a factor in sensors designed with different focal plane arrays and may suffer a spatial shift between the images, resulting in spectral channels that are not spatially coregistered. Figure 7.69 shows the effect of shifting the

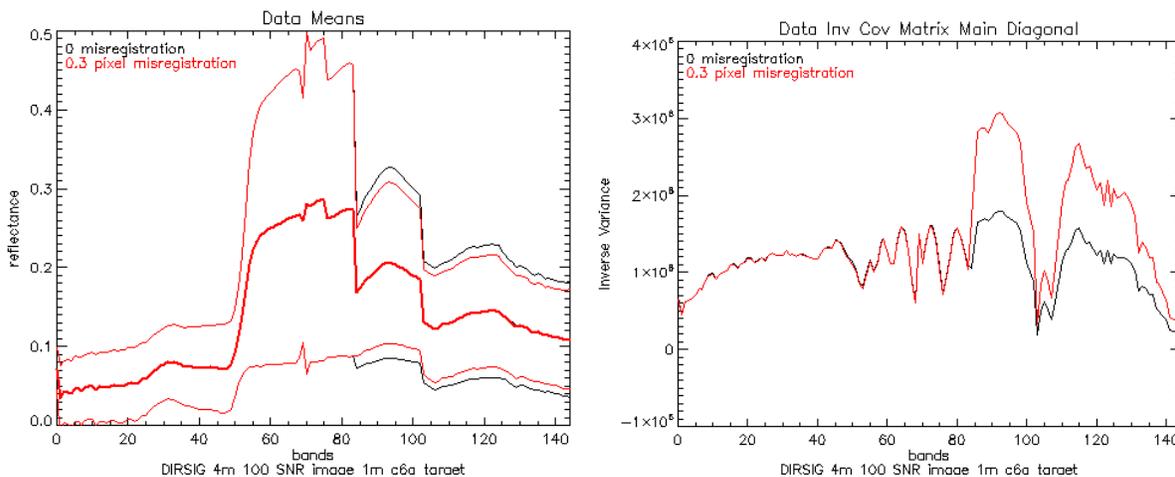


Figure 7.69: Image mean and inverse covariance diagonal illustrating misregistration

DIRSIG 4 m 100 SNR image at the VNIR/SWIR interface (channel 84 in these plots) by a fraction of a spatial pixel so that each pixel in the SWIR bands is shifted slightly to the right of those in the VIS bands as one views the image. We see clearly that a 0.3 spatial shift to the right does not affect the image mean, but does impact the variances associated with the image. These are slightly smaller (left side of Figure 7.69), and result in larger inverse covariance diagonal (right side of Figure 7.69). Thus, as each pixel vector becomes replaced with one third of its neighbor to the left, the amount of spectral variance in each band of the SWIR decreases.

As we increase the amount of misregistration up to a whole pixel, the utility gradually decreases, as shown in Figure 7.70 for two sizes of target c6. In general we see utility decreasing

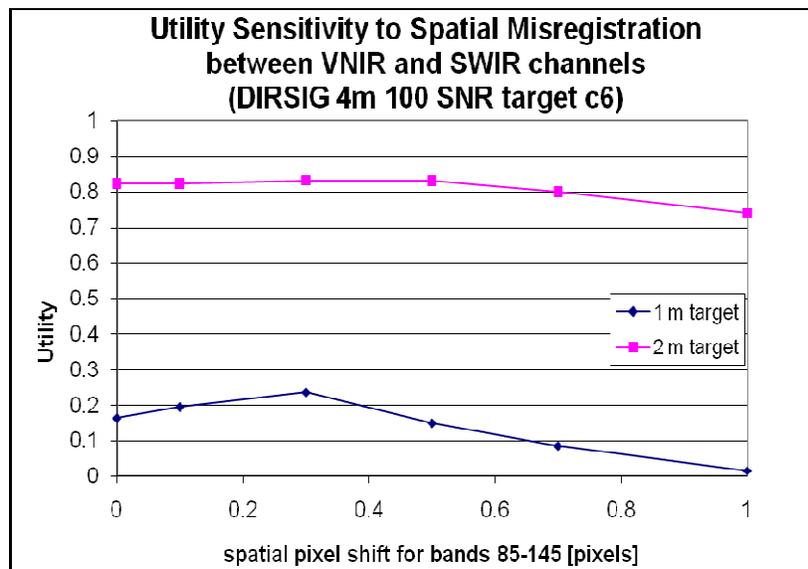


Figure 7.70: Utility plotted against pixel misregistration

with increasing misregistration, but the 1m target displays an increase at 0.3 pixel misregistration. The utility decreases because the threshold defined by the specified PFA gets larger as the amount of misregistration increases. The 1 m target displays a slight increase in utility at 0.3 pixel shift. An interesting effect is that a little misregistration actually seems to improve utility. This is because we are trying to match a reference target spectrum to image pixels that have been spatially misregistered at the VNIR/SWIR interface. The image covariance is based on misregistered pixels, but the target mean minus the image mean is going to give a result that is based on an unshifted target mean vector.

The spatial orientation of the false alarms with the misregistration is interesting. Those pixels identified as false alarms appear to shift with the misregistration. Figure 7.71 shows the spatial distribution of the first twenty false alarms, with green indicating the false alarms in the

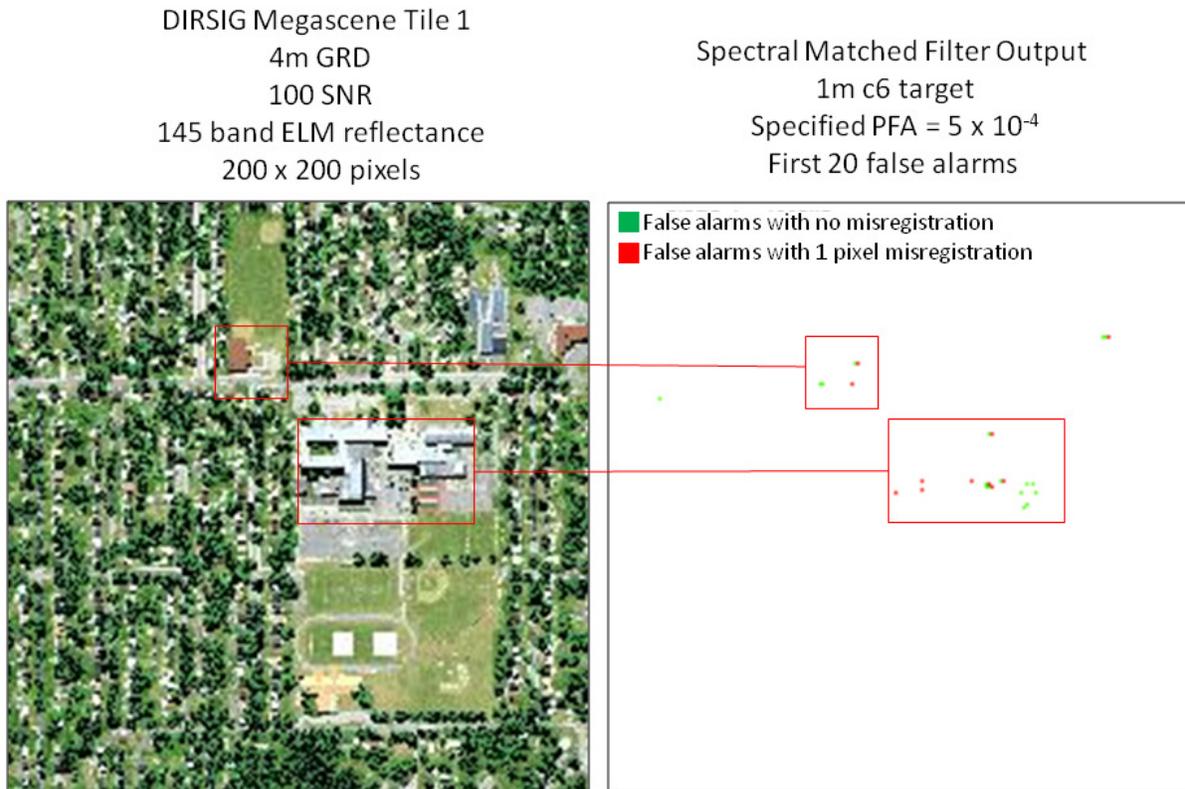


Figure 7.71: Spatial distribution of false alarms in original and misregistered DIRSIG images

original (unshifted) image and red indicating the false alarms when there is 1 pixel misregistration of the SWIR bands to the right.

We observe a few points about the sensitivity of utility to misregistration. First, the spectral character of the scene is important. In a spatially (and spectrally) uniform background, misregistration would not matter very much since mixing spectra of neighbors does not change the image covariance. On the other hand, if there is a scene material variability, then there will be more impact. We could repeat this experiment by considering a spectrally uniform image instead of a complicated one, and should see the utility be insensitive to misregistration. Second, the direction (whether to spatially shift pixels right, left, up, down, or at an angle) of misregistration will affect the sensitivity of utility based on image content. Third, the part of the spectrum that is shifted will make a difference relative to the target spectrum being sought by the detection algorithm. In this case, for the DIRSIG image with the c6 target, the utility does not seem very sensitive to misregistration in the SWIR. We note, however that our definition of utility is narrow and depends on a detector that is sensitive to target-image vector mismatches, so we are careful not to generalize here.

7.3.2.5 Spectral Shift

A sensor may suffer from spectral miscalibration, so that the center of channels is not what was originally designed. We simulate this effect by shifting the image spectra by a portion of a channel. When we create the target present version of the image, we also shift the random target to be implanted by an equal amount to replicate the true situation in a target detection task with spectrally shifted imagery. We investigate the effect of spectral shift on utility with a 308 x 308 pixel subset of the FR I run 05 image with no targets in it. The mean of the image subset with one standard deviation is shown in Figure 7.72 for no shift, 0.3 channel shift, and 1 channel shift.

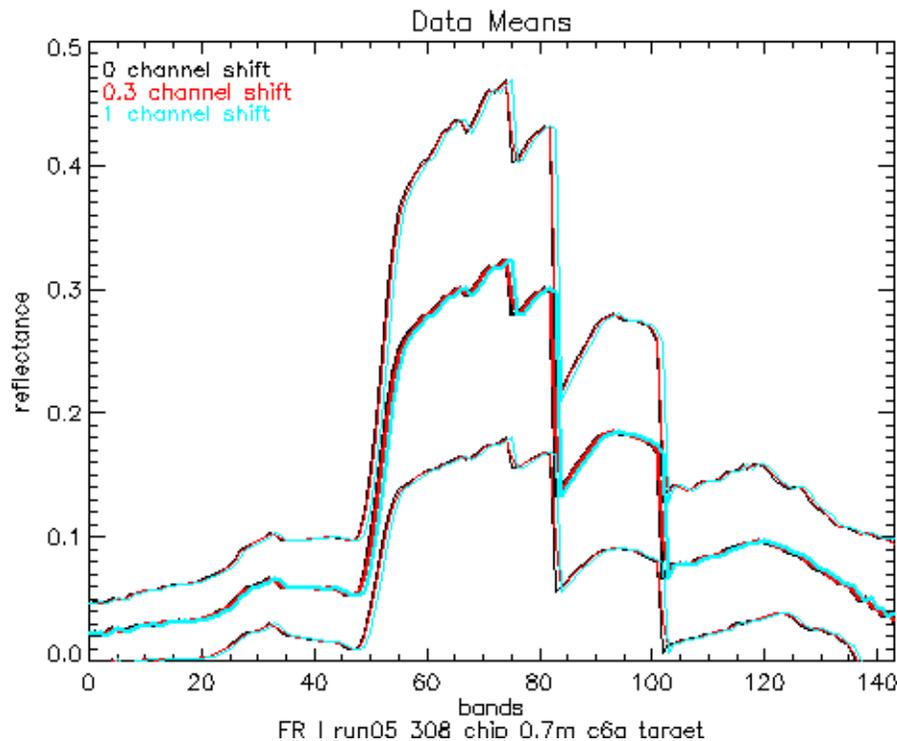


Figure 7.72: FR I Run 05 subset mean vector with different amounts of spectral shift

While the effect of the spectral shift is not immediately apparent in the image mean, it does produce a change in the inverse covariance matrix estimated from the shifted image and by extension, the filter vector used in the spectral matched filter. Figure 7.73 shows the result of no shift, 0.3 channel shift, and a whole channel shift on the inverse covariance matrix diagonal (left) and on the resulting filter vector (right). The fact that the diagonal of the inverse covariance is virtually the same for no shift and the whole channel shift is plausible, but the larger inverse at 0.3 channel shift is not clear. It is not surprising that the larger inverse covariance and 0.3 channel shift produces the filter vector that it does, with large weights concentrated in the area of largest inverse variance.

Chapter 7. Results

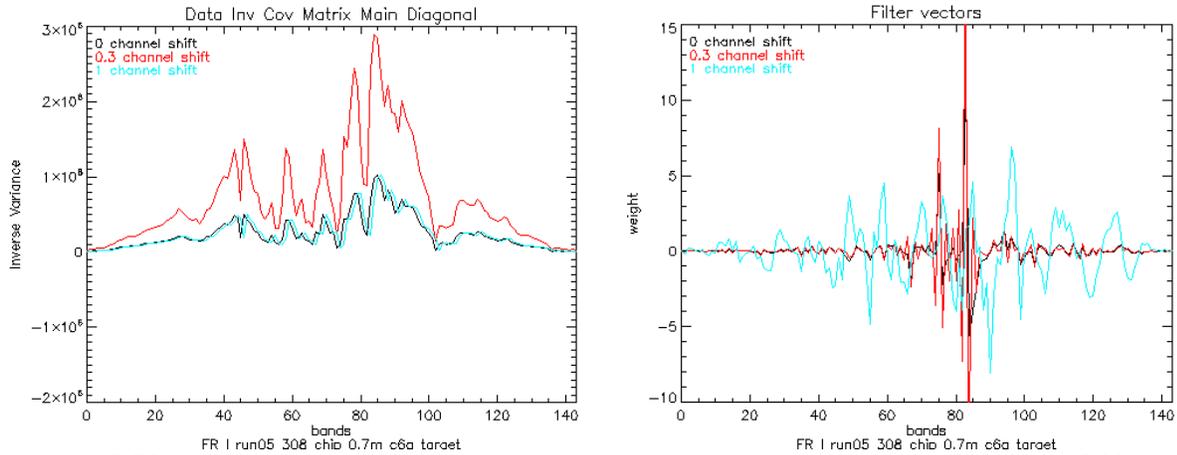


Figure 7.73: Effect of spectral shift on inverse covariance matrix diagonal (left) and filter vector (right)

The results of the utility at different spectral shifts are seen in Figure 7.74, along with the target absent and present distributions at no spectral shift, 0.3 channels, and whole channel shift. The trend is an initial increase in utility up to 0.3 channel shift and then a decrease with more

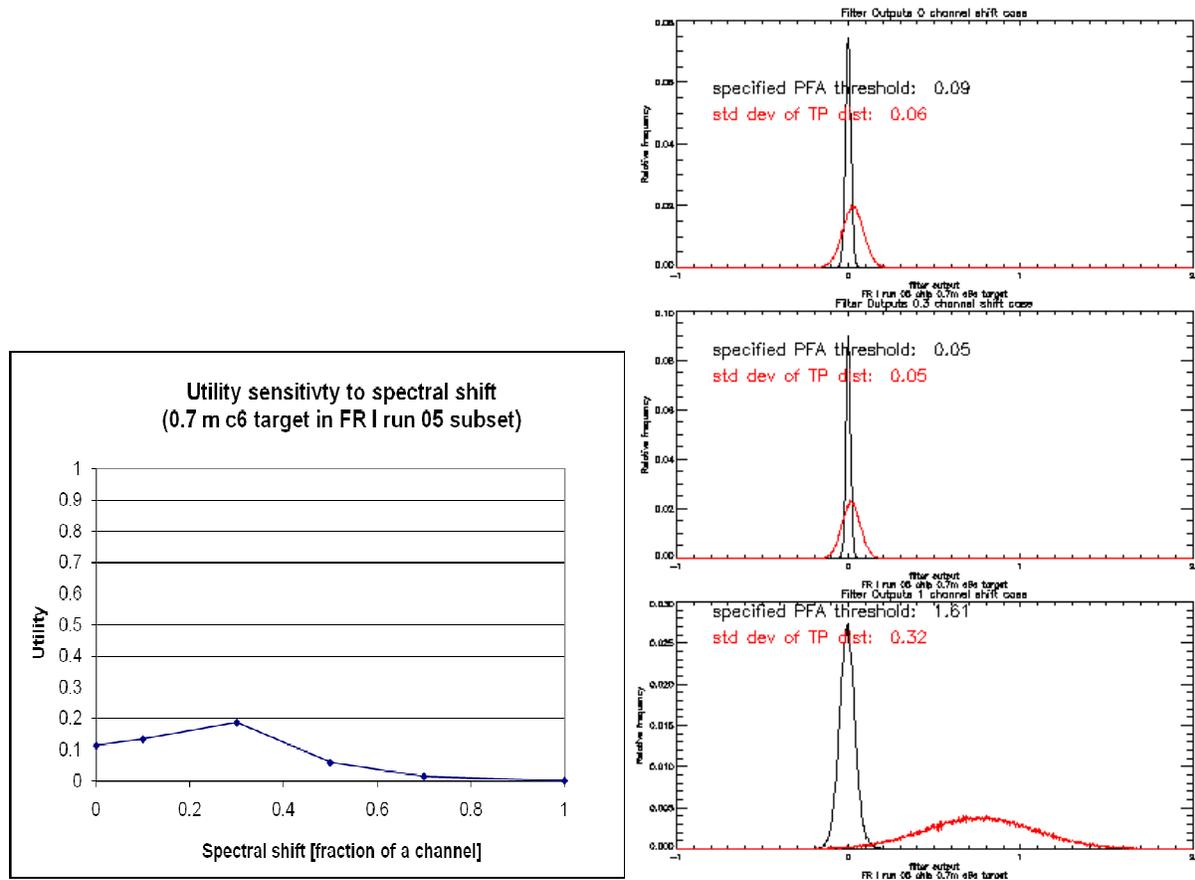


Figure 7.74: Utility plotted against spectral shift (left) and filter output distributions at three spectral shifts (0 – top right, 0.3 – middle right, 1 – bottom right)

spectral shift. The shift of the PFA threshold (47 pixels for this 308 x 308 image) from a larger to smaller value means that more of the TP distribution is captured before the specified PFA is attained, and these two cases' distributions look the same. But in the whole channel shift case, the distributions are more spread out and the PFA thresh has shifted to a much larger value, only getting a little bit of the TP distribution. The results of this inquiry are somewhat inconclusive. We cannot easily explain how the spectral shift is changing the image interaction with the target to produce decreased utility. This experiment was repeated with the DIRSIG image, and in that case, the utility increased with spectral shift.

7.3.3 Preprocessing Parameters

This Section explores the image chain parameters that the image analyst controls in order to improve the exploitation of information from the image. We are interested in the actions that an image analyst typically takes to make the image ready for target detection algorithm processing. The first and absolutely critical step is to transform the image from digital counts or calibrated radiance units into reflectance. The underlying assumption of the target implant method and the majority of target detection processing is that the image is in reflectance space. We begin by examining three different methods of atmospheric compensation applied to the FR I run 05 image, and note their effect on the utility. We also introduce perturbations to the gain and offset parameters of the ELM approach and observe the resulting impact on utility. Next, we look at the spectral range. Although this might more appropriately be termed an image acquisition parameter, since the upper limit of the wavelengths under consideration is determined by the sensor characteristics, we consider it here, since the analyst could conceivably choose to look at some subset of bands for application-specific reasons. We next look at the definition of those channels impacted by atmospheric absorption, which are commonly called "bad bands," and usually come in a bad bands list in the header file associated with an image. These could have an impact on the performance of an algorithm, so we examine the utility sensitivity to them. Finally, we look at the effect on utility of cleaning the data to deal with retrieved reflectance values that may be negative or greater than one.

7.3.3.1 Compensation Method

The HYDICE Forest Radiance canonical data set (MIT Lincoln Lab, 2004) includes reflectance images obtained by three atmospheric compensation methods: ELM, FLAASH, and ATREM.

We quantify these methods using the difference in each spectral channel between the retrieved reflectance for calibration panels and the ground truth reflectance for the panels. The error is reduced to a scalar by taking the mean over all spectral channels of the absolute difference between the true calibration panel reflectance and the retrieved reflectance. Figure 7.75 shows the ground truth measured reflectance and three retrieved reflectance spectra for the R2 (4% reflectance) panel deployed in

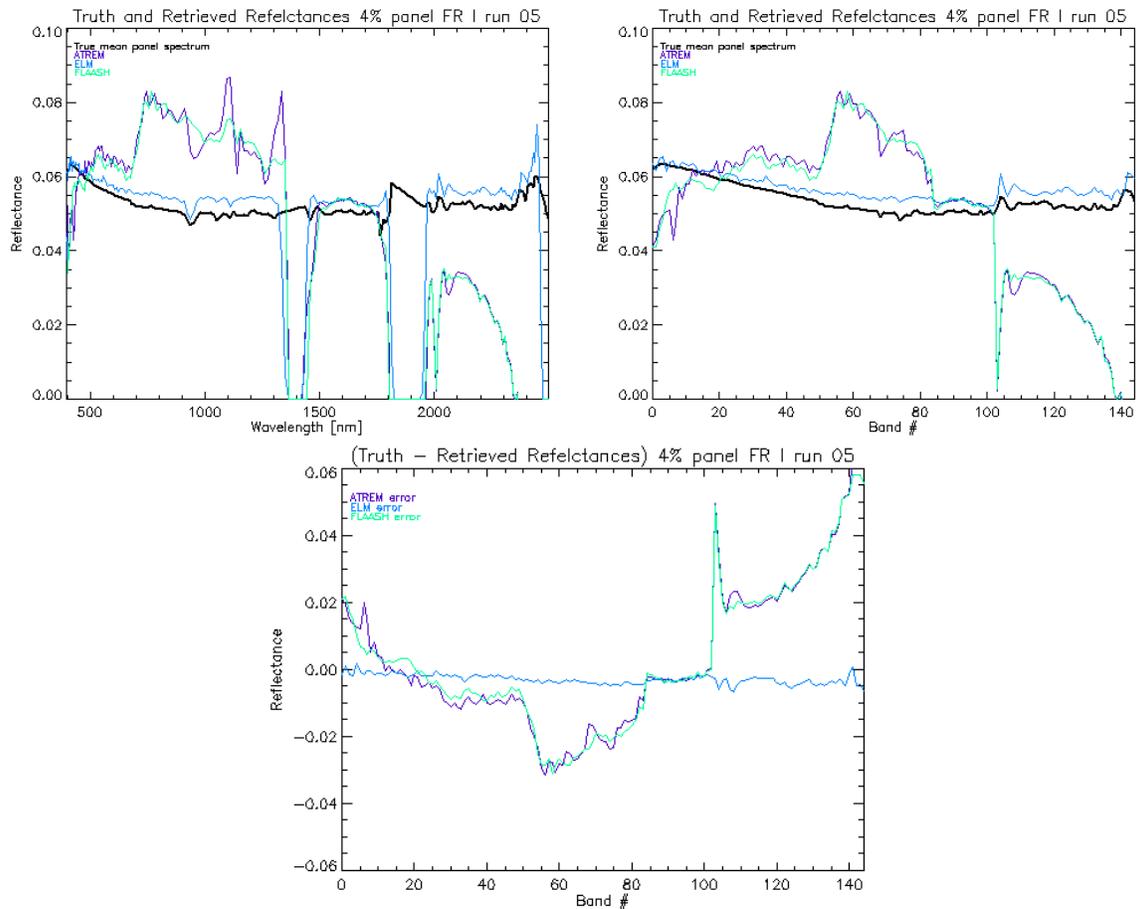


Figure 7.75: Comparison of retrieved reflectance spectra for 4% calibration Panel in FR I Run 05 image

the FR I run 05 image. The top left plot uses the full wavelength range of HYDICE, and includes the atmospheric absorption features. The heavy black line is the true mean spectrum of the panel. We see that ELM does the best job of retrieving the reflectance. The top right plot shows the reflectance spectra plotted against only the 145 channels free from atmospheric absorption effects. The bottom plot shows the error between each of the three compensation methods and

the ground truth spectrum. We see that the largest errors in the FLAASH and ATREM are in channels greater than 100 (wavelengths larger than 2000 nm).

We calculate the utility based on a 0.7 m c6 target based on the each of the three reflectance methods and plot them against the mean absolute error of the difference between retrieved and true reflectance spectra. This is shown in Figure 7.76 for the R2 panel and the R4 (16% reflector) panel for the original FR run 05 image (left) and the target-free image (right). The most notable characteristic in Figure 7.76 is that larger errors in retrieved reflectance cause

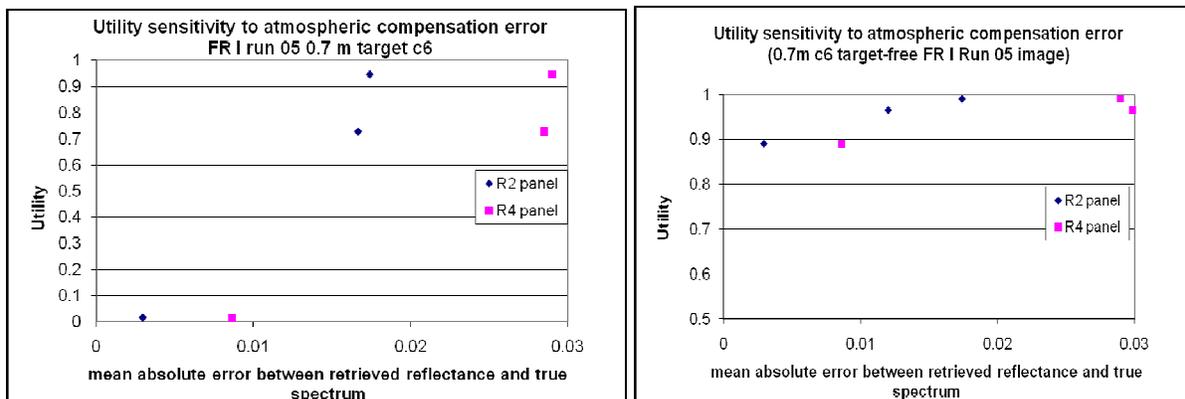


Figure 7.76: Utility plotted against retrieved reflectance error for FR I Run 05 image

higher utility. The order of the mean absolute errors from smallest to largest is ELM, FLAASH, and ATREM. Recall from Section 7.3.1.1 that the utility calculation is very sensitive to the presence of target pixels in the image. In this case, the image has 172 pure c6 pixels which greatly depress the utility. Comparing the utility values in the left and right sides of Figure 7.76 confirms that target presence pushes down the utility. The more subtle point is that the ELM is the lowest utility because ELM has retrieved spectra that “look” like the targets and FLAASH and ATREM do not due to the spectral distortion that they introduce in the retrieved reflectances. Thus, as the target-implant method creates the target present case, it injects a good reference target spectrum in every pixel of the image which in the case of ATREM and FLAASH looks very different from the background pixels. This creates an easier detection situation than in ELM, where the injected target is more readily confused with the existing target spectra. When in-scene target pixels are removed, the utility difference between compensation methods is less pronounced, but there is still an increasing trend in utility with compensation method error. We see that while having a good atmospheric compensation makes utility more sensitive to the background, but also see that our target-implant methodology will report higher utilities in situations where the image is spectrally distorted. The important point is that we need to ensure

that our reflectance image is as accurate as possible, because operationally, there is no way of knowing whether we are dealing with a spectrally distorted image or a good one. The target implant method is based on injecting a good reflectance spectrum to create the target present situation, so in cases where there is a disparity between the image spectra and the target injected, whether due to scene content or bad preprocessing, the utility will be higher because the detection will be relatively easy. Note again that conclusions about the utility sensitivity are predicated on the specific operating point, and we conclude that utility sensitivity to atmospheric compensation method is smaller for target free images than for those containing targets.

7.3.3.2 ELM parameters

In this section, we investigate the effect on utility of changes in the ELM gain and offset parameters. We pursue a simple approach, multiplying the existing gain and offset coefficients by a constant factor so that the coefficient vectors are scaled up or down in all channels. Although this may not be indicative of the true mechanism by which ELM might behave since it does not mimic any spatial dependencies, it does give insight into the sensitivity of utility to perturbations of the two parameters most important to ELM. An alternative approach would be to introduce a mismatch between the ground truth reflectance and sampled image radiance units in the formulation of the ELM gain and offset. We opted for the simple scaling approach since it lent itself to quantification on the x-axis of our plots of utility against the parameter of interest.

Figure 7.77 shows utility plotted against the mean absolute error between true reflectance panel spectra and the retrieved reflectance spectra after scaling of the ELM gain and offset.

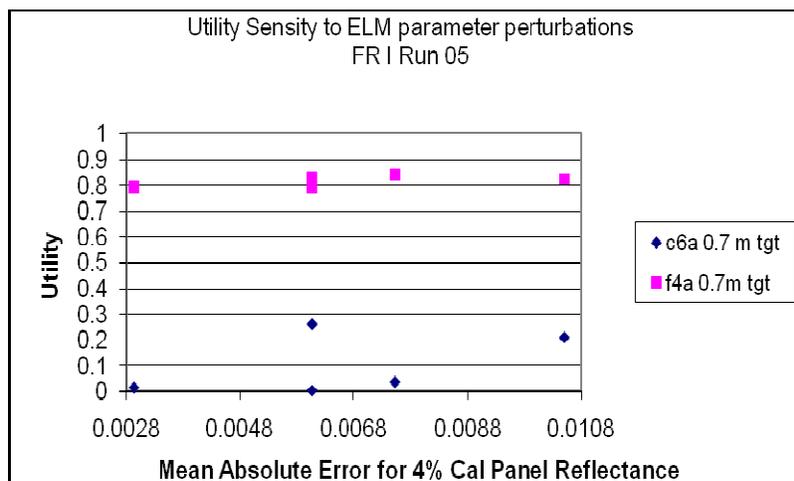


Figure 7.77: Utility plotted against reflectance errors for various ELM gain and offset scalings

We calculate the utility with two targets, c6 and f4 in the FR I run 05 image. Note that we have applied the same scaled ELM parameter coefficients to the random target vectors injected into the target present image so that we avoid the problem we encountered in the previous section with image distortion. The x-axis values are the mean absolute errors resulting from 10% scaling up and down of the gain and offset parameters independently. If we accept the mean absolute error between the retrieved and true reflectance as a good metric of “goodness” of a particular set of ELM coefficients, then we would conclude that utility is relatively insensitive to the mean absolute error resulting from bad ELM coefficients. In general, it appears that the f4 target is less sensitive to these changes than the c6 target. We also see that the utility for the c6 target increases as the retrieved reflectance deviates more from the true reflectance. This follows the trend observed in the previous section, in which the more accurate reflectance retrieval for target c6 produces lower utility due to the depression of utility when targets we seek are resident in the image when we use the target implant method of utility assessment.

Another approach to better understand the effect of perturbations to the ELM parameters is to use the scalar multiplier for the gain and offset as the dependent variable in our plot rather than the mean absolute error between retrieved and true reflectances. Figure 7.78 shows this for a 1 m c6 target in the DIRSIG 4 m 100 SNR image. The change in offset clearly pushes the utility

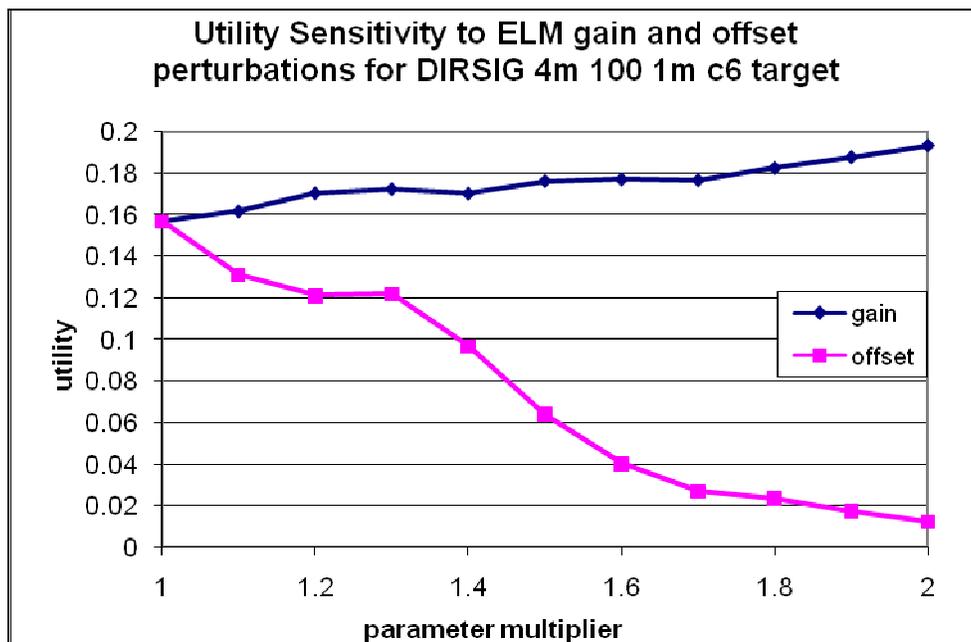


Figure 7.78: Utility plotted against scale factor for elm gain and offset parameters in the DIRSIG image

down as it is scaled with larger values whereas the increase in gain causes a slight increase in utility. Figure 7.79 shows the image mean spectrum for the baseline ELM parameters on the left and twice scaled ELM offset (top right) and gain (bottom right) relative to the reference c6 target spectrum. We see that the increased offset primarily impacts the first twenty channels of the

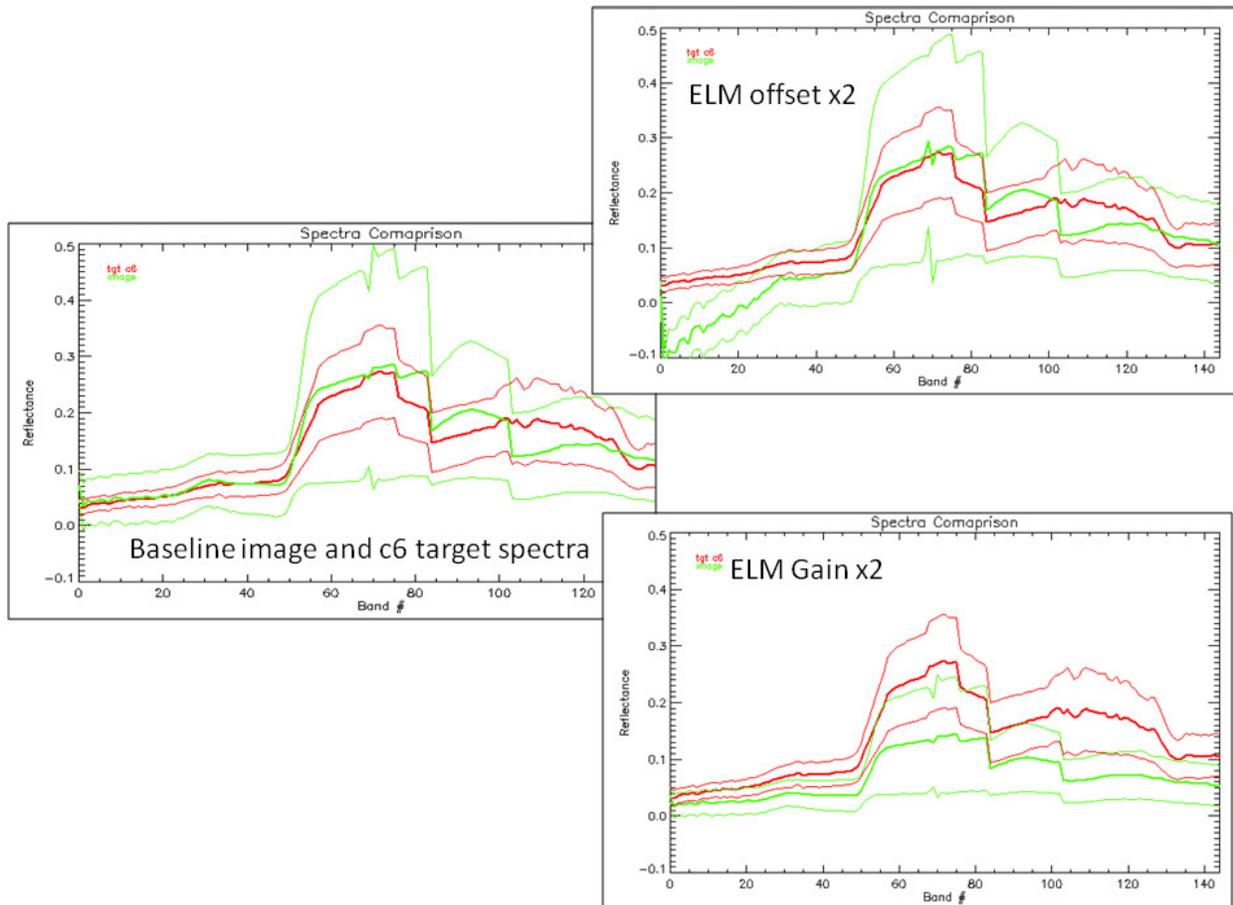


Figure 7.79: Effect of scaling ELM gain and offset on image mean spectrum

image mean, causing very small (negative) reflectances due to the magnifying effect of the offset compensating for the solar radiance in that part of the spectrum. The increase in the gain has the effect of depressing the image mean so that the target appears stronger than in the baseline case, leading to an easier detection situation and higher utility.

7.3.3.3 Spectral range

It seems intuitive that including more of the spectral range of a spectral image should help with detection, since more of the characteristic features of the target are available for use in the detection algorithm. We investigate the effect of the spectral range on utility by imposing an

Chapter 7. Results

upper limit on the range (750, 1000, 1500, and 1700 nm) of the spectral channels included in the image and target in the target implant method. Figure 7.80 shows the image mean along with the mean spectra and standard deviations of three target spectra (left) and the Mahalanobis distances plotted as a function of spectral range (right). The spectra (on the left) clearly illustrate the

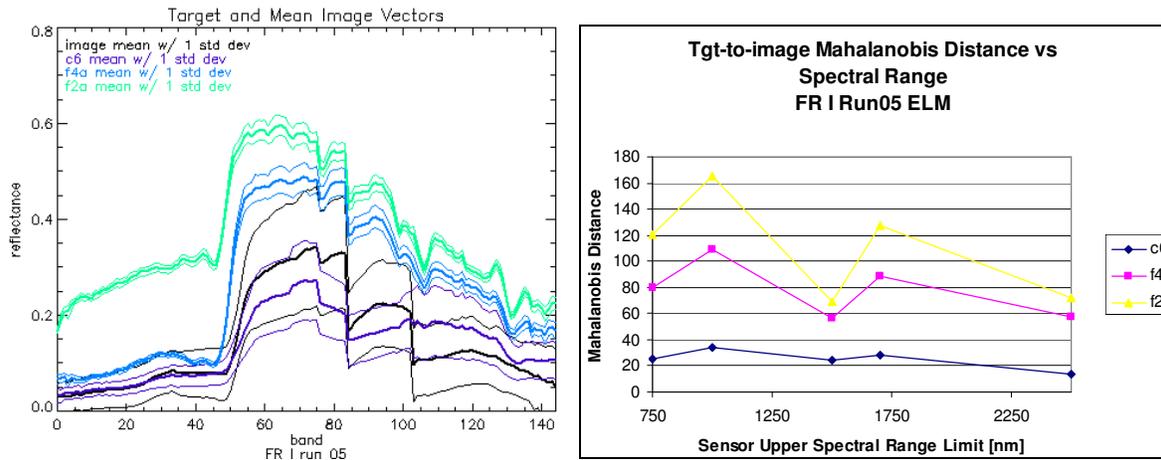


Figure 7.80: Image and three target spectra and Mahalanobis distances in FR I Run 05 image

supposition that spectral character that might help in target detection is more evident with a wider range of channels. The Mahalanobis distances (on the right), tell a different story, indicating less distinction between target and image as spectral range increases. This contradiction highlights the fact that simple observations of a high dimensionality vector space may not be indicative of the vector operations that are occurring in the linear filter operation that forms the detector.

Figure 7.81 shows the utility plotted against spectral range for several targets and target

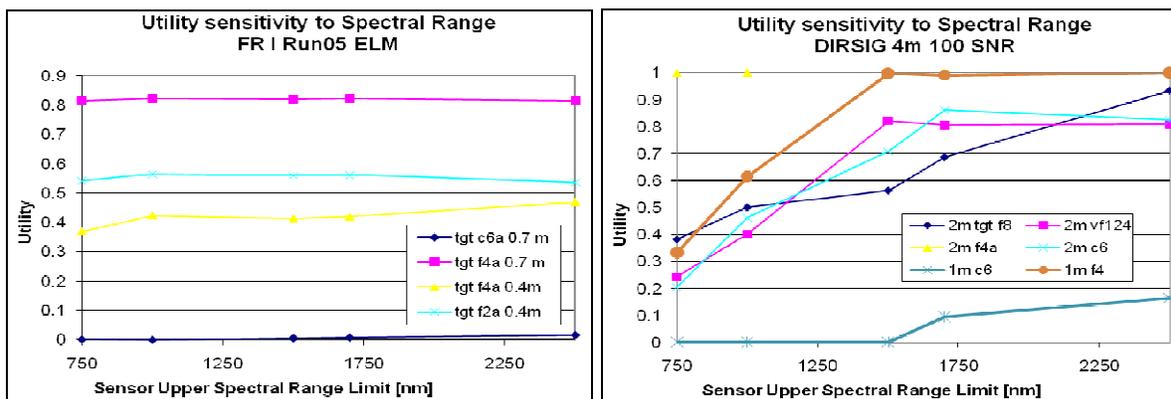


Figure 7.81: Utility plotted against spectral range in FR I run 05 and DIRSIG images

sizes in the FR I Run 05 image and the DIRSIG 4 m 100 SNR image. While the trend in utility plotted against spectral range is inconsistent in the FR image with both increasing and decreasing

utilities with spectral range, it is clear that the utility is not very sensitive in the FR image to changes in spectral range. The DIRISG image utility displays considerable sensitivity to the spectral range and has the increasing trend in utility with larger spectral ranges that we expect. This observation is again a statement about the importance of the operating point based on the image and target combination in determining the sensitivity of utility to a particular parameter variation. The trends observed in the utility do not track the trend in Mahalanobis distance, which is most likely due to the limitations of the Mahalanobis distance alone to characterize utility for real spectral image distributions.

7.3.3.4 Bad band definition

An important step in preprocessing an image prior to applying a target detection algorithm is to choose to operate in a feature space that does not include those spectral channels in which the signal is degraded by atmospheric absorption along the path to the sensor. Typically, these “bad bands” are identified as part of the sensor data product, so that the analyst can exclude them from processing. Because these bands essentially contain noise, they will either be rendered very large or very small by the atmospheric compensation routine. If such values were to be included in the vectors involved in the target detection filtering operations, they could cause spurious results, since they do not contain information about the scene but could potentially contain values that might cause the inverse covariances to become unstable or the vector multiplications to produce very large or small results. In the HYDICE FR collect, 65 of the 210 channels have been identified as “bad” in the header file. To see how these bad bands might manifest themselves, we look at the filter vector formed using all 210 bands as opposed to 145. Figure 7.82 shows the filter vectors (right) as well as the first 20 false alarms that result from application of the filter vectors to the 145 channel case (bottom left) and 210 channel (top left) cases. We see that the false alarms in both cases are the same, although the filter operator for each case is very different. The 210-channel filter vector places large weights on the atmospheric absorption region channels. These appear to have the effect of compensating to keep filter output relatively constant regardless of the presence of bad bands.

Figure 7.83 shows the impact on utility of the inclusion of bad bands in the target detection for the FR I run 05 image and the DIRSIG 4 m 100 SNR images. In the case of the FR image, we conclude that the utility is insensitive to the number of bad bands included in the utility calculation. In the DIRSIG image, we see a slight increase in the utility for the c6 target as

Chapter 7. Results

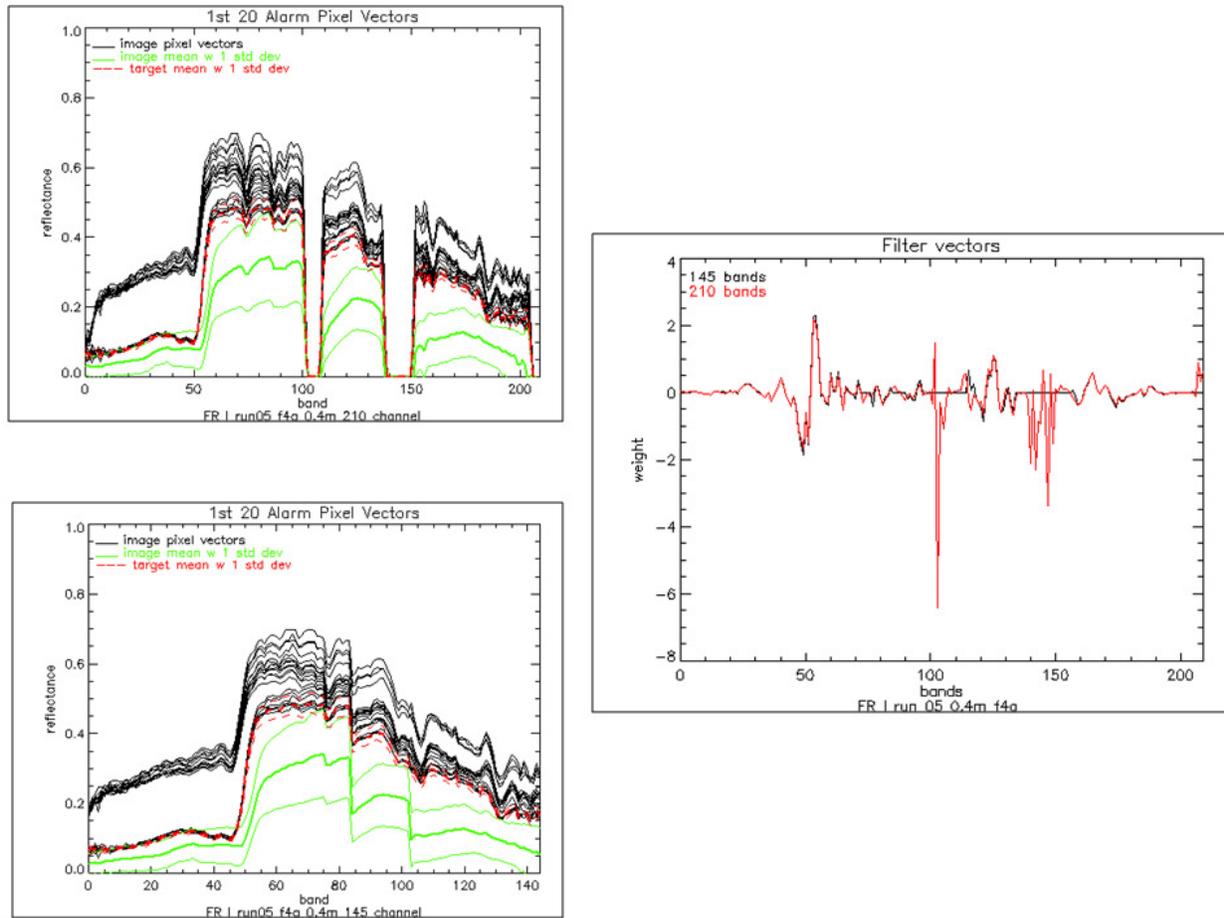


Figure 7.82: False alarm spectra and filter vector for 0.7 m c6 target in FR I Run 05

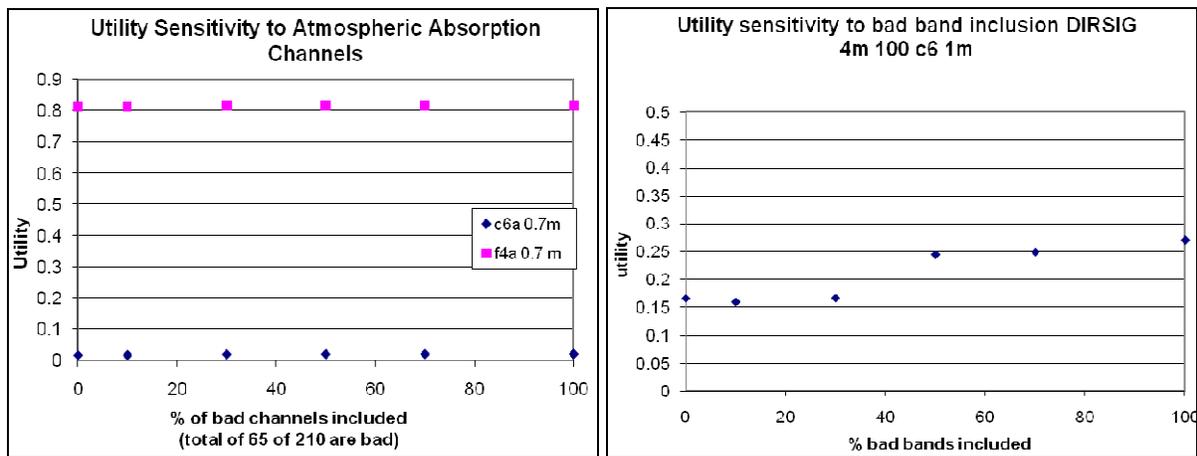


Figure 7.83: Utility plotted against percent of bad bands included

the number of bad bands increases. The f4 target utility is not plotted, but shows a constant utility of 1 for all bad band inclusions. Between the 30% and 50% points of bad bands included, the target spectrum absorption bands have small values in the absorption bands, but the image

absorption bands have large values, leading to a situation with a very different target and image spectrum, which makes for easier detection and thus higher utility. This discrepancy requires further investigation, and again indicates the unpredictable effect on the utility metric of mismatch between target and image spectra.

7.3.3.5 Bad reflectance values

One of the vestiges of atmospheric compensation routines is reflectances that may be larger than 1 or less than 0 for depending on the compensation algorithm. We investigate the impact of these bad reflectance values on the utility. As an example, 1.5% of the pixel elements in the ELM FR I run 05 image are less than zero and greater than one. This translates to 3.5% of the pixel vectors in the image being impacted. The input reference target spectra have no spurious reflectance values in feature space. Thus, we have the situation of where “fixing” the data to be more correct may make it more like the target so that utility might go down since it does not look as different as if it included the spurious values. We look at three cases of options to “clean” these spurious reflectance values in the FR image: case 1 is to include the pixels in the processing (which is the baseline we use), case 2 is to set outlier values to 0 and 1, and case 3 is to ignore the pixel vectors impacted by these bad values, which will change the number of pixel vectors evaluated. Table 7.12 shows the utilities in the FR I Run 05 image for these three cases applied to two targets in to the ELM, ATREM, and FLAASH reflectance images. We see that the utility remains the same or decreases as the values are set to 0 and 1, while it increases for ATREM and FLAASH and decreases for ELM. This is a subtle effect that defies easy explanation. Further investigation is required to better understand this effect.

	c6a 0.7m Utility	f4a 0.7m Utility
baseline ELM	0.02	0.81
set to 0 and 1	0.02	0.81
ignore	0.01	0.71
baseline FLAASH	0.73	0.96
set to 0 and 1	0.72	0.96
ignore	0.87	1.00
ATREM baseline	0.95	1.00
set to 0 and 1	0.92	1.00
ignore	0.98	1.00

Table 7.12: Effect on Utility of Data Cleansing for FR I Run 05 ELM, ATREM, and FLAASH Images

7.3.4 Target Detection Scenario Parameters

The parameters in the target detection scenario are those that the image analyst controls in order to answer specific informational questions about the utility of an image. The parameter behaviors are posed as questions about the impact on utility of changes in a particular parameter. Each of the subsequent sections will examine these questions that an analyst might pose and attempt to establish answers based on the FR and DIRSIG images. We first investigate the effect of target variability on the utility. Next we look at the impact of different types of targets on utility. Then we look at target size and the specified PFA over which to calculate the utility. We choose to not discuss the sensitivity to the type of algorithm here, since it was covered in Section 7.1.4.1.

7.3.4.1 Target variance

The variability of the target is an important consideration in designing a detection algorithm that is robust. We employ a simple method of scaling the target covariance matrix in order to simulate a more variable target. We are able to do this because we have access to target statistics from the FASSP library. We would expect that as the target variability goes down, detection performance will increase, but the behavior is not always this clear-cut. The determinant of target detection performance (utility in our case) is based on how the TA and TP distributions interact as the variance of the target is scaled. There are two observed trends: that the utility actually goes up with increased target covariance scaling and that the utility goes down with increased target covariance scaling. There are three major drivers of the relationship between the TA and TP distributions: target type, target fraction, and specified PFA. The target type (and its interaction with the image background) determines how the target absent distribution will look in terms of having large-valued outlying false alarm pixel responses. The target fraction determines the mean of the target present distribution, and as target implant fraction increases, the separability between the TA and TP distributions will increase. The specified PFA determines the threshold that will determine where on the TP curve the utility is calculated. Larger specified PFA means a lower corresponding threshold on the filter output, and that we will have to move through more of the TP distribution as we sweep the threshold on the filter output from large to small values, which makes the role of the outlier FAs less important than at small PFAs.

With these ideas in mind, we first examine the first case – increasing utility with target variance. Figure 7.84 shows this with the DIRSIG 4 m image and five 1 m targets. We observe

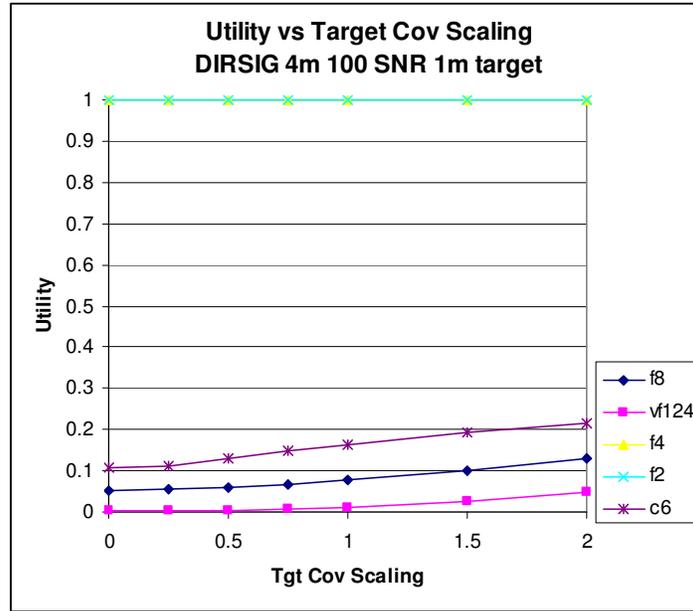


Figure 7.84: Utility as a function of target covariance scaling

that f2 and f4 are easy detections, so we see no discernable sensitivity to the target covariance scaling, but all of the others at this relatively small implant fraction (0.063), are showing the utility goes up with more target variability. We look at the TA and TP distributions for an explanation. Figure 7.85 shows these distributions for target c6 at covariance scaling of 1 (top row), 2 (middle row), and 0 (bottom row). Looking at the distributions in the left column, we

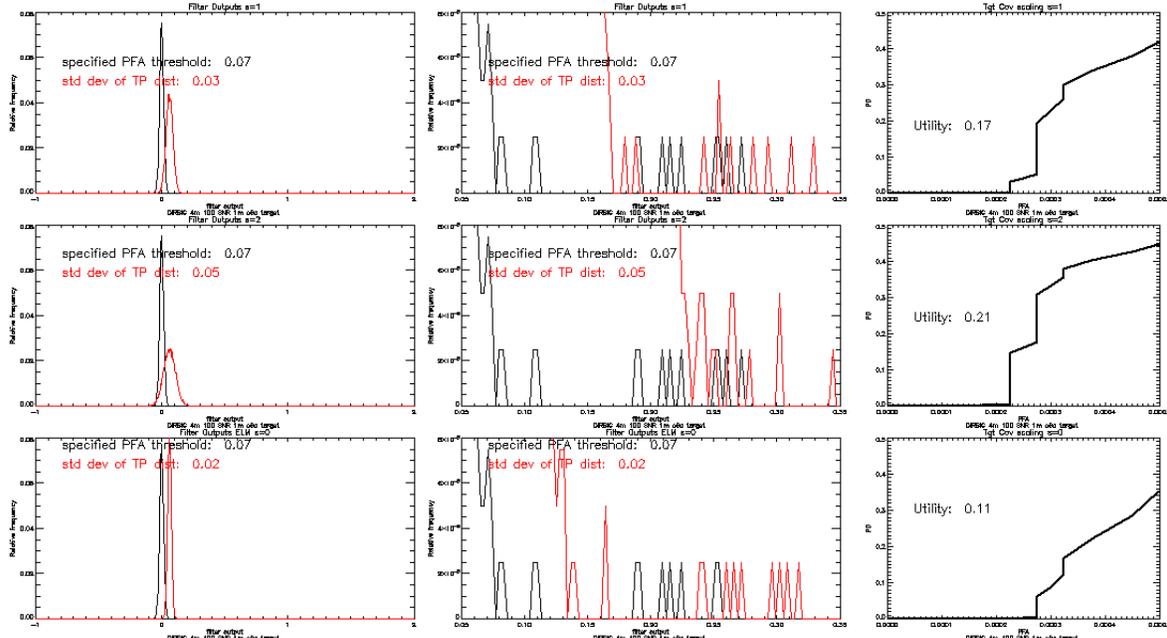


Figure 7.85: Distributions and ROC curves for 1 m c6 utility in DIRSIG image for three different target variability scalings

would think that smaller target covariance means tighter TP distributions, which should create an easier detection situation. But if we take a closer look at the right tails of the TA and TP distributions, including the 172 specified PFA false alarms, we see that the large false alarms are hurting performance by creating more PFA as we sweep the threshold before we reach our PD. Since utility is an area measure and “keeps track” of the “history” of sweeping the threshold (“how we got there”), it is significantly impacted by this “push” from the early FAs, as the ROC curve on the bottom right illustrates. Thus, even though we have a tight TP distribution with less target variability, because of the small fraction and the outlier FAs, we suffer from smaller utility. This reinforces the importance of the particular operating point, since we have several factors interacting.

Now we show the case in which utility decreases with increasing target covariance scaling. This is shown in Figure 7.86 for the same image with targets f4 and vf124. In this plot,

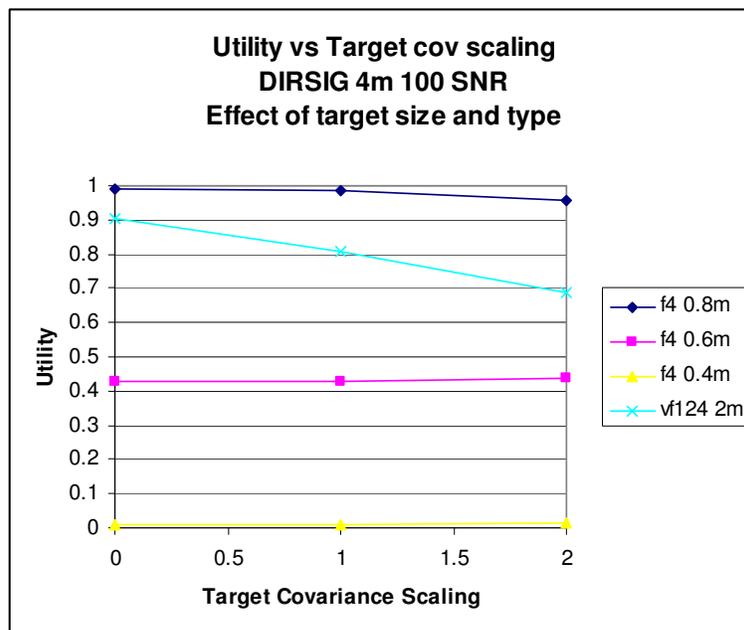


Figure 7.86: Utility plotted against target covariance scaling for two targets in DIRSIG image

f4 is an easy target, so by adjusting the fraction, we create a more difficult detection situation and see that as TA and TP move closer together, the same thing is happening with the early FAs causing utility to go up. However, if we look at vf124, which is a very difficult target, but increase the fraction to increase TA and TP separability, we see a different trend. In Figure 7.87, we plot the TA and TP distributions in order to better understand the increase in utility. Looking at the right edges of the TP distributions, we see why utility goes down as target variability

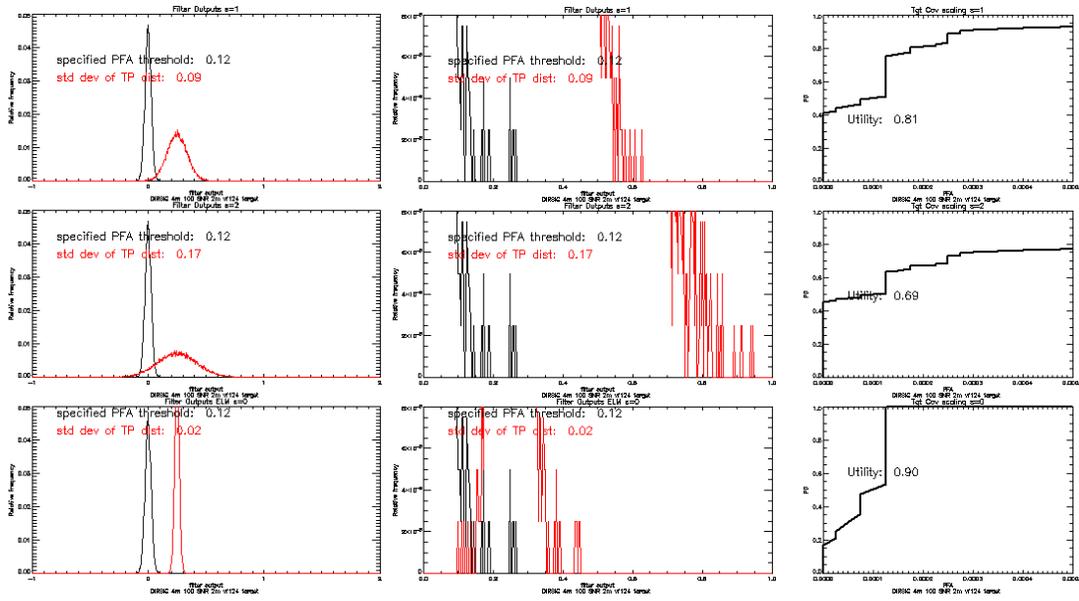


Figure 7.87: Filter output distributions for 2 m vf124 target in DIRSIG image

increases. In this case, a skinnier TP distribution means that we get through virtually the whole TP distribution before attaining our specified PFA, thus assessing a higher utility. We conclude that utility displays moderate sensitivity to the target variability, but the sensitivity and direction of utility change is determined by the unique operating point of the target detection scenario rather than a simple explanation of only the width of the target present distribution.

7.3.4.2 Target type

We have encountered the effect of target type on utility in all of the parameters that we have considered thus far. Here, we seek to isolate the effect on utility of the target type. First, we take a close look at three targets (c6, f2, and vf124) representative of a wide range of detection difficulties. Figure 7.88 shows the filter vectors associated with these targets in the FR I run05 image. The easy (f2) target has smaller coefficients, whereas the more difficult targets have the larger coefficients. This is driven by the inverse covariance of the data, and the difference between mean image and target vectors. Since the inverse covariance is the same in this case, the difference in means will be the largest contributor to the filter weights. The greater variability in the filter weights combined with the inherent variability of the target type translates to a bigger spread of filter outputs and fatter TA and TP distributions, which we see in Figure 7.89, with the top row representing target c6, the middle row target f2, and the bottom row target vf124 at 0.7m size in FR I run 05. We see that both difficult targets have significantly wider target present

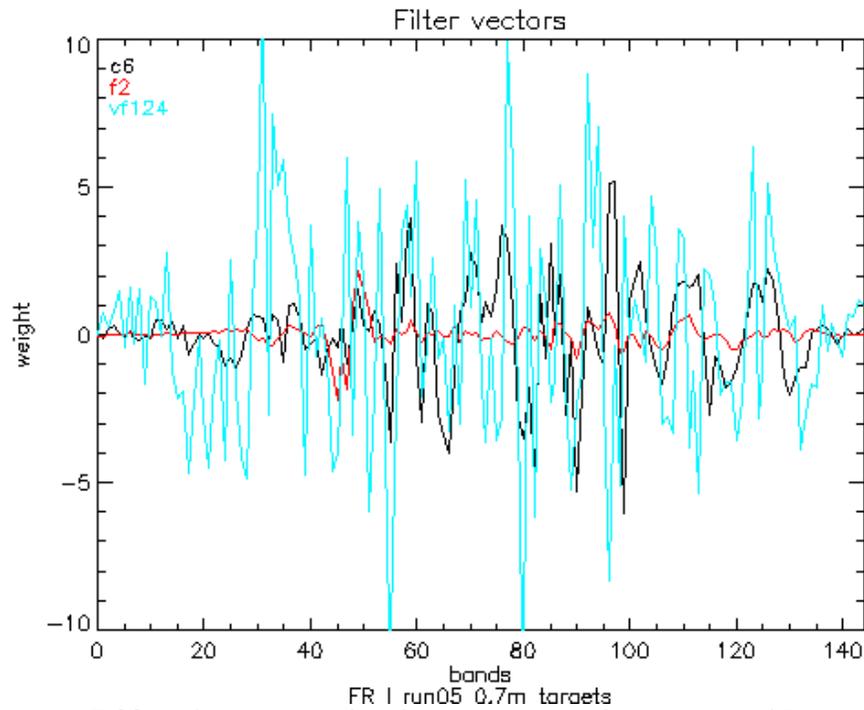


Figure 7.88: Filter vectors for three targets in the FR I Run 05 image

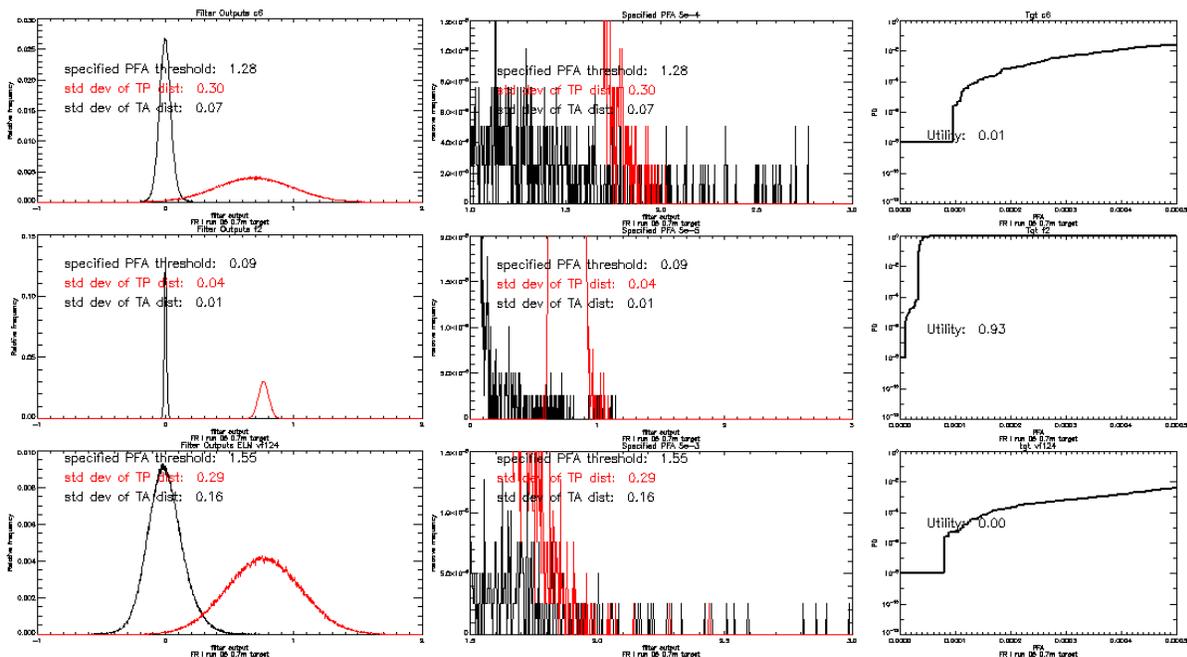


Figure 7.89: Distributions and ROC curves for three targets in FR I Run 05

distributions than the easy target (f2). This causes the first false alarms to make a large impact on the resulting utility calculation. Even though c6 encounters more false alarms above the largest TP output, the area under TP attained by the point at which specified PFA is reached is greater than that of the vf124 target. This is primarily due to the fatter TA distribution of the vf124. This

inherent variability is a characteristic of a particular target type, and is one way in which we could categorize targets.

Another way to categorize the sensitivity of utility to target type is to use the Mahalanobis distance to quantify target type relative to a particular background image. We calculate the utility of 19 0.4 m and 0.7 m targets in the FR I Run 05 image and plotted these against the Mahalanobis distance. We repeated this for our DIRSIG image using 1 m targets. Figure 7.90 shows this plot. In general, as the statistical distance increases between target and image mean, the utility goes up. Keep in mind that these results are for a single target implant fraction, and as we increase or decrease the implant fraction, it will compress the utility at upper or lower limits and change the apparent sensitivity of the utility to the statistical distance that describes the target type. In this case, the utility sensitivity was calculated by considering the slope of the linear fit line associated with all 19 targets.

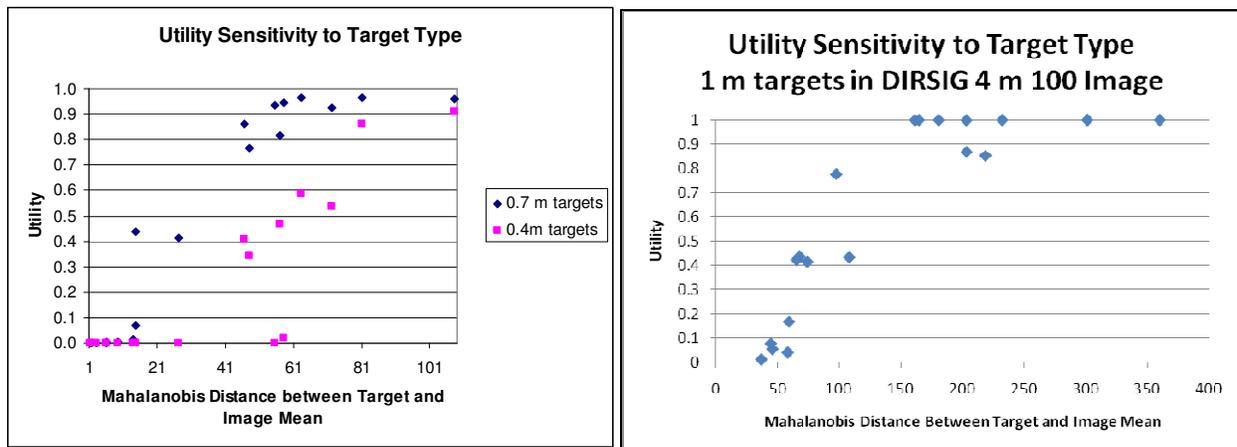


Figure 7.90: Utility of 19 targets plotted against their Mahalanobis distance for the FR I run 05 image (left) and the DIRSIG image (right)

7.3.4.3 Target size

We have seen the large impact that target size plays on the utility in the exploration of other parameters. Most obviously, it controls the separation between the target absent and present distributions. In Figure 7.91, we show the results of utility plotted against target size for the FR I and DIRSIG images for several targets. We see that the utility sensitivity to target size will be determined by the type of target and the range of target sizes that are considered. There are target sizes in which the utility will be insensitive, such as when the detection is very easy of very

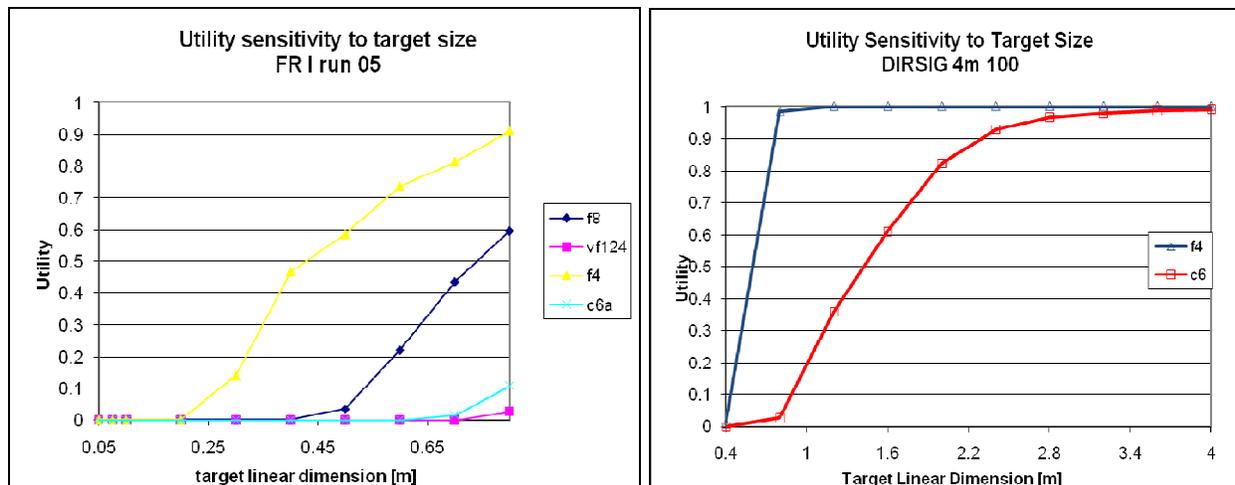


Figure 7.91: Utility versus target size for FR I Run 05 and DIRSIG 4m 100 SNR images

difficult. There is also a transition region of target sizes in which the utility is responding with large changes to target size changes. The explanation for the different behaviors is found in the explanations for target variability and target type, discussed in Sections 7.3.4.1 and 7.3.4.2. This interaction is complicated and requires examination at the filter output level, and we have not found any generalizable results at this point.

7.3.4.4 Specified PFA

The specified PFA is a parameter that the analyst specifies based on the requirements of the target detection task. We demonstrate the effect of specified PFA by considering the TP distribution closely as we change the specified PFA over three values using target c6 at 0.7m size in the FR I Run05 image. Figure 7.92 shows the details of the right tails of the TA and TP distributions along with the threshold corresponding to the specified PFA and the ROC curves. The top row of the Figure shows the specified PFA of 5×10^{-4} , the middle row shows 5×10^{-5} and the bottom row shows 5×10^{-3} . We see that at the very small PFA of 5×10^{-5} , we get nothing but false alarms by the time we hit the 20 pixels representing that specified PFA. At the largest specified PFA, we are able to accrue many detections before we reach the specified PFA threshold. Thus, while the shape of the distributions is not any different for the specified PFA, the location of our limit for calculating the probabilities that are used to create the ROC curve are changing and make a significant difference in the assessed utility.

Figure 7.93 shows the utility plotted against different specified PFA settings for several targets in the FR I Run 05 and DIRSIG images. We see that the utility is sensitive to changes in

the specified PFA in a manner that depends on the target type and image. We see all targets in the FR image displaying sensitivity, but only the c6 target in the DIRSIG image.

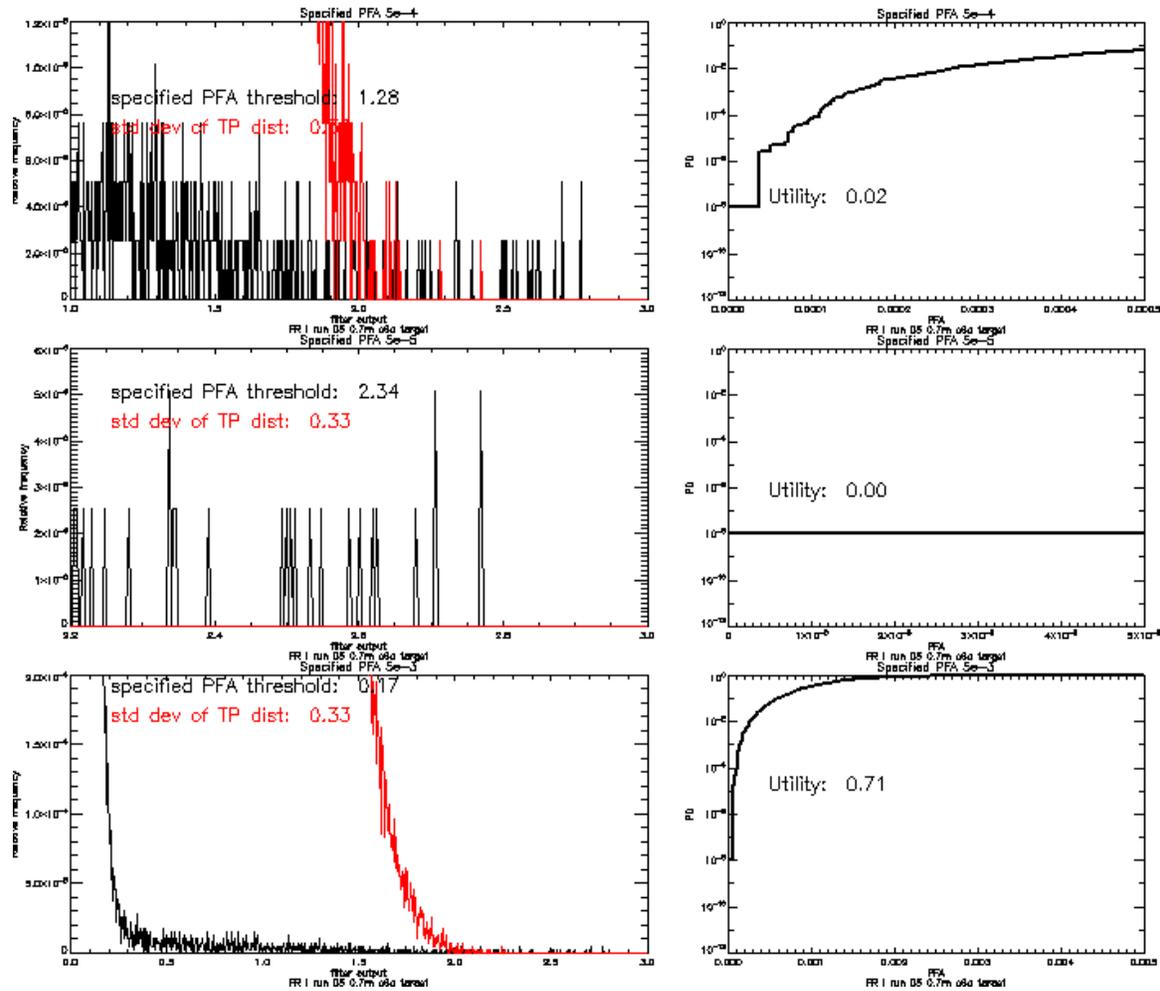


Figure 7.92: Filter output distributions and ROC curves for three specified PFAs

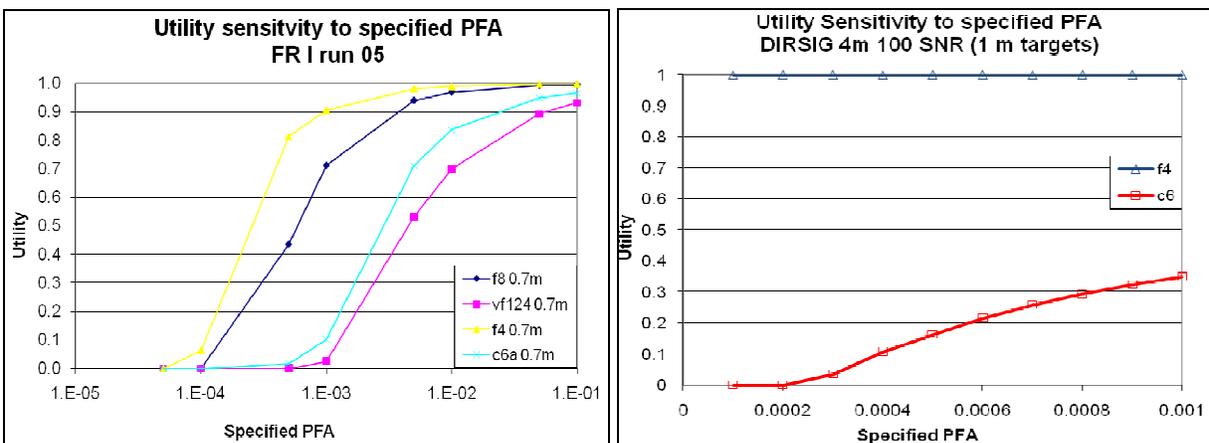


Figure 7.93: Utility plotted against Specified PFA in FR I and DIRSIG images

7.3.5 Ranking of Parameters

This section is an attempt to summarize and compare the utility sensitivity to the different image chain parameters considered in the previous sections. In Section 6.7.1, we addressed the methodology followed in coming up with a quantifiable measure of sensitivity and a system for ranking the utility sensitivity to the parameters. There are a few challenges to making equitable comparisons between utility sensitivities. The first is that the x-axis in each plot of utility versus a parameter is different. We normalize the parameter range using the maximum value to allow all the x-axes to be on a scale from 0 to 1, like the utility axis. Next, the question of where in the utility versus parameter plot to calculate the utility is an important one in that it will make a big difference in the sensitivity. This is the “operating point” that we alluded to in the previous discussions in Section 7.3. We have tried as much as possible to use the mid range of the parameter axis, having tried to select a reasonable range of that parameter variation to do the analysis. We have discussed the concept of the operating point on the utility versus parameter plot extensively, and this calculation of utility sensitivity is where we need to be careful to choose an operating point that is representative of a larger set of image chain parameters. That is difficult because of the small number of images and targets that we investigate. In cases where the data is available, we average the sensitivity to give a more widely applicable answer, however the construction of this sensitivity analysis designed to use a single target and fraction target detection scenario as much as possible. The second challenge is that we need a simple and fair measure of sensitivity that does not unduly bias any particular parameter. The simplest metric is the three-point instantaneous slope of the utility versus parameter plot at the mid range point. In spite of these challenges, we have managed to assign sensitivities to each of the parameters considered and have used that to create a rank ordering of those parameters that cause the greatest utility sensitivity.

We first considered the utility sensitivities associated with the FR image. We could not look at the sensor parameters here, since we could not control them independently. Table 7.13 shows the parameter under investigation, sensitivity score, and relative rank on the left and ordered rank of each parameter on the right side. Table 7.14 shows the sensitivity score and ranks for the DIRSIG image. We did not create FLAASH and ATREM reflectance images in DIRSIG, so these parameters are not in the consideration. However, the DIRSIG images allowed us to consider the sensor parameters.

Category	sensitivity	Rank	Ordered Rank	Parameter
Scene			1	specified PFA
target presence	0.68	4	2	target type
class type	0.44	5	3	Compensation method
Preprocessing			4	target presence
Compensation method	1.12	3	5	class type
ELM gain	0.02	9	6	target covariance scale
ELM offset	0.02	10	7	target size
bad band inclusion	0.00	11	8	spectral range
spectral range	0.03	8	9	ELM gain
Detection Scenario			10	ELM offset
target covariance scale	0.06	6	11	bad band inclusion
target type	1.15	2		
target size	0.04	7		
specified PFA	11.25	1		

Table 7.13: Sensitivity Scores for FRI Run 05 Image Parameters

Category	c6	f4	Average sensitivity	Rank	Parameter	Ordered Rank
Scene					GRD	1
target presence	0.53	0.11	0.32	6	spectral range	2
class type	0.14	0.00	0.07	13	target size	3
number of classes	0.28	0.00	0.14	10	SNR	4
Sensor					target type	5
GRD	0.95	1.09	1.02	1	target presence	6
SNR	0.89	0.60	0.75	4	ELM offset	7
spectral resolution	0.34	0.00	0.17	9	specified PFA	8
misregistration	0.24	0.00	0.12	11	spectral resolution	9
spectral shift	0.07	0.00	0.03	17	number of classes	10
Preprocessing					Misregistration	11
ELM gain	0.07	0.00	0.03	16	bad band inclusion	12
ELM offset	0.56	0.00	0.28	7	class type	13
bad band inclusion	0.20	0.00	0.10	12	target covariance scale	14
spectral range	0.27	1.45	0.86	2	bad data	15
bad data	0.08	0.00	0.04	15	ELM gain	16
Detection Scenario					spectral shift	17
target covariance scale	0.12	0.00	0.06	14		
target type			0.55	5		
target size	1.60	0.00	0.80	3		
specified PFA	0.54	0.00	0.27	8		

Table 7.14: Sensitivity Scores for DIRSIG Image Parameters

In Table 7.14, we consistently applied two targets to every parameter and took the average of the utility versus sensitivity plot slope as the sensitivity figure of merit. It is clear that c6 showed more sensitivity than target f4. Interestingly, there are not many commonalities in trends in the rankings in Tables 7.13 and 7.14. The closest top parameter is target type, ranked second in the

FR image and fifth in the DIRSIG image. Perhaps it is safest to not draw any conclusions from these results but to leave it as a matter for further investigation, in which the method and range of parameters considered is improved upon from this initial attempt. Figure 7.94 provides a graphical depiction of the rankings and the sensitivities categorized by color representing portion of the image chain.

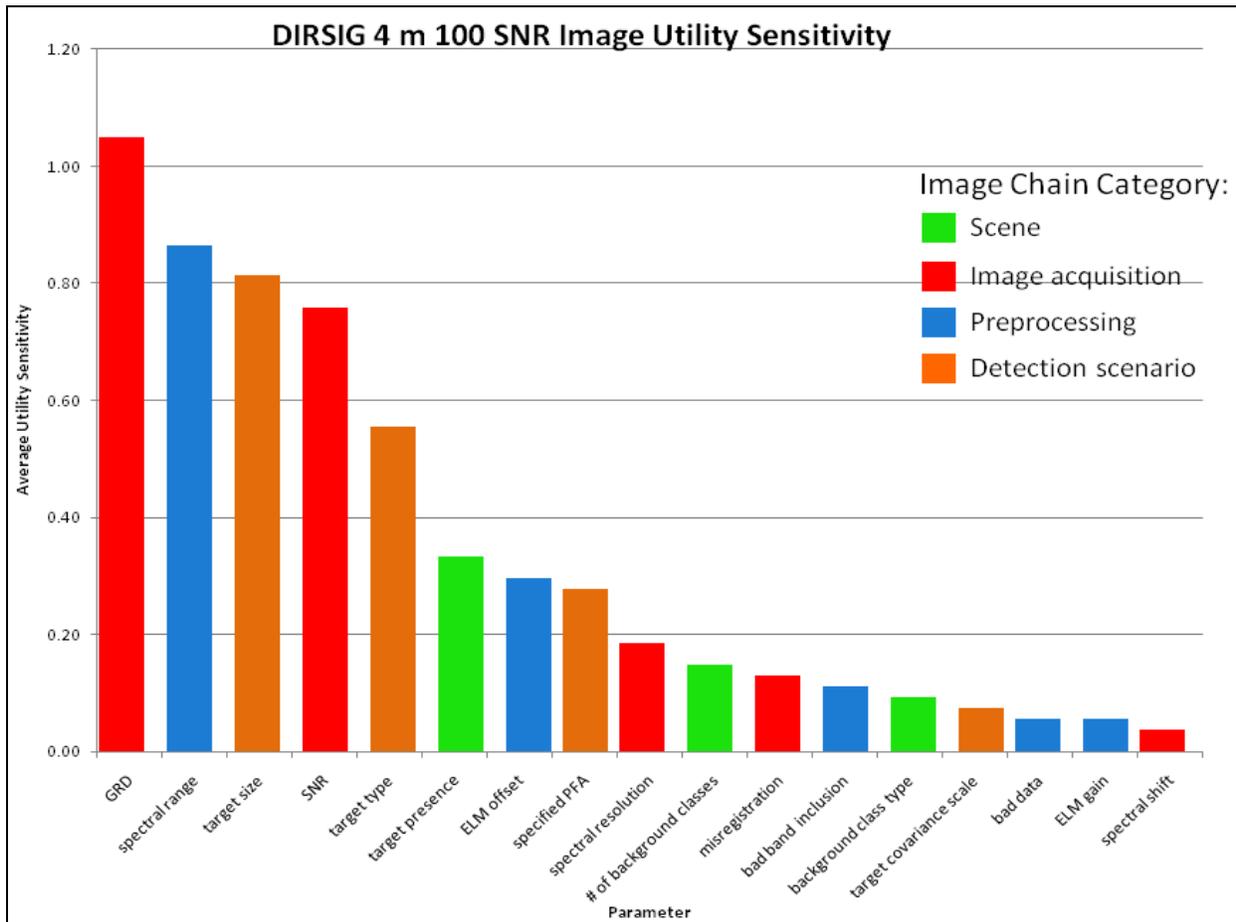


Figure 7.94: Ranking of parameters based on utility sensitivity for the DIRSIG image

7.3.6 Summary of Sensitivity Analysis

The sensitivity analysis has increased our understanding of the behavior of the utility metric. We have examined the major parameters in the scene, image acquisition, preprocessing, and target detection scenario categories of the image chain. We have applied various ad hoc techniques to measure the sensitivity of the utility to each of the parameters under investigation. There are instances in which the confidence in our determination of sensitivity is deemed low due to the difficulty of designing a representative test. We have seen the dramatic effect of the presence of target pixels in the image on the utility and need to be aware of this in actual implementation.

Chapter 7. Results

The same applies for situation in which the image has been distorted relative to the target that is being sought. The utility will report higher than it really should due to the mismatch in target and image. These effects do not mean that the utility metric is broken, simply that further investigation is required to fully understand the behaviors. This sensitivity analysis was based on two images and a handful of targets using one detection algorithm. A wider ranging study is needed before conclusions can be drawn.

Chapter 8

Conclusions

8.1 Summary

This dissertation has explored the concept of spectral image quality. It established a construct that distinguished between image fidelity and image utility on the basis of the answer to the questions, “How closely does the image capture the scene?” and “How useful is the image in extracting information of interest about a scene?” The difference between assessing the utility of an image by operating on the image itself and predicting its utility by operating on representative image parameters was highlighted. The scope of the research was narrowed to focus on image utility in the specific modality of spectral imaging and the application of subpixel target detection.

The two overarching goals of the research were to 1) establish a framework for discussing general image quality which incorporates our definitions of fidelity, utility, assessment, and prediction and is specifically applicable to spectral imagery and 2) develop a capability to assess and predict the utility of spectral images for target detection applications. Implied in the second goal is that we try to understand the behavior of our utility metric in order to allow its informed application to a wide variety of images and detection scenarios.

Towards the first goal, a review of traditional approaches to image fidelity and utility measures drawn primarily from panchromatic optical remote sensing was conducted in Chapter 3. These measures were categorized according the method in which they were obtained. In our organization of the various image quality techniques, the process of predicting utility was generalized as different models using information drawn from image fidelity measures to predict image utility measures. Chapter 4 introduced the imaging modality of spectral imaging, highlighting its characteristics in order to draw the distinction between spectral and panchromatic imagery. Spectral imagery application areas were discussed and a review of the theory and implementation of target detection algorithms was presented. The means of assessing and predicting target detection performance for target detection in spectral imagery were also discussed. Chapter 5 applied the same categorization as in Chapter 3 to spectral imagery fidelity and utility measures. It then described the means

of assessing spectral image utility and discussed various spectral image quality approaches found in the literature.

Chapter 6 described the approach to accomplishing the specific research objectives. The utility metric was described, along with the method and experiments by which spectral image utility could be assessed for different image types. The prediction of utility and methods to quantify the accuracy of the prediction relative to the assessed utility were discussed, as well as the comparison of our spectral image utility methods with the other spectral image quality approaches discussed in Chapter 5. Finally, the experiments to explore the sensitivity of the assessed utility to a number of image chain parameters were presented along with a means to quantify and rank the relative sensitivity of the utility to each parameter.

Chapter 7 discussed the results attained from the exploration of the three major aspects of spectral image utility. The first set of results emphasized the viability of our spectral image utility metric by demonstrating the behavior of utility assessed with the target implant method in simple scenarios. We found that the utility metric confirmed intuition in how the usefulness of the image for subpixel target detection behaved with parameters such as the scene, the GRD, spectral resolution, and the SNR. We demonstrated the ability of the assessed utility to be used as a means of labeling images for a specific task or set of tasks. We examined twenty images produced by six different sensors and DIRSIG to validate the utility metric and demonstrate its potential usefulness to the spectral image analyst community. Using these images, we showed the impact of sensor type, scene composition, and other factors that contribute to the ultimate assessed image utility. The second part of the results focused on the prediction of utility. Three real images produced by different sensors were employed to demonstrate how the utility could be predicted by using three basic statistical models of the filter output. The accuracy of these predictions was quantified using two metrics. The sensitivity of the accuracy of the predicted utility to prediction model parameters was explored for the three models, and the tradeoff in prediction accuracy and time savings from the predictions was discussed. This part of the research also included the comparison of our image-derived utility prediction method with the other spectral image quality methods using ten real images, leading us to a fuller understanding of the application and behavior of each of the spectral image quality methods. The third major group of results

focused on the sensitivity of assessed utility to image chain parameter variations. We introduced a simple method for quantifying the utility sensitivity and then conducted experiments involving parameters in each major part of the image chain: scene, image acquisition, preprocessing, and information exploitation. Using a real image and a simulated image, we determined the utility sensitivity of each image to variations in sixteen different image chain parameters. The results are by no means an exhaustive exploration of the subject. The primary purpose for presenting them in this dissertation was to briefly demonstrate one potential approach to establishing a tractable grip on the difficult topic of spectral image utility.

8.2 Original Contributions

This research is exploratory in its nature and empirical in its approach. Gaining traction on the issue necessitates the establishment of a larger framework, which can then be used to divide future areas of inquiry in a logical manner. Thus, our first contribution is the attempt to taxonomize approaches to image quality using the image fidelity, utility, assessment, and prediction constructs. The hope is that through such an organization of the many techniques, a robust framework through which to view image quality will emerge. We would like to think of this research as the foundation of future research that will one day establish a commonly accepted, simple, helpful metric for spectral images. The need for such a metric and a means to assess and predict it for a wide variety of images will become more important as spectral imaging matures and finds more acceptance in operational applications. We would hope that our spectral image utility metric research provides the seeds of something that will grow into something akin to NIIRS and the GIQE in the panchromatic image analyst community. Considering that decades elapsed between the implementation of NIIRS and the introduction of GIQE, the time horizon towards realizing our goal may yet be some years in the future.

This research represents a contribution to advancing the state-of-the-art in spectral imaging by providing a method to assess the utility of any spectral image for detecting any type and size of subpixel target and summarize the utility over a range of target detection scenarios that may be pertinent to a particular information exploitation task. The ability to determine the detectability of a target in an image about which we have no ground truth and

no targets in the scene is a big step forward in addressing the question of how useful an image might be for finding a particular target of a certain size. The application of this target implant methodology to the creation of an image utility metric has given us the foothold needed to label arbitrary spectral images. This provides an alternate approach to carefully planned data collects, simulated imagery, or parametric modeling in order to assess the utility of an image. The relatively easy application of this method allows a large number of images to be consistently labeled and cataloged.

The spectral image utility prediction method that we introduce, the image-derived approach, is self-contained in that the efficacy of the prediction may be directly compared to the image utility assessed by the target implant method. While predicting the utility of a notional imaging scenario is performed very nicely by the parametric analytical model, FASSP, the image-derived prediction method estimates the statistical parameters needed to perform the prediction directly from the image being evaluated, rather than a notional description of an imaging scenario. The combination of this variant of the FASSP approach and the target implant method into a self-consistent methodology for predicting and assessing image utility is an original contribution. The fact that the utility prediction is up to 40% faster than utility assessment, makes it attractive in creating a searchable image catalog that can provide responsive and pertinent information based on image analysis requirements. Fundamental exploration of the behavior of target detection algorithms with a variety of target and background combinations is facilitated by the image-derived utility prediction method.

The detailed analysis of utility sensitivity to image chain parameters contributes a new tool for grasping the impact of the many competing effects of image chain parameters on the utility of an image. Understanding of the impact of target-to-background contrast, background complexity, sensor parameters such as spatial, spectral, and radiometric resolution, and sensor artifacts on the utility of an image represents a new way to approach many inquiries about sensor or detection algorithm performance. The target-implant method serves as a test bed that allows investigation of any parameter of interest to be fully investigated. The understanding from these investigations will lead to better characterization of the behavior of our utility metrics and hopefully the effect of the parameters themselves.

8.3 Lessons Learned from Research Objectives

This research was very exploratory in its nature. As such, there was not a clear expectation of exactly what the results should look like apart from our intuition of how target detection algorithms behaved. The results were found to be very specific to image and target combination, and generalization was not always possible. In some cases we felt comfortable generalizing, particularly where intuition was confirmed. In other cases, where it was too difficult to make the call, we did not press the point too much. In each of the succeeding subsections, we will highlight the major lessons learned in each area of exploration. These lessons include confirmations of intuition, but also observations on how perhaps the experiment might have been set up differently.

8.3.1 Image Utility Framework

We learned that the task-based approach of Barrett and Myers with some modifications is the most appropriate unifying framework that allows discussion of image utility. By adding our semantic definitions of image fidelity and utility at the appropriate locations and denoting the activities of assessing and predicting image utility, we have a framework that applies to spectral image utility. We further refine the framework by adding choices pertinent to the optical remote sensing community. We also show that by viewing any prediction model of spectral image utility as a simple input-output process, the task-based approach applies to the various image quality methods in the literature. Thus, we believe that the modified task-based approach appears to be the best suited framework to address the issue of image quality based on our research.

8.3.2 Target Implant Image Utility Assessment

We demonstrated the viability of the utility metric and our means for assessing it in this part of the research. We learned that we can use the assessed utility of different images as a means of comparing the images on an equitable basis. We have given image analysts the ability to make meaningful statements such as, “Using the target-implant utility assessment method, I find that this image ranks somewhere between 0 and 1 for finding a specific target of a particular size using the spectral matched filter.” The hope is that the image analysts can communicate the usefulness of images easily in this manner. We found that the assessed

utility metric based on normalized area under the ROC curve captures a “history” of false alarm performance in the very low PFA operating areas that are important in target detection applications. We saw that the trend in assessed utility is that *for a given target size and target type*, utility goes up with increased spectral resolution, better SNR, and smaller GRD. We also saw that simpler, more homogeneous scenes yield a higher utility than complicated mixed scenes. It was also very obvious that while the assessed utility is a simply calculated metric, it expresses a complicated situation that depends on many parameters that are not easily put into an easy relationship between one parameter and the utility. We confirmed that there are many simultaneous factors that drive our assessed utility.

From the analyst’s perspective, the target detection scenario parameters are important to consider they control the assessed utility (you get what you ask for). The analyst should be aware of these factors, some of which are enumerated here. One is the type of detector used to the utility assessment, as some are designed for subpixel targets and take target variability into account, while others do not. Another is the specified PFA used to calculate the utility. A lower specified PFA will put more emphasis on the earliest false alarms. This will cause utility to be lower. The choice of specified PFA will be driven by operational requirements of the analysis task. The target type will play a major role, with dim, spectrally flat targets being more difficult and hence producing lower utilities. The target size is another critical factor that determines the utility, and we saw that this controls the separation between target absent and present distributions. The interaction of target type, size, and variability create a unique detection situation that the analyst must appreciate. Target size is also the important factor that allows us to compare images of different GRDs, allowing the utility reported for each image to be indicative of the true fraction of a subpixel target in the image. We concede that our solution to dealing with target sizes larger than GRD is suboptimal, since we are not rewarding the utility for a multipixel target, but we maintain that our approach to “flat lining” the utility for target sizes larger than GRD is the only fair way to compare different GRD image utilities given our subpixel detection task.

We also showed the value of our utility metric to the image analyst community by the ability to easily summarize many different detection scenarios into a single metric. The use of weighting functions that can be tuned by the analyst to their specific requirements allow questions to be answered in a summary sense, averaging over many target sizes, target types,

specified PFAs, and detection algorithms, or in a very specific sense, giving the utility for a specific target of a specific size, and one specified PFA with one algorithm. The flexibility that this affords the analyst in “mining” the utility space has great potential, and we only briefly illustrated the many possible applications of this idea.

8.3.3 Image-derived Image Utility Prediction

The results pertaining to prediction of the utility using the image-derived prediction approach allow us to see the value inherent in being able to accurately predict the assessed utility. We confirmed that predicting using statistical parameters is a viable way to represent how a detection algorithm that is independent of spatial location behaves. The accuracy of the prediction can be characterized using the absolute error between the predicted and assessed utilities. The methodologies are directly comparable, and we see that distributional models that best characterize the behavior of the filter output at low PFAs will track assessed utility with the least error. The exceedance metric is another way of comparing the predicted and assessed results by looking specifically at the components of the ROC curve. We apply this metric to pick the optimum T-distribution degree of freedom (DOF) based on the minimum exceedance metric over the low to specified PFA range in the PFA vectors. We also look at the exceedance metric associated with the PD vectors. We did not see a clear correlation between the exceedance metric and the utility error. But, the PFA exceedance metric is important because it gives us a method of semi-automatically selecting the optimum T-distribution DOF. In this sense, the T-distribution DOF that controls the heaviness of the “tail” of the PFA vector is a good indicator of the image difficulty. The heavy tail will correspond to the situation that has significant early false alarms whereas the light tail is the Gaussian situation, which oversimplifies the complexities of spectral imagery. Thus, just looking at the PFA and PD vectors explains why the T-distribution does a better job at tracking the utility than normal or Sum of Gaussian class distributions. The composite PD vector combines the heavy tail of the T-distribution with the realization that the target is assumed to be coming from a normal distribution, so mixing them linearly improves the accuracy. The three images we use to look at prediction represent a range of high, low, and mid range utilities. We see that the T-distribution and composite models produce the smallest

utility error and summarize this with the mean absolute distance from the perfect match line on the scatter plot of assessed and predicted utilities.

We confirmed that prediction utility is faster than assessing it. Key factors that contribute to the 40% time differential are the time to implant target and run the filter on each pixel for the assessed utility, the time to estimate global parameters is the same for both assessed and predicted, and the time to estimate the T-DOF is a major factor in the prediction approach, since our selection of the optimum DOF requires that we run a target absent case using the target implant method. We note that the time for estimating the optimal T-distribution DOF can be reduced by decreasing the sample support used to estimate the global image parameters at apparently minimal sacrifice of prediction accuracy.

Our investigation of the sensitivity of prediction accuracy to prediction parameters led us to conclude that the utility error was significantly sensitive to the T-DOF chosen, and that a clear optimal DOF is associated with minimizing the utility error. The number of classes used for sum of Gaussian does not significantly impact the utility error. The sample support used for the normal distribution model likewise does not appear to significantly impact the utility error. These last two observations are primarily due to the fact that the sum of Gaussian and normal models are always the worst performers in terms of utility error.

Comparison of our prediction methods (and assessment as a reference) with other spectral image quality approaches led to a mixed bag of results. No clear trends manifested themselves in terms of various techniques behaving in a certain manner. Part of the difficulty in interpreting the results is that implementing each of the methods required some ad hoc approaches and use of disparate input parameters. This makes comparison of the approaches difficult. In general, though, it appears that our spectral image utility, which is defined for a specific target detection situation, uses a wider range of the possible utility values (0 to 1). Most importantly, the details of implementing each of the spectral image quality approaches deepened our understanding of the requirements and characteristics of each method.

8.3.4 Image Utility Sensitivity to Image Chain Parameters

We learned a significant amount from the investigation of image chain parameters. Perhaps the most important lesson that we learned was that our utility metric has certain situations in which it responds in a manner not indicative of the true detection situation, but rather a

vestige of the design of the target implant method. We call this observed phenomenon target-image mismatch due to other-than-natural causes. The result is that the implanted target looks very different from the image we are injecting it into. This will make an easier detection situation, and the utility will be higher than it really should. This occurred with our investigation of atmospheric compensation, where FLAASH and ATREM spectra were clearly not producing the best reflectances yet were assessing higher image utilities. Although not discussed in the body of the dissertation, the application of empirical atmospheric compensation methods to an image for which calibration panels are not in the scene can lead to the same effect. We also see this effect when we inaccurately characterize the sensor spectral response characterization or when the bad bands list applied to the target is not exactly the same as that applied to the image, creating a spectral shift effect. We see a depression of utility in a situation which is caused by a natural effect – targets of interest are in the scene. If there are significant target pixels of the target we are trying to match in the image, then it will drive up false alarms, which will hurt utility. This is the double-edged sword of this utility metric being very sensitive to the early false alarms

Having established the areas in which the metric behaves in a non-intuitive fashion, we were able to conduct experiments that explored each parameter independently to note their effect on the assessed utility. We devised a method to examine the effect on utility of perturbing one parameter while holding all others constant. Though not always achievable in practice, we cover a wide range of parameters in the image chain. Some are more realistic than others, since we only have artificial control over the scene and sensor parameters. Others very directly relate to the choices that an image analyst would make in processing the image. We look at the parameter, try to pick a nominal operating range, and then call the middle of that range the baseline. We then normalize that parameter with respect to the maximum value so that we can equitably compare all parameters. We then find the instantaneous slope of the utility about this point with its two neighbors on either side. This is equivalent to noting the effect of a slight perturbation of the parameter. The steepness of that plot will tell us how sensitive the utility is to this parameter. One drawback is that it is very sensitive to where we are on the utility curve (in other words, if we have either a very easy or very difficult detection scenario, utility will either be close to 1 or close to 0 and the slope will be very small). If we are stuck in one of the extremes, chances are that small changes in

the parameter are not going to make any significant difference in the utility. A priori, we do not know where the most sensitive operating point will be.

It is difficult to make conclusions when only using a few images, and we should not try to conclude too much from this before looking at more images, but it looks like specified PFA, target type, and compensation method are the top three parameters to which utility is most sensitive. Diagnostics for explaining behavior include spatial orientation of the early false alarms, spectral patterns in the early false alarm pixel vectors, target absent and present distributions and relative areas under them up to the specified PFA number of pixels. We learned that target type is important because it controls the relative width of both target present and absent distributions. An easy target built into the filter results in tight target absent and present distributions. Scene spectral composition has an impact that is always related to the specific target. Although we do not quantify the sensitivity to algorithm type, this is a significant factor in the utility. We have not looked at enough images or algorithms to draw conclusion as to what type of algorithm will do consistently best or worst. We already discussed target presence and its importance in the previous paragraph, but we note that this will hurt the utility in two ways: the target pixels will cause false alarms and will contaminate the estimate of the covariance matrix. Slowly adding target of interest pixels back into a target free image has a dramatic impact on the assessed utility, whereas when there are interfering signatures, the effect is not very significant.

The spatial resolution effectively changes the implant fraction when we change GRD and keep target size constant. Larger GRD will always do worse because it is like working at a very small fraction. Sensor SNR confirms intuition that lower SNR leads to lower utility. The spectral resolution with compensation for constant SNR adjustment results in a smaller-valued covariance matrix, which means larger inverse covariances. This implies larger filter weights, which may be one factor in creating more variability in the filter output, translating to more overlap in TA and TP distributions, and lower utility. Misregistration causes more false alarm pixels to have larger values. False alarms are occurring next to neighbors. Spectral shift does not have a clear story relating to utility, although the general trend is to degrade utility with more shift.

The perturbation of ELM coefficients shows that utility is relatively insensitive to these parameters, but not in a consistent manner. Also, the type of atmospheric compensation

does not appear to have a significant impact. The utility seems relatively insensitive to the number of bad bands. We expect utility to go up with increased spectral range, which it appears to do.

8.4 Future Work

The nature of the topic necessitated a very focused topic area. A natural division of future work tasks seemed to be those that were obviously within the defined narrow scope and those that were not.

8.4.1 Within Current Research Scope

In the category of assessing image utility and demonstrating the viability of the utility metric, we believe that an important and helpful addition would be to find a way to deal with different GRD images on the same plot more gracefully, rather than employing the “flat line” approach for sizes larger than GRD. We have started to examine a way of “rewarding” the utility for multipixel detection situations by applying a spatial confidence figure of merit based on the Johnson criteria found in the semantic transform of GSUM. We would like to assess the utility of images of the same scene made by different spatially and temporally co-registered sensors. This would facilitate a real sensor-to-sensor comparison using our spectral utility metric. Likewise, we would like to investigate more images of the same scene by the same sensor at different times. All sensor parameters would remain the same, but the atmosphere and illumination conditions would be different. Finally, we would like to use images of many different scenes in the same vicinity (so that atmosphere isn’t too different) at very close to the same time. This would isolate the effect on utility specifically due to scene composition.

For research pertaining to predicting the utility, we would recommend more exploration of possible models for describing the output, beyond the basic four that we have proposed. We would like to refine the selection of the optimal T-DOF without needing to assess the utility. A more deliberate comparison of spectral image quality methods would be a very rich area of research, and the lessons gleaned from such an endeavor would undoubtedly help in the formulation of a more robust metric or extension of existing ones.

In those experiments pertaining to characterizing the utility sensitivity to image chain parameters, we believe that further investigation is required in several areas: understanding and mitigating situations in which the target does not match the image due to image artifacts (spectral distortions), the effect of target presence in the scene on utility, scene spectral composition (more realistic experiments), radiometric calibration, more atmospheric and image collection parameters. This last category would be a good one to work with DIRSIG to generate a many cases. We also recommend a continued search for a more consistent and robust methodology to characterize the sensitivity that does not depend on the operating point as much as the current method. The sensitivity analysis should be conducted with more images to see what kind of trends emerge. We should look at the effect of using targets from other than the FASSP reference library, and we must account for target variability. In general, we would like to explore if there is a difference between using image-derived target statistics and using those captured by a hand-held spectrometer? Finally, we would like a way to look at the non-linear pixels (shadow) and their effect as they are added to the image and use another means of perturbing ELM parameters that does not rely on the calibration panels in the scene.

8.4.2 Beyond Current Scope

In looking beyond the current scope of research, we would to devise a means of assigning a confidence in our utility assessments and predictions and alert the analyst when the utility metric is not trustworthy. We would extend this approach to other application areas like classification, unmixing, and anomaly detection, to include detectors that are spatially dependent. In order to make this more accessible to the image analyst community, we would make the current assessment methodology into a web-based utility to allow users to “test” the utility of any image. For extensions to prediction, we would like to extend the prediction to nonlinear detectors in order to increase the applicability of prediction approach. We would also search for other parameters that might not necessarily have to be derived from the image in order to make the prediction.

References

- [1] Adams, J. B. and A. R. Gillespie (2006). *Remote Sensing of Landscapes with Spectral Images: A Physical Modeling Approach*, Cambridge University Press.
- [2] Aiazzi, B., L. Alparone, A. Barducci, S. Baronti, and I. Pippi (2001). "Information-Theoretic Assessment of Sampled Hyperspectral Imagers," *IEEE Transactions on Geoscience and Remote Sensing*, v39(7), pp. 1447-1458.
- [3] Alder-Golden, S. M., A. Brek, L. S. Bernstein, S. Richtsmeier, P. K. Acharya, M. W. Matthew, G. P. Anderson, C. Allred, L. Jeong, and J. Chetwynd (1998). "FLAASH, A MODTRAN4 Atmospheric Correction Package for Hyperspectral Data Retrievals and Simulations," *Proceedings of the 7th Annual JPL Airborne Earth Science Workshop*, Pasadena, CA, pp. 9-14.
- [4] Avcibas, I., B. Sankur, and K. Sayood (2002). "Statistical evaluation of image quality measures," *Journal of Electronic Imaging*, v11(2), pp. 206-223.
- [5] Barrett, H. H., J. Yao, J. P. Rolland, and K. J. Myers (1993). "Model observers for assessment of image quality," *Proceedings of the National Academy of Science*, v90, pp. 9758-9765.
- [6] Barrett, H. H. and K. J. Myers (2004) *Foundations of Image Science*, John Wiley and Sons Inc., Hoboken, NJ.
- [7] Barten, P. G. J. (1990). "Evaluation of subjective image quality with the square-root integral method." *Journal of the Optical Society of America*, v7(10), pp. 2024-2031.
- [8] Beaven, S. G., A. D. Stocker, and P. V. Villeneuve (2003). "Target Detection using In-Scene Illumination Compensation," *Proc. Military Sensing Symposium -- CC&D Conference*.
- [9] Bernstein, L. S., S. M. Adler-Golden, R. L. Sundberg, R. Y. Levine, T. C. Perkins, A. Berk, A. J. Ratkowski and M. L. Hoke (2004). "A New Method for Atmospheric Correction and Aerosol Optical Property Retrieval for Vis-SWIR Multi- and Hyperspectral Imaging Sensors: QUAC (QUick Atmospheric Correction)," *Proc. 13th JPL Airborne Earth Science Workshop*, pp. 1-12.
- [10] Biberman, L. M. (1973). "Image Quality," in *Perception of Displayed Information*, ed by Lucien M. Biberman, Plenum Press, NY.
- [11] Bicknell, W. E., C. J. Digenis, S. E. Forman, D. E. Lencioni (1999). "EO-1 Advanced Land Imager", *SPIE Conference on Earth Observing Systems IV*.

References

- [12] Cafer, C. E., S. R. Rotman, J. Silverman, and P. W. Yip (2002). "Algorithms for point target detection in hyperspectral imagery," *SPIE Proceedings of Imaging Spectrometry VIII*, v4816, pp. 242-257.
- [13] Cafer, C. E., M. Stefanou, E. Nielsen, A. Rizzuto, O. Raviv, and S. Rotman (2007). "Analysis of false alarm distributions in the development and evaluation of hyperspectral point target detection algorithms," *Optical Engineering*, v46(7), pp. 076402-1-15.
- [14] Chang, C., J. Liu, B. Chien, H. Ren, C. Wang, C. Lo, P. Chung, C. Yang, D. Ma (2000) "Generalized constrained energy minimization approach to subpixel target detection for multispectral imagery," *Optical Engineering*, v39(5), pp. 1275-1281.
- [14] Chang, C., D. C. Heinz (2000). "Constrained Subpixel Target Detection for Remotely Sensed Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, v38(3), pp. 1144-1159.
- [15] Christophe, E., D. Leger, C. Mailhes (2005). "Comparison and Evaluation of Quality Criteria for Hyperspectral Imagery," *IS&T/SPIE 17th Annual Symposium on Electronic Imaging*.
- [16] Cocks, T., R. Jenssen, A. Stewart, I. Wilson and T. Shields (1998). "The HyMap Airborne Hyperspectral Sensor: The System, Calibration and Performance," *1st EARSEL Workshop on Imaging Spectroscopy*, pp. 1-7.
- [17] Damera-Venkata, N., T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik (2000). "Image Quality Assessment Based on a Degradation Model," *IEEE Transactions on Image Processing*, v9(4), pp 636-650.
- [18] Descour, M. and E. Dereniak (1995). "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results," *Applied Optics*, v34(22), pp. 4817-4826.
- [19] Digital Globe (2008) <http://www.digitalglobe.com/index.php/85/QuickBird>
- [20] Dosselmann, R. and X. D. Yang (2005). "Existing and Emerging Image Quality Metrics," *IEEE Canadian Conference Proceedings*, pp. 1906-1913.
- [21] Driggers, R., P. Cox, J. Leachtenauer, R. Vollmerhausen, and S. Scribner (1998). "Targeting and Intelligence Electro-Optic Recognition Modeling: A Juxtaposition of the Probabilities of Discrimination and the General Image Quality Equation," *Optical Engineering*, v37(3), pp. 789-797.
- [22] Farrand, W. H. and J. C. Harsanyi (1997). "Mapping the Distribution of Mine Tailings in the Coeur d'Alene River Valley, Idaho, through the Use of a Constrained Energy Minimization Technique," *Remote Sensing of the Environment*, v59 pp. 64-76.

References

- [23] Feng, X. (1995). *Design and Performance of a Modular Imaging Spectrometer Instrument*, Rochester Institute of Technology Ph.D. Dissertation.
- [24] Fiete, R. D. (1999). "Image quality and λ FN/p for remote sensing systems," *Optical Engineering*, v38(7), pp. 1229-1240.
- [25] Fiete, R. D. and T. A. Tantalo (1999). "Comparison of SNR image quality metrics for remote sensing systems," *Optical Engineering*, v38(5), pp. 815-820.
- [26] Gao, B.-C., K. B. Heidebrecht, and A. F. H. Goetz (1993). "Derivation of Scaled Surface Reflectances from AVIRIS Data," *Remote Sensing of the Environment*, v44(2/3), pp. 165-177.
- [27] Grangery, E. M. and K. N. Cupery (1972). "An optical merit function (SQF), which correlates with subjective image judgments," *Photographic Science and Engineering*, v16(3), pp. 221- 230.
- [28] Hailstone, R (2005). *Basic Principles of Imaging Science I Course Notes*, Rochester, NY.
- [29] Holst, G. C. (2006). *Electro-Optical Imaging System Performance*, 4th ed, SPIE Optical Engineering Press, Bellingham, WA.
- [30] Huck, F. O., C. L. Fales, R. Alter-Gartenberg, and Z. Rahman (1992). "Information, Entropy and Fidelity in Visual Communication," *SPIE Proceedings on Visual Information Processing*, v705, pp. 145-154.
- [31] Huck, F. O., C. L. Fales, and Z. Rahman (1997). *Visual Communication: An Information Theory Approach*, Kluwer Academic Publishers, Boston, MA.
- [32] Huck, F. O., C. L. Fales, R. Alter-Gartenberg, S. K. Park, and Z. Rahman (1999). "Information theoretic assessment of sampled imaging systems," *Optical Engineering*, v38(5), pp. 742-762,
- [33] Image Resolution Assessment and Reporting Standards (IRARS) Committee (1995). *Multispectral Imagery Interpretability Rating Scale Reference Guide*, Washington, DC.
- [34] Irvine, J. M. (1997). "National Imagery Interpretability Rating Scales (NIIRS): Overview and Methodology," *SPIE Proceedings on Airborne Reconnaissance XXI*, v3128, pp. 93-103.
- [35] Janssen, R. (2001). *Computational Image Quality*, SPIE Press, Bellingham, WA.
- [36] Kanungo, T., M. Y. Jaisimha, J. Palmer, and R. M. Haralick (1995). "A Methodology for Quantitative Performance Evaluation of Detection Algorithms," *IEEE Transactions on Image Processing*, v4(12), pp. 1667-1674.

References

- [37] Kerekes, J. P. and D. A. Landgrebe (1991). "An Analytical Model of Earth-Observational Remote Sensing Systems," *IEEE Transactions on Systems, Man, and Cybernetics*, v21(1), pp. 125-133.
- [38] Kerekes, J. P. and J. E. Baum (2002). "Spectral Imaging System Analytical Model for Subpixel Object Detection," *IEEE Transactions on Geoscience and Remote Sensing*, v45(5).
- [39] Kerekes, J. P. (2004). *Model-Based Analysis of Sub-pixel Target Detection Sensitivity of Reflective Hyperspectral Imaging Systems (HTAP-20)*, Lincoln Laboratory, Lexington, MA.
- [40] Kerekes, J. P. and S. M. Hsu (2004a). "Spectral Quality Metrics for Terrain Classification," *Proceedings of SPIE on Imaging Spectrometry X*, v5546, pp. 382-389.
- [41] Kerekes, J. P. and S. M. Hsu (2004b). "Spectral Quality Metrics for VNIR and SWIR Hyperspectral Imagery," *Proceedings of SPIE on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, v5425, pp. 549-557.
- [42] Kerekes, J. P., A. P. Cisz, and R. E. Simmons (2005). "A Comparative Evaluation of Spectral Quality Metrics for Hyperspectral Imagery," *Proceedings of SPIE on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, v5806, pp. 469-480.
- [43] Kerekes, J. P. (2006). "Model-based Exploration of HSI Spaceborne Sensor Requirements with Application Performance as the Metric," *Proceedings of the 2006 International Geoscience and Remote Sensing Symposium*.
- [44] Kerekes, J. P., D. Messinger, P. Lee and R. Simmons (2006), "Comparisons between spectral quality metrics and analyst performance in hyperspectral target detection," *Proceedings of SPIE on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*, v6233, pp. 62330W-1-9.
- [45] J.P. Kerekes (2008). "Receiver Operating Characteristic Curve Confidence Intervals and Regions," *IEEE Geoscience and Remote Sensing Letters*, v5(2) , pp. 251-255.
- [46] Keshava, N. (2003). "A Survey of Spectral Unmixing Algorithms," *Lincoln Laboratory Journal*, v14(1), pp. 55-78.
- [47] Landgrebe, D. (2002). "Hyperspectral Image Data Analysis," *IEEE Signal Processing Magazine*, v19(1), pp17-28.
- [48] Landgrebe, D. A. (2003). *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley and Sons, Hoboken, NJ.

References

- [49] Leachtenauer, J. C., W. Malila, J. Irvine, L. Colburn, and N. Salvaggio (1997). "General Image-Quality Equation; GIQE," *Applied Optics*, v36(32), pp. 8322-8328.
- [50] Leachtenauer, J. C., and R. G. Driggers (2001). *Surveillance and Reconnaissance Imaging Systems: Modeling and Performance Prediction*, Artech House, Boston, MA.
- [51] Manolakis, D., C. Siracusa, D. Marden, and G. Shaw (2001). "Hyperspectral Adaptive Matched Filter Detectors: Practical Performance Comparison," *SPIE Proceedings on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VII*, v4381, pp. 18-33.
- [52] Manolakis, D., D. Marden, J. Kerekes, and G. Shaw (2001). "On the Statistics of Hyperspectral Imaging Data," *SPIE Proceedings on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VII*, v4381, pp. 308-316.
- [53] Manolakis, D., and G. Shaw (2002). "Detection Algorithms for Hyperspectral Imaging Applications," *IEEE Signal Processing Magazine*, v19(1), pp.29-43.
- [54] Manolakis, D., D. Marden, and G. Shaw (2003). "Hyperspectral Image Processing for Automatic Target Detection Applications," *MIT Lincoln Laboratory Journal*, v14(1), pp.79-116.
- [55] Manolakis, D. (2005). "Taxonomy of detection algorithms for hyperspectral imaging applications." *Optical Engineering*, v44(6), pp. 066403-1-11.
- [56] Marden D. and D. Manolakis (2004). "Statistical modeling of hyperspectral imaging data and their applications," *Tech. Rep. HTAP-16*, Lincoln Laboratory, MIT, Lexington, MA.
- [57] Martin, L., J. Vrabel. and J. Leachtenauer (1999). "Metrics for Assessment of Hyperspectral Image Quality and Utility," *Proceedings of the International Symposium on Spectral Sensing Research*, Las Vegas, NV.
- [58] Massachusetts Institute of Technology Lincoln Laboratory (2004). *Canonic Data Set Documentation*.
- [59] Metz, C. E. (1986). "ROC methodology in radiologic imaging, *Investigative Radiology*, v21, pp. 720-733.
- [60] Miyahara, M., K. Kotani, and V. R. Algazi (1998). "Objective Picture Quality Scale (PQS) for Image Coding," *IEEE Transactions on Communications*, v46(9), pp. 1215-1226.
- [61] Nill, N. B. and B. H. Bouzas (1992). "Objective image quality measure derived from digital image power spectra," *Optical Engineering*, v31(4), pp. 813-825.

References

- [62] O'Sullivan, J. A., R. E. Blahut, and D. L. Snyder (1998). "Information-Theoretic Image Formation," *IEEE Transactions on Information Theory*, v44(6), pp. 2094-2123.
- [63] Richards, J. A. and X. Jia (1999). *Remote Sensing Digital Image Analysis: An Introduction*, 3rd Ed, Springer-Verlag, Berlin, Germany.
- [64] Riehl, K. and L. Maver (1996). "A comparison of two common aerial reconnaissance image quality measures," *SPIE Proceedings on Airborne Reconnaissance XX*, v2829, pp. 242-254.
- [65] Schaum, A. and A. Stocker (2004). "Advanced Algorithms for Autonomous Hyperspectral Change Detection," *Proceedings of the 33rd Applied Imagery Pattern Recognition Workshop*.
- [66] Schott, J.R., S. D. Brown, R. V. Raqueno, H. N. Gross, and G. Robinson (1999). "An Advanced Synthetic Image Generation Model and Its Application to Multi/Hyperspectral Algorithm Development", *Canadian Journal of Remote Sensing*, v25(2).
- [67] Schott, J. R. (2007). *Remote Sensing: The Image Chain Approach*, 2nd ed, Oxford University Press, NY.
- [68] Schwartz, C. R., A. C. Kenton, W. F. Pont, and B. J. Thelen (1995). "Statistical parametric signature/sensor/detection model for multispectral mine target detection," *SPIE Proceedings on Detection Technologies for Mines and Minelike Targets*, v2496, pp. 222-238.
- [69] Shaw, G. A., and H. K. Burke (2003). "Spectral Imaging for Remote Sensing", *MIT Lincoln Laboratory Journal*, v14(1), pp. 3-28.
- [70] Shen, S. (2003). "Spectral Quality Equation Relating Collection Parameters to Object/Anomaly Detection Performance," *SPIE Proceedings on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, v5093, pp. 29-36.
- [71] Simmons, R., T. Elder, D. Stewart, E. Cincotta, C. Kennedy, and R. Van Nostrand (2005). "General Spectral Utility Metric for Spectral Imagery," *SPIE Proceedings on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, v5806, pp. 457-468.
- [72] Stein, D. W. J., S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker (2002). "Anomaly Detection from Hyperspectral Imagery," *IEEE Signal Processing Magazine*, pp.58-69.

References

- [73] Sun, Q. and M. D. Fairchild (2003). "Image Quality for Visible Spectral Imaging," *Proceedings of PICS Conference*, Rochester, NY, pp. 210-214.
- [74] Sweet, J., J. Granahan, and M. Sharp, M. (2000). "An Objective Standard for Hyperspectral Image Quality," *Proceedings of AVIRIS Workshop*.
- [75] Sweet, J. (1999). "An Objective Standard for Hyperspectral Image Quality," *Proceedings of the International Symposium on Spectral Sensing Research*.
- [76] Van Trees, H. L. (2001). *Detection, Estimation, and Modulation Theory Part I*, John Wiley and Sons, NY.
- [77] Vane, G., R. O. Green, T. G. Chrien, H. T. Enmark, E. G., and W. M. Porter (1993). "The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)" *Remote Sensing of Environment*, v44(2-3), pp. 127-143.
- [78] Yuhas, R. H., A. F. H. Goetz, and J. W. Boardman (1992). "Discriminating Among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm," *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, v1, pp. 147-149.