**Michigan
Technological
University**

Michigan Technological University
Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's Reports

2018

Target detection, tracking, and localization using multi-spectral image fusion and RF Doppler differentials

Casey D. Demars

Michigan Technological University, cddemars@mtu.edu

Copyright 2018 Casey D. Demars

Recommended Citation

Demars, Casey D., "Target detection, tracking, and localization using multi-spectral image fusion and RF Doppler differentials", Open Access Dissertation, Michigan Technological University, 2018.
<https://doi.org/10.37099/mtu.dc.etdr/720>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etdr>

Part of the [Signal Processing Commons](#)

TARGET DETECTION, TRACKING, AND LOCALIZATION USING
MULTI-SPECTRAL IMAGE FUSION AND RF DOPPLER DIFFERENTIALS

By

Casey D. Demars

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Electrical Engineering

MICHIGAN TECHNOLOGICAL UNIVERSITY

2018

© 2018 Casey D. Demars

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Electrical Engineering.

Department of Electrical and Computer Engineering

Dissertation Advisor: *Dr. Michael C. Roggemann*

Committee Member: *Dr. Laura Brown*

Committee Member: *Dr. Timothy Havens*

Committee Member: *Dr. Christopher Middlebrook*

Department Chair: *Dr. Daniel Fuhrmann*

Contents

List of Figures	ix
List of Tables	xvii
Preface	xix
Acknowledgments	xxi
List of Abbreviations	xxiii
Abstract	xxv
1 Introduction	1
1.1 Motivation	1
1.2 Approach	7
1.3 Summary of Key Results	9
1.4 Organization	11
2 Multi-Spectral Detection and Tracking of Multiple Moving Tar-	
gets in Cluttered Urban Environments	13

2.1	Abstract	13
2.2	Introduction	14
2.3	Detection and Segmentation Algorithm	21
2.4	Target Tracking	29
2.5	Experiment	33
2.5.1	Detection Results	34
2.5.2	Tracking	41
2.6	Conclusion	44
3	Target Localization and Tracking by Fusing Doppler Differentials from Cellular Emanations with a Multi-spectral Video Tracker .	45
3.1	Abstract	45
3.2	Introduction	46
3.3	Cell Phone Emanations	53
3.4	Multi-spectral Video Tracker	61
3.5	Data Association	64
3.6	Experiments	66
3.7	Conclusion	73
4	Multi-spectral image shift-estimation error calculations using sim- ulated phenomenology	75
4.1	Abstract	75
4.2	Introduction	76

4.3	Image Registration Algorithms	81
4.3.1	Area-based	81
4.3.2	Feature-based	83
4.4	Fisher Information Error	85
4.4.1	Cramer-Rao Lower Bound	85
4.4.2	Shift-estimation Error for Multi-spectral Imagery	88
4.5	Evaluation Metrics	91
4.5.1	Tie-point Shift Estimation Error	91
4.5.2	Bias Gradient	92
4.5.3	Cumulative Distribution Function	94
4.6	Synthetic Multi-spectral Imagery	95
4.7	Numerical Calculations of Error Metrics	101
4.7.1	Area-based Registration Algorithm Errors Maps	105
4.7.1.1	DFT	105
4.7.1.2	ECC	107
4.7.1.3	MI	108
4.7.1.4	Area-based Algorithm Comparison Using the CDF	109
4.7.2	Feature-based Shift-estimation Error Maps	111
4.7.2.1	Lowe	112
4.7.2.2	EHD	113
4.7.2.3	PCEHD	114

4.7.2.4	LGHD	115
4.7.2.5	Feature-based Algorithm Comparison Using the CDF	115
4.7.3	Summary of Algorithm Performance for Selected Sub-fields .	117
4.8	Conclusion	118
5	Conclusion	123
5.1	Summary of Key Results	124
5.2	Suggested Future Work	126
	References	129
A	Letters of Permission	143
A.1	SPIE letter of permission	143
A.2	OSA letter of permission	144

List of Figures

2.1	Flowchart of the detection and tracking algorithm using fusion of of multi-spectral imagery and SIFT features.	19
2.2	Example multi-spectral frames from synthetic DIRSIG data set. (a) - visible, (b) - NIR, (c) - MWIR, (d) - LWIR	20
2.3	Example foreground images detected using the GMM model at frame 600. (a) - VIS, (b) - NIR, (c) - MWIR, (d) - LWIR	26
2.4	Example fused foreground maps using all spectral bands (TOT) for a single frame. (a) is the fused foreground maps, $FG_{FUS}(x, y)$. (b) is the foreground map after applying a Gaussian Filter. (c) shows the foreground after a threshold has been applied to the filtered image.	28
2.5	Zoomed area highlighting the algorithm for fusing the foreground pixels from multi-spectral imagery.	28
2.6	Detection and false alarm rates for detecting moving targets using the labeled spectral combinations in the fusion algorithm.	37

2.7	Detection rate of scene targets using segmented objects from the foreground detection algorithm or features in the instance that no foreground object was detected.	38
2.8	Example showing how detected pixels are scored as true positives or false alarms. (a) is an example vehicle in MWIR. (b) illustrates detected foreground pixels; blue indicates true positive and red indicates false alarm.	39
2.9	Detection rates of pixels that belong to targets vs. rate of false pixels that do not belong to any targets.	40
2.10	Mean centroid displacement of detected targets for all detected targets for all frames.	41
2.11	Examples frames from the tracking sequence (Video 1, MPEG, 14 MB). (a) - Frame - 150, (b) Frame - 300, (c) Frame - 450, (d) Frame - 600	43
3.1	Geometry of the multi-spectral and RF sensors in reference to the plane in which the moving vehicles lie.	50
3.2	Coordinates of the multi-spectral and RF sensors in reference to the plane in which the moving vehicles lie.	50
3.3	Flowchart of the fusion processing to associate a specific moving target between a multi-spectral video tracker and a constellation of RF sensors.	51

3.4	Flowchart of RF sensor processing to isolate and extract Doppler shifts.	52
3.5	Range of the moving ground targets from the video sensor for the given video sequence.	54
3.6	Range-rate of the moving ground targets from the video sensor for the given video sequence.	56
3.7	Doppler Shift at 1 GHz carrier frequency for the moving ground targets as measured from location of the video sensor for the given video sequence.	57
3.8	Example ranges between a transmitting target (GT #1) and the locations of the RF sensor constellation.	58
3.9	Example range-rate between a transmitting target (GT #1) and the locations of the RF sensor constellation.	59
3.10	Example Doppler shifts for 1 GHz carrier frequency between a transmitting target (GT #1) and the locations of the RF sensor constellation.	59
3.11	Maximum Doppler differential for GT #1 using all RF sensor combinations (left Y-axis) vs. its corresponding ground speed (right Y-axis).	60
3.12	Example multi-spectral frames from synthetic DIRSIG data set for (a) Visible, (b) NIR, (c) MWIR, and (d) LWIR spectral bands.	62

3.13	RMSE of the centroid estimate using the multi-spectral video tracker for the ground-truth video targets.	64
3.14	The process of associating the DD at a given frame measured from a constellation of RF sensors with the output of a multi-spectral video tracker using two statistical metrics; 1) absolute difference and 2) root- mean-square difference.	66
3.15	Correct association rate for matching specific cellular emanation mea- sured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the absolute DD.	68
3.16	Correct association rate for matching specific cellular emanation mea- sured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the absolute DD.	68
3.17	Correct association rate for matching specific cellular emanation mea- sured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the RMSD DD.	70
3.18	Correct association rate for matching specific cellular emanation mea- sured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the RMSD DD.	70
3.19	Confidence ratio for matching specific cellular emanation measured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the RMSD DD.	72

3.20	Confidence ratio for matching specific cellular emanation measured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the RMSD DD.	73
4.1	Process of converting DIRSIG radiance maps to imagery using the ODEM.	98
4.2	Synthetically generated imagery of an airfield set of (a) VIS and (b) LWIR bands. Images are 512×512 pixels with a ground sample distance of 0.056 meters. For reference, the image color scale in digital units is shown in the upper-left and a 32×32 sub-field is drawn in the lower-right of each image.	100
4.3	DIRSIG radiance map in (a) VIS and (b) LWIR displaying discrete thermal scarring as a result of the temporal sampling in the THERM solver. Note these radiance maps are free of an optical or detector effects to enable thermal scarring to be visualized.	100
4.4	Process of generating error maps from shift-estimation error metrics for multi-spectral imagery. For each pixel, sub-fields surrounding the pixel are extracted from the image pair, and error metrics are calculated from an ensemble of noise-corrupted sub-field pairs. The resulting metric is inserted into the error map at the pixel location.	102

4.5	RT of shift-estimate UFIE for the VIS (a), LWIR (b), and VIS registered to LWIR (c) sub-fields centered at the indicated field angles. In (c), black pixels indicate that the associated FIM has negative elements.	104
4.6	Samples VIS-LWIR (a-c) and VIS-VIS (d-f) DFT algorithm shift-estimation errors for 32×32 -pixel sub-fields centered at the indicated field angles.	106
4.7	Sample VIS-LWIR (a-c) and VIS-VIS (d-f) ECC algorithm shift-estimation errors for 32×32 -pixel sub-fields centered at the indicated field angles.	108
4.8	Sample VIS-LWIR (a-c) and VIS-VIS (d-f) MI algorithm shift-estimation errors for 32×32 -pixel sub-fields centered at the indicated field angles.	109
4.9	CDF plot of the RMSE for shift-estimation of 32×32 -pixel sub-fields using area-based algorithms.	110
4.10	SIFT keypoints VIS (left) and LWIR (right) images with lines connecting matched features using the (a) Lowe, (b) EHD, (c) PCEHD, and (d) LGHD descriptors. Blue lines connecting features indicate an error displacement of less than 5 pixels and red lines connect features indicate an error displacement greater than 5 pixels.	112

4.11	Sample shift-estimates errors using the highest weighted feature match from the Lowe descriptor for a 32×32 -pixel sub-field centered at the indicated field angle. “Black” pixels indicate sub-fields where a minimum of 200 feature point were not matched.	113
4.12	Sample shift-estimates errors using the highest weighted feature match from the EHD descriptor for a 32×32 -pixel sub-field centered at the indicated field angle. “Black” pixels indicate sub-fields where a minimum of 200 feature point were not matched.	114
4.13	Sample shift-estimates errors using the highest weighted feature match from the PCEHD descriptor for a 32×32 -pixel sub-field centered at the indicated field angle. “Black” pixels indicate sub-fields where a minimum of 200 feature point were not matched.	114
4.14	Sample shift-estimates errors using the highest weighted feature match from the LGHD descriptor for a 32×32 -pixel sub-field centered at the indicated field angle. “Black” pixels indicate sub-fields where a minimum of 200 feature point were not matched.	115
4.15	Cumulative distribution plot of RMSE for 32×32 -pixel sub-fields using different descriptors for feature matching on the synthetic VIS and LWIR imagery.	117
4.16	Sub-field locations for comparison of area- and feature-based algorithm metrics.	118

4.17 Performance comparison of area and feature-based techniques showing	
(a) BFIE, (b) the variance increase due to the bias gradient, (c) RSSB,	
(d) RSV, and (e) RMSE for four 32×32 -pixel sub-fields.	119

List of Tables

2.1	Spectral bands examined in this work along with their respective wave-lengths.	18
2.2	Parameters used for the GMM foreground detector for the different spectral bands.	24
2.3	Empirically derived background thresholds used in the GMM for the different spectral combinations.	27
2.4	Empirically calculated target false alarm rates produced for given thresholds of the foreground map.	35
2.5	Empirically calculated target detection rates produced for given thresh-olds of the foreground map.	36
2.6	Results summary for the proposed multi-spectral fusion algorithm tracking moving vehicles through the full motion video sequence. . .	42
4.1	Summary of metrics for shift-estimation of 32×32 -pixel sub-fields using area-based registration with VIS and LWIR imagery.	111

4.2	Summarized RMSE metrics for 32×32 -pixel sub-fields using four different feature descriptors. Note the Lowe descriptor did not have any sub-fields with a sub-pixel RMSE.	116
-----	---	-----

Preface

This dissertation is comprised of three journal articles authored by myself and my advisor, Dr. Michael Roggemann, along with contributions from Dr. Timothy Havens, Adam Webb, Dr. David Tyler, David Allen, and Jeff Dank. Chapter 2 is from the article [1] that was published in Optical Engineering on December 14, 2015 as “Multi-spectral detection and tracking of multiple moving targets in cluttered urban environments.” As the first author, I examined techniques to improve detection and tracking of targets by fusing multi-spectral imagery. I developed the algorithm to fuse the multi-spectral image sets, produced the results, and prepared the manuscript. The material in Chapter 3 is under submission with MDPI Sensors as “Target Localization and Tracking by Fusing Doppler Differentials from Cellular Emanations with a Multi-Spectral Video Tracker.” For this work I developed a method to fuse a multi-spectral video tracker with a constellation of RF receivers for localization and tracking of a specific target with a cell phone. I developed the code, performed the simulations, and produced the manuscript. Dr. Roggemann proposed the idea of fusing the multi-spectral image tracker with a cellular emanation to localize and track a target of interest. Dr. Roggemann, Adam Webb, and Dr. Havens provided assistance in conceptual understandings and aided in the writing of the manuscript. The final paper in Chapter 4 is from the article [2] that has been published in OSA Applied Optics as “Multi-spectral image shift-estimation error calculations using simulated

phenomenology.” As the main author of this work, I developed a novel framework for evaluating shift-estimation errors in multi-spectral imagery using synthetic data. I generated the synthetic data used in this work, wrote the code for testing numerous image registration algorithms, and compiled the results. The original idea of evaluating error in image registration algorithms was conceived by Dr. Tyler. The original airfield set was constructed by David Allen and Jeff Dank and modified by myself by adjusting the scene geometry and attributing the scene with thermal material properties. FORTRAN code to calculate the Fisher information for multi-spectral imagery was written by Dr. Tyler while I wrote a parallel wrapper in Python to calculate the Fisher information for a large number of sub-fields. The work in this dissertation is a result of my own effort, but the quality of the work was greatly improved through the assistance of the prior mentioned individuals.

Acknowledgments

First, I send the utmost thanks to my advisor, Dr. Michael Roggemann, who believed in me and provided me the opportunity and support to pursue my doctoral work. Without him I would not be where I am today; I will forever be indebted for this. I thank Dr. Laura Brown, Dr. Timothy Havens, and Dr. Christopher Middlebrook for serving as my committee members and contributing to my dissertation. I am very thankful for the support of Dr. David Tyler who has been a great mentor and provided many meaningful conversations on image registration and Fisher information. I would like to recognize the countless people at Michigan Technological University who generously helped me in many ways. Finally, I thank my family for their support and motivation through this journey.

List of Abbreviations

CDF	Cumulative Distribution Function
CRLB	Cramer-Rao Lower Bound
DD	Doppler Differential
DoG	Difference-of-Gaussian
FDOA	Frequency Difference of Arrival
FI	Fisher Information
FIE	Fisher Information Error
FOV	Field of View
FWHM	Full Width Half Maximum
GPU	Graphics Processing Unit
GMM	Gaussian Mixture Model
GT	Ground Truth
MWIR	Mid-Wave Infrared
NIR	Near Infrared
LWIR	Long-wave Infrared
RF	Radio Frequency
RMS	Root-Mean-Square
RMSD	Root-Mean-Square Difference

RMSE	Root-Mean-Square Error
RS	Root-Sum
RSS	Root-Sum-Squares
RSSB	Root-Sum-Squares Bias
RSV	Root-Sum Variance
SWIR	Short-Wave Infrared
TPSE	Tie-Point Shift-Estimation
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
UCRLB	Unbiased Cramer-Rao Lower Board
UFIE	Unbiased Fisher Information Error

Abstract

It is critical for defense and security applications to have a high probability of detection and low false alarm rate while operating over a wide variety of conditions. Sensor fusion, which is the the process of combining data from two or more sensors, has been utilized to improve the performance of a system by exploiting the strengths of each sensor. This dissertation presents algorithms to fuse multi-sensor data that improves system performance by increasing detection rates, lowering false alarms, and improving track performance. Furthermore, this dissertation presents a framework for comparing algorithm error for image registration which is a critical pre-processing step for multi-spectral image fusion.

First, I present an algorithm to improve detection and tracking performance for moving targets in a cluttered urban environment by fusing foreground maps from multi-spectral imagery. Most research in image fusion consider visible and long-wave infrared bands; I examine these bands along with near infrared and mid-wave infrared. To localize and track a particular target of interest, I present an algorithm to fuse output from the multi-spectral image tracker with a constellation of RF sensors measuring a specific cellular emanation. The fusion algorithm matches the Doppler differential from the RF sensors with the theoretical Doppler Differential of the video tracker output by selecting the sensor pair that minimizes the absolute difference or

root-mean-square difference. Finally, a framework to quantify shift-estimation error for both area- and feature-based algorithms is presented. By exploiting synthetically generated visible and long-wave infrared imagery, error metrics are computed and compared for a number of area- and feature-based shift estimation algorithms.

A number of key results are presented in this dissertation. The multi-spectral image tracker improves the location accuracy of the algorithm while improving the detection rate and lowering false alarms for most spectral bands. All 12 moving targets were tracked through the video sequence with only one lost track that was later recovered. Targets from the multi-spectral tracking algorithm were correctly associated with their corresponding cellular emanation for all targets at lower measurement uncertainty using the root-mean-square difference while also having a high confidence ratio for selecting the true target from background targets. For the area-based algorithms and the synthetic air-field image pair, the DFT and ECC algorithms produces sub-pixel shift-estimation error in regions such as shadows and high contrast painted line regions. The edge orientation feature descriptors increase the number of sub-field estimates while improving the shift-estimation error compared to the Lowe descriptor.

Chapter 1

Introduction

1.1 Motivation

Materials and manufacturing processes continue to advance and there is an increase in the performance of sensors including higher spatial resolution, greater range resolution, faster frame rates, and wider bandwidth. Although these advances have contributed to improvements in system performance, fundamental limits imposed by physics make it impossible for even the most advanced sensors to perform certain functions. The combination of data from two or more sensors, referred to as sensor fusion, is an area that has been explored for automatic targeting systems to improve detection statistics, location accuracy, and target classification [3, 4]. While

the advantages of fusing data are apparent, the process of fusing the data is not straightforward and algorithm development is generally difficult. Hall pointed out that sensor fusion can decrease reliability of a system if not done properly [5]. Sensor fusion requires algorithm design to be specific to the application and when implemented in a robust manner, fusion increases the probability of detection, lowers the false alarm rate, and increases accuracy of the detection location [6, 7].

There are a number of motivations to implement sensor fusion. First, scenarios exist where an independent sensor is incapable of performing a particular function, whereas adding sensors overcome limitations to perform that function. For example, image sensors in the visible band measure reflected spectral information for a given field of view, but do not measure the range to those targets. By fusing image data with range measurements from a LIDAR, a 3D point cloud of spectral content and range data can be constructed and used for further data exploitation. Another motivation for sensor fusion is to increase system performance by adding additional sensor measurements. This can increase the signal-to-noise ratio and increase accuracy of resulting data-products. Finally, fusion of multiple lower-cost sensors can provide the same or exceed the functionality of a single, higher-cost sensor. In imaging applications, high-resolution and high-bandwidth sensors can be expensive due to the required materials and complexity of the manufacturing process; the use of multiple imaging sensors to provide these functions can reduce system cost.

Sensor fusion has been a significant area of research for military applications as they generally require high probability of detection, with low false alarms rates while performing under the widest possible range of conditions. To improve detection of buried explosives, Pinar *et al.* presented methods to fuse forward-looking ground penetrating radar with visible and long-wave infrared imagery (LWIR). The authors operate independently on each modality and combine features using a kernel support-vector machine. Fasano *et al.* improved tracking of unmanned aerial systems (UAS) by fusing radar and electro-optical data using an extended Kalman filter [8]. The authors reduce both false alarms rates and improve tracking accuracy. Radar and lidar were used by Hollinger *et al.* to improve detection of partially obscured objects [9]. Detections from the two sensor were combined in a fusion process and filtered to remove false alarms.

Image fusion is a popular subset of sensor fusion that creates a data-product by combining data such as multi-spectral, polarimetric, or depth imagery. The fusion general occurs at one of three stages of processing: pixel-level, feature-level, and decision-level. Pixel-level fusion creates a single image that is a function of pixels from the individual images and commonly used by an image-analyst or down-stream processing. Feature-level fusion allows algorithms to be tailored for each image type and fuse the resulting by-products that include numerous classifications of features, such as foreground maps, histograms, edge contours, and texture features. In decision-level fusion, processing is performed on each image to where a decision is made, such

as object classification and location. These decisions are fused based on a confidence metric to give an overall system level decision.

Image fusion has been utilized in a number of applications including medical [10], remote sensing [11, 12, 13], and surveillance [14]. A common medical application fuses image data from different image sources, including magnetic resonance imaging (MRI), computed tomography (CT) scans, and positron emission tomography (PET) scans, to improve diagnosing of diseases and injuries [10]. These applications use pixel-level fusion to create an image that exploits the body structure targeted by each modality. In remote sensing, a common application of image fusion is classifying land-use from panchromatic and hyper-spectral sensors. Hyper-spectral sensors produce images for a number of spectral bins which contain information about the ground cover, but spatial resolution of the images is relatively low. Panchromatic sensors image a single spectral band at a high spatial resolution. The fusion is performed at the pixel-level where the hyper-spectral image is up-sampled using the panchromatic image to create an image with high spectral content at a high spatial resolution. Surveillance applications utilize multi-spectral imagery to increase probability of detection and increase tracking performance; visible and thermal imagery are often fused. These two bands offer complementary information as visible imagery contains reflected spectral information while thermal imagery contains emissive information.

Registration is a critical pre-processing step for fusing data collected at different time

samples, fields of view, and/or reference frames. For example, two sensors may have different locations and controlled by different clock cycles, resulting in both spatial and temporal mis-registration. The registration process aligns the data by finding common tie-points between them and calculates the transformation needed to align the data to a common coordinate system in which the fusion process can occur. There is increased difficulty in the registration process when the dimensionality of the data sources vary, making a common reference system challenging to define. For example, one sensor may produce data in two-dimensions, such as an (x, y) -pixel measurement from an image frame, and the other sensor may produce data in one-dimension such as an angle-of-arrival measurement from a radio frequency sensor.

Spatial registration of image data, referred to as image registration, is particularly important for image fusion. Image registration estimates the geometric transformation required to align two images so that corresponding pixels overlap. A number of registration algorithms have been demonstrated to produce sub-pixel accuracy for images of the same spectral-band using area-based [15] and feature-based [16] algorithms. Imagery that contains phenomenological differences, such as multi-spectral or polarimetric, is particularly useful for fusion as the images contain differing information, but this also makes the registration process more challenging due to intensity differences between the two images. For visible and thermal spectral bands, contrast is often inverted between the two bands; for example, the color white produces a high reflectively (high radiance) in the visible bands, but results in low solar absorption

and a cooler temperature (low radiance) in the thermal bands.

Algorithm development for detection and tracking algorithms is challenging and difficult to evaluate the performance without ground-truth information which is expensive and time consuming to collect. Algorithm development is significantly more challenging in sensor fusion where accurate ground-truth information can be nearly impossible to obtain due to the aforementioned registration requirements. Synthetic data is a useful tool for algorithm development as the ground-truth information is known and can be used to evaluate algorithm performance, but is only effective if it's representative of the real data. A dataset developed for algorithm training and evaluation, known as "The SYNTHIA Dataset," is a collection of synthetically generated images of urban scenes where each image is accompanied by a ground-truth image that labels the classifications of the pixels in the image [17]; the authors note that this imagery requires adaption to be comparative to actual sensor data. For simulating multispectral images, DIRSIG is a physics-based rendering tool that calculates aperture reaching radiance from 0.4 - 20.0 μm and can be used to form synthetic imagery as would be collected by a real system [18].

1.2 Approach

In this dissertation, I present approaches of fusing data from multiple sensors to improve detection and tracking performance for a collection of targets in multi-spectral full-motion video, along with localizing targets of interest. I also present a framework for evaluating area- and feature-based algorithms for shift-estimation of multi-spectral imagery.

First, an algorithm was developed to fuse multi-spectral image data to improve detection and tracking of multiple moving targets in cluttered urban environments using synthetically generated data of four spectral bands; visible, near-infrared (NIR), mid-wave infrared (MWIR), and LWIR. For all spectral bands, a background model is used on the time history of the pixel intensities using a Gaussian mixture model for foreground pixel detection. Foreground pixels from all spectral bands are weighted, summed, and filtered to create a fused foreground map and image segmentation is performed on the fused map to detect target candidates. A database of tracked targets is used to store position history and features from the scale-invariant feature transform for associating targets between frames and recovering lost tracks. The proposed tracking algorithm are created track sequences for all 12 targets from a synthetic full motion video set.

To localize and track a specific target of interest, I propose to fuse a new combination of sensors; a multi-spectral video tracker and a constellation of RF sensors measuring a cellular emanation. The RF sensors measure the Doppler shift in the transmitted carrier frequency caused by radial motion of a cellular emanation; Doppler differentials (DD) between all sensor pairs are calculated. The data is fused by associating the DD from the RF sensors with the theoretical DD computed from the multi-spectral tracker output. The DD from the sensors are matched by comparing the absolute difference at each time instance and the root-mean-square difference.

Finally, I present a framework for quantitative comparison of area- and feature-based image registration algorithms using synthetically generated multi-spectral imagery; visible and LWIR are used in this work. Shift-estimate errors are calculated for 32×32 -pixel sub-fields using area- and feature-based algorithms; errors include estimator bias, variance, bias gradient, root-mean-square error, and bias gradient. Parallel computing is utilized to generate error images and cumulative distribution plots for all sub-fields from the full field of the synthetic imagery. Error metrics are compared to variances derived from relevant Fisher information matrices. The synthetic imagery is generated using computer graphics software to create sets of 3D models attributed with material reflectively and thermal properties, then rendered using the DIRSIG tool-set, and finally, a custom optics and detector effects model is applied.

1.3 Summary of Key Results

I present the results of the multi-spectral image fusion algorithm for improving detection and tracking in cluttered urban environments using synthetically generated visible, NIR, MWIR, and LWIR video sequences that include moving vehicles with known ground-truth locations. The fusion of the multi-spectral foreground maps improved the accuracy of target centroid estimation while increasing the detection rates and lowering the false alarm rates for most spectral bands. LWIR produced the highest detection rate, but produced the highest false alarm rate. The spectral combination VIS-MWIR had only a slightly lower detection rate relative to LWIR, but significantly reduced the false alarm rate. Using features as the tracking mechanism, NIR-MWIR had the largest percentage of tracks attributed to features. For centroid location accuracy, NIR-LWIR produced the lowest displacement of less than 0.1 meters. The tracking algorithm successfully tracked all 12 moving targets, with only one instance of a lost track that was later recovered.

To localize and track a specific target of interest, track sequences of moving targets produced by the multi-spectral tracking algorithm were fused with DDs measured from a constellation of RF sensors. Two association metrics were used to match the video tracker results with the DD measurements; absolute DD and root-mean-square difference. The absolute DD performs well with low measurement uncertainty for

targets that have motion patterns that are not similar to other targets. The RMSD performs better particularly for low uncertainty cases, but maintains good performance through significant measurement uncertainty levels for targets with favorable motion patterns. The confidence ratio using the RMSD performs well for nearly all targets with low measurement uncertainty and maintains high rates for over half of the targets through all uncertainty levels.

I present quantitative metrics, including bias, variance, root-mean-square error, and bias gradient, for evaluation of shift-estimation using area- and feature-based registration for multi-spectral imagery. For area-based algorithms and the synthetic image pair, the MI algorithm had significantly higher error and computational cost than both the DFT or ECC algorithms. Also, over field angles where the difference between reflective and emissive phenomenology resulted in contrast reversals at edges, the ECC algorithm performed noticeably better than DFT; however, where edges did not have contrast reversal between VIS and LWIR, such as shadow edges, the DFT algorithm tended to outperform the ECC. For feature-based algorithms and the synthetic image pair, the Lowe SIFT descriptor produced the worst results, with no sub-fields having an RMSE less than 1 pixel. The PCEHD descriptor had the highest percentage of errors between 0.5 and 1 pixel, and the LGHD descriptor had feature matches in the largest number of sub-fields. Finally, I was able to calculate variances derived from relevant Fisher Information matrices for both unbiased estimation and non-zero algorithm bias gradient. Over 4 sub-fields selected from the image pair, the

feature-based algorithms appeared to depend less on the true shift (lower bias gradient), resulting in relatively smaller variance added to the unbiased Fisher information error.

1.4 Organization

The remainder of this dissertation is composed of three journal articles that have been published or still under the review process. The papers were edited to fit the format and flow of this document. Chapter 2 is from the paper titled “Multi-Spectral Detection and Tracking of Multiple Moving Targets in Cluttered Urban Environments” that was published by SPIE in Optical Engineering, Vol. 54, No 12, December 2015 [1]. This paper presents an approach to fusing multi-spectral imagery to increase the detection and tracking performance of an airborne surveillance system. Chapter 3 is from the paper titled “Target Localization and Tracking by Fusing Doppler Differentials from Cellular Emanations with a Multi-Spectral Video Tracker” that is currently under the review process with MDPI Sensors. The paper presents an approach to fuse a multi-spectral image tracker with RF sensors to localize and track a target of interest. Chapter 4 is from the paper titled “Multi-spectral image shift-estimation error calculations using simulated phenomenology” that has been accepted for publication in OSA Applied Optics [2]. The paper presents evaluation metrics for area- and feature-based shift estimation algorithms using synthetic visible and long-wave

infrared imagery. Chapter 5 presents conclusions and highlights the accomplishments in the work, and provides suggestions for future research.

Chapter 2

Multi-Spectral Detection and Tracking of Multiple Moving Targets in Cluttered Urban Environments

2.1 Abstract

I present an algorithm to improve the performance for target detection and tracking by fusing foreground pixels from multi-spectral imagery. In all spectral bands, pixel

intensities are modeled using a Gaussian mixture model, and pixels not belonging to a background distribution are classified as foreground pixels. Foreground pixels from the spectral bands are weighted and summed into a single foreground map and filtered to give the fused foreground map. Foreground pixels are grouped into target candidates and associated with targets from a tracking database by matching features from the scale-invariant feature transform. The performance of the presented algorithm was evaluated with a synthetically generated data set of visible, near infrared, mid-wave infrared, and long-wave infrared video sequences. With a fused combination of the spectral bands, the proposed algorithm lowers the false alarm rate while maintaining high detection rates. All twelve vehicles were tracked throughout the sequence, with one instance of a lost track that was later recovered.

2.2 Introduction

Automatic detection and tracking of moving targets in full motion video (FMV) from aerial imaging systems such as unmanned aerial vehicles (UAV) and satellites are of significant interest in the defense and security communities [19, 20, 21]. These aerial platforms can remain undetected from prospective targets and encompass a large surveillance area. Satellites have been a primary “spy” tool for decades and continue to provide for their respective nations, but their coverage is limited by orbital mechanics, and is hence not always sufficiently timely; nor can a satellite generally

be launched on demand to address a short term tactical matter in the field. Vast amounts of research have been invested in UAV surveillance, and UAVs have been a significant resource for intelligence gathering [19, 22, 23]. Large areas, such as open waters or borders, can be surveyed for intrusions, regions can be assessed for building of weapon facilities, or urban areas can be checked for potential threats.

The urban environment is of interest for this work. Urban environments provide significant challenges to the problem of automatically detecting and tracking moving vehicles. These areas generally contain complicated clutter and a collection of different targets, for example, humans, buildings, roads, and vehicles. Each of these different entities also vary greatly in shape and size that challenge automatic target detection and recognition algorithms. Trees, buildings, tunnels, and other formations result in object occlusions that affect the appearance of the targets and sometimes completely block the targets from view for a few to several consecutive frames.

Multi-spectral imagery is useful for increasing the robustness of detection and tracking targets through cluttered urban environments. Images in the visible spectrum ($0.4 \mu m - 0.7 \mu m$) provide reflected spectral information that creates contrast between targets, and between targets and the background. The visible spectrum requires good illumination during the daytime hours. Imaging in the long-wave infrared (LWIR) band ($8 \mu m - 14 \mu m$) is dependent on the temperature and thermal emissivity of the target, but is not dependent on solar illumination, and thus provides night time

imaging capabilities. The effects of atmospheric aerosols also plays a role in these imaging modalities. For example, Mie scattering significantly hinders the performance of visible imaging in the presence of fog aerosols[24]. The wavelengths of the mid-wave infrared (MWIR) and LWIR bands are longer than the visible wavelength, making them less susceptible to the attenuation due to Mie scattering, and thus provide some immunity to the effects of fog and other aerosols on image quality[24].

The approach of multi-spectral detection and tracking fuses information obtained from images in different spectral bands to improve detection statistics. Various approaches have been taken in algorithm development for detecting and tracking using multi-spectral imagery where the fusion framework takes place in three stages of the processing: pixel-level [25, 26, 27, 28, 29], feature-level [30, 31, 32], and decision-level [33]. Fusion at the pixel-level creates a single image that is a composition of the pixels in the multi-spectral images. It is often used to create a single image that is interpreted by an operator [34, 35]. The combination of pixels into a single image is difficult as there is not always a correlation of the pixel values from the different spectral images, and it has been found that a mild anti-correlation exists between the visible and LWIR bands [36]. Feature-level fusion combines the byproduct of processing of individual spectral bands. These processing products include numerous classifications of features, such as foreground maps, histograms, edge contours, and texture features. Processing of individual spectral bands allows feature extraction

algorithms to be optimized for each band. In decision-level fusion, processing is performed on each independent spectral band where a decision is made, such as object size and location. These decisions are fused based on band-specific confidence levels to give an overall decision.

To exploit benefits of each spectral band, feature and decision level fusion allow algorithm development tailored for their respective bands. Algorithms fusing background models from different image modalities to create a common foreground for target detection have been demonstrated [30, 31, 32, 33]. Chen and Wolf model the foreground in both visible and LWIR imagery with the mixture-of-Gaussians model, while using an adaptive learning rate that is based on the decision of each spectral band [31]. They also fuse the two spectral bands for their appearance model to increase the performance of target association. Torresan *et al.* perform the background subtraction on each individual spectral modality and merge the results by picking a master and slave foreground map based upon the confidence of each modality [33]. By modeling each background pixel's intensity as a single Gaussian distribution, Davis and Sharma extract regions-of-interest by the intersection of the visible and LWIR foreground maps [30]. Salient contours from the regions-of-interest are then calculated from both visible and LWIR images and fused to create a single contour saliency map.

The aforementioned works consider visible and LWIR bands; this work additionally exploits near infrared (NIR) and MWIR bands, and the combinations of spectral

bands. Table 2.1 shows the spectral bands used and their associated wavelengths. The main contribution of this work is an algorithm to fuse multi-spectral data sets to reliably detect and track moving targets with high probability and low false alarm rate. I focus on detection and tracking of vehicles through an urban scene that includes partial occlusions and crowded traffic intersections. A block diagram of the proposed algorithm is shown in Fig. 2.1. To compensate for fluctuating pixel intensities in each spectral band, background models using a Gaussian mixture model (GMM) adapt to the evolving scenes and detect foreground pixels. Foreground pixels from different spectral bands are fused into a foreground region and filtered to obtain a single foreground map that represents pixel regions belonging to target candidates. Features based on the scale-invariant feature transform (SIFT) are extracted from these target regions and used for two purposes [15]: detecting targets missed by the segmentation detection, and associating targets from a tracking database constructed from prior frames. Lastly, locations for each target are estimated and the GMM mixture is updated.

Table 2.1
Spectral bands examined in this work along with their respective wavelengths.

Spectral Band	Wavelengths (μm)
Visible	0.4 - 0.7
Near Infrared (NIR)	0.8 - 1.2
Mid-Wave Infrared (MWIR)	3 - 5
Long-Wave Infrared (LWIR)	8 - 14

To develop and evaluate the algorithm, a UAV imaging scenario was synthetically

Figure 2.1: Flowchart of the detection and tracking algorithm using fusion of multi-spectral imagery and SIFT features.

generated from the Digital Imaging and Remote Sensing Image Generation (DIRSIG) toolset [18]. DIRSIG is a mature and widely used simulation package for $0.4 \mu m$ - $20 \mu m$ wavelengths. An urban scene with 12 vehicles was simulated at visible, NIR, MWIR, and LWIR wavelengths. A normal traffic scenario was simulated using the open source tool Simulation of Urban MObility (SUMO) to provide realistic traffic maneuvers. Figure 2.2 shows a 2000×2000 pixel frame from each spectral band. By visual inspection, the appearance of target vehicles varies between the scenes, providing different intensity information.

The remainder of this chapter is organized as follows. Section 2.3 presents the method for foreground extraction using the GMM and the region growing process to group disjoint pixels. I also discuss the method for fusing the spectral modalities in Sec. 2.3.

Figure 2.2: Example multi-spectral frames from synthetic DIRSIG data set. (a) - visible, (b) - NIR, (c) - MWIR, (d) - LWIR

Section 2.4 presents the association target candidates with track sequences. Experimental results on the performance of the algorithm are presented in Sec. 2.5. In Sec. 2.6 I present conclusions.

2.3 Detection and Segmentation Algorithm

In this section I describe the detection and segmentation algorithms; I then present the fusion process used to combine foreground maps to build pixel regions that represent target candidates.

Pixel intensities fluctuate due to changes in illumination, movement from both background and target objects, and pixel noise. This does not allow a single value to characterize the time history of the intensity of a single pixel for a given video sequence. To compensate for these changes, background modeling techniques are used to describe the probability distribution of the pixels intensity by empirically deriving the parameters from the video sequence. The GMM has been successfully demonstrated to compensate for the fluctuations in pixel intensities [37, 38, 39]. In a scene where the sensor is fixed, keeping the viewpoint stationary, I use statistical information extracted from the time history of the intensity fluctuations to understand the probability distribution of intensity at each pixel, and use these distributions to make hypotheses about the label of each pixel. Each pixel in the scene is classified as a foreground or background pixel, and I update the parameters of the GMM during each frame. I now describe this process in detail.

I define $X(x, y; t)$ as the pixel intensity at location (x, y) and time t . The goal is to

classify this pixel as a background or foreground pixel by fitting it to a distribution model. The distribution of the time history of the intensity, $P(X(x, y; t))$, is modeled as a sum of weighted Gaussian distributions:

$$P(X(x, y; t)) = \sum_{j=1}^K w_{j,t}(x, y) \mathcal{N}(X(x, y; t), \mu_{j,t}(x, y), \Sigma_{j,t}(x, y)), \quad (2.1)$$

where K is the number of Gaussian distributions, $\mu_{j,t}(x, y)$ is the mean of the distributions, and the covariance matrix, which is assumed to be diagonal, is given by $\Sigma_{j,t}(x, y) = \sigma_{j,t}^2(x, y)I$, where I is the identity matrix. The weighting factor $w_{j,t}(x, y)$ represents the portion of which the j^{th} Gaussian comprises the entire model, and is dependent on the number of occurrences for the particular distribution. This weighting has range $0 < w_{j,t} \leq 1$, and is normalized such that $\sum_{j=1}^K w_{j,t} = 1$. The Gaussian probability density function is

$$N(X(x, y; t), \mu_{j,t}(x, y), \Sigma_{j,t}(x, y)) = \frac{1}{(2\pi)^{n/2} |\Sigma_{j,t}(x, y)|^{1/2}} \times \exp \left(-\frac{1}{2} (X(x, y; t) - \mu_{j,t}(x, y))^T \Sigma_{j,t}(x, y)^{-1} (X(x, y; t) - \mu_{j,t}(x, y)) \right) \quad (2.2)$$

From the K distributions, it must be determined what number of distributions are classified as belonging to the background. The top B weighted distributions are

selected as the background, where

$$B = \arg \min_b \left(\sum_{j=1}^b w_{j,t}(x, y) > Thr \right). \quad (2.3)$$

The threshold Thr is user defined with range $[0, 1]$ and is dependent on the scene.

In a complex scene with multiple moving targets where pixel distributions vary among targets and among targets and background, more Gaussian models will be present and thus require a higher Thr . In the scenes tested with this algorithm, few objects were moving and typically only one Gaussian mode was needed to describe the background. By executing the GMM algorithm with a series of parameters on the test data set, the optimal Thr was empirically derived for each spectral band by comparing correctly detected pixels to falsely detected pixels. The resulting values of Thr are shown in Table 2.2. In the algorithm, LWIR had the lowest Thr at 0.5, which is attributed to the distributions of the target intensities being similar, along with being lower than the background surrounding the targets.

To evaluate whether the current pixel intensity $X(x, y; t)$ is a background or foreground pixel, I calculate the a priori probability of that pixel intensity belonging to each of the K distribution components. If the intensity value falls within 2.5 standard deviations of any background distribution, it is labeled background, otherwise it is labeled as foreground. Following the classification of the pixel, the distribution

parameters are updated as [39]:

$$w_{j,t+1}(x, y) = w_{j,t}(x, y) + \alpha (1 - w_{j,t}(x, y)), \quad (2.4)$$

$$\mu_{j,t+1}(x, y) = \mu_{j,t}(x, y) + (\alpha/w_{j,t}(x, y)) (X(x, y; t) - \mu_{j,t}(x, y)), \quad (2.5)$$

$$\sigma_{j,t+1}^2(x, y) = \sigma_{j,t}^2 + (\alpha/w_{j,t}(x, y)) \times \left((X(x, y; t) - \mu_{j,t}(x, y))^T (X(x, y; t) - \mu_{j,t}(x, y)) - \sigma_{j,t}^2(x, y)^2 \right), \quad (2.6)$$

where α is the learning rate. In a scene where objects typically move slowly, the update equations should also update at a slower rate and require a smaller α . After experimentation I found the optimal α for each data set, shown in Table 2.2. All spectral bands used a low α , with the lowest value in the LWIR band which can be attributed to no shadows being present.

Table 2.2
Parameters used for the GMM foreground detector for the different spectral bands.

	VIS	NIR	MWIR	LWIR
Learning Rate	0.0004	0.0005	0.0004	0.0003
Threshold	0.7	0.8	0.7	0.5

The GMM algorithm produces intermediate foreground maps in all spectral bands that do not represent the complete target region and do not necessarily correlate with one another. This is a consequence of discrepancies in the foreground modeling, and

is caused by low SNR between the target and background. Examples of intermediate foreground maps at frame 600 are shown in Fig. 2.3. In the NIR band, the bottom target has a high number of foreground pixels in comparison to the other bands. In the MWIR band, the target on the left has a low number of foreground pixels whereas the other bands have a high number of pixels. The fusion of foreground maps from multi-spectral video creates combined foreground maps that accurately estimates the centroid of the target with a low false alarm rate. This is distinct from previous efforts [30, 31, 32, 33, 34, 35] in that I have considered additional spectral bands for foreground fusion, and use SIFT features for unique target identification and detection of missed targets.

I define a fused foreground map, $FG_{FUS}(x, y)$, as the sum of individual weighted foreground maps, where w represents the weighting and the subscript represents the respective band,

$$FG_{FUS}(x, y) = w_{VIS}FG_{VIS}(x, y) + w_{NIR}FG_{NIR}(x, y) + w_{MWIR}FG_{MWIR}(x, y) + w_{LWIR}FG_{LWIR}(x, y). \quad (2.7)$$

$FG_{FUS}(x, y)$ is spatially filtered with a 3×3 Gaussian filter with $\sigma = 0.5$. Thresholding of $FG_{FUS}(x, y)$ is performed to remove pixels that have a low foreground

Figure 2.3: Example foreground images detected using the GMM model at frame 600. (a) - VIS, (b) - NIR, (c) - MWIR, (d) - LWIR

probability of belonging to the foreground, $FG_{FUS}(x, y) < th$. The spectral combinations and their respective thresholds are shown in Table 2.3. Thresholds were chosen by the lowest false alarm rate produced by the detection algorithm. False alarm rates for the series of tested thresholds are presented in the results section in Table 2.4.

Table 2.3

Empirically derived background thresholds used in the GMM for the different spectral combinations.

Combination	Background Threshold
VIS	—
NIR	—
MWIR	—
LWIR	0.2
VIS-NIR	1.0
VIS-MWIR	—
VIS-LWIR	1.6
NIR-MWIR	0.4
NIR-LWIR	1.2
MWIR-LWIR	1.0
VIS-NIR-MWIR	1.2
VIS-NIR-LWIR	2.0
VIS-MWIR-LWIR	2.0
NIR-MWIR-LWIR	1.2
VIS-NIR-MWIR-LWIR (TOT)	2.0
VIS-NIR-MWIR-3*LWIR (TOT3)	3.0
VIS-3*LWIR	3.4

Shown in Fig. 2.4 (a) is an example of a fused foreground. A smoothing of the combined foreground map is applied using a 2D Gaussian filter and shown in Fig. 2.4 (b). The filtering results in filling of gaps where pixels were missed from the foreground map without over dilating the region. The final foreground map is shown in Fig. 2.4 (c). A zoomed area on a car region depicting the foreground fusion process is shown in Figure 2.5. The effect of thresholding the fused and filtered foreground map is illustrated; the target shadow is removed from the foreground region.

The final step of creating the pixel regions that represent the detected candidates is an image closing, which consists of a dilation followed by an erosion. The structuring

Figure 2.4: Example fused foreground maps using all spectral bands (TOT) for a single frame. (a) is the fused foreground maps, $FG_{FUS}(x, y)$. (b) is the foreground map after applying a Gaussian Filter. (c) shows the foreground after a threshold has been applied to the filtered image.

Figure 2.5: Zoomed area highlighting the algorithm for fusing the foreground pixels from multi-spectral imagery.

element of this procedure is a disk with a radius of four pixels. The dilation operation fills in voids between pixel segments and grows the size of the region. In the erosion operation, I attempt to remove any unnecessary region growth that is a by-product of the dilation. Pixel regions that do not exceed an area of 200 pixels are filtered to remove objects that may not represent vehicle-sized objects.

2.4 Target Tracking

The association of targets involves relating a track sequence from prior frames with target candidates detected in the current frame. This task is trivial in the case where targets stay separated and no occlusions exist. However, in actual practice and in this data set, targets become merged or occluded, making distinguishing between targets difficult.

The SIFT feature were chosen for identification due to their robustness with respect to changes in rotation and scale, and their invariance to change in camera viewpoints and illumination changes [15]. Due to the reliance on these features to uniquely identify targets, they must be robust for long-term tracking applications. A disadvantage of SIFT is the heavy computations required for the keypoints, where typical processing times are tenths of seconds to multiple seconds per frame in a normal CPU implementation [40, 41]. Developments in graphics processing units (GPU) and field programmable gate arrays (FPGA) have created opportunities for real-time algorithms. SIFT implementations have been developed for both GPUs [41, 42, 43] and FPGAs[40], where results demonstrate real-time SIFT calculations.

SIFT features are composed of a keypoint that gives sub-pixel location and orientation of the feature, along with a descriptor that is calculated based on local pixel texture.

In the SIFT algorithm, keypoints are first identified at multiple scales. A scale space of the image is created with varying amounts of blur applied to each image using the Gaussian kernel. The blurred image is defined as

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y), \quad (2.8)$$

where the Gaussian kernel $G(x, y, k\sigma)$ with variance $k\sigma$ is

$$G(x, y, k\sigma) = \frac{1}{2\pi k\sigma^2} \exp\left(-\frac{x^2 + y^2}{2k\sigma^2}\right). \quad (2.9)$$

Within the scale space, Difference of Gaussian (DoG) images are calculated by

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma). \quad (2.10)$$

The local extrema in the DoG images at each scale are found by comparing the pixel value with its 8 surrounding pixels and the 9 neighboring pixels from each of the nearest blurred images. To create an invariance to scale, the extrema must exist on multiple scales. A filtering step of the detected extremas in the DoG images is implemented based on the intensity; an extrema with a low intensity is susceptible to changes in illumination and is therefore unstable and removed from the keypoints.

A reference orientation for subsequent processing is given to the keypoint to provide

invariance to rotation. From the blurred image in which the extrema was located, the gradient magnitude $m(x, y, k\sigma)$ is calculated by

$$m(x, y, k\sigma) = \sqrt{(L(x+1, y, k\sigma) - L(x-1, y, k\sigma))^2 + (L(x, y+1, k\sigma) - L(x, y-1, k\sigma))^2} \quad (2.11)$$

and the orientation, $\theta(x, y, k\sigma)$, by

$$\theta(x, y, k\sigma) = \tan^{-1} \left(\frac{L(x, y+1, k\sigma) - L(x, y-1, k\sigma)}{L(x+1, y, k\sigma) - L(x-1, y, k\sigma)} \right). \quad (2.12)$$

A histogram of 10 degree bins is created of the orientations, and the magnitudes added to the histograms are the gradient magnitudes that are Gaussian weighted with a variance of $1.5k\sigma$. The peaks of the histograms are detected, where the highest peak and any peaks above 80% of the highest peak are selected as orientations for the new keypoints. Peaks in the histogram represent dominant directions of the local gradients.

Unique identifications are generated for each keypoint, referred to as descriptors. A 16×16 region around the keypoint, with respect to the calculated orientation, is divided into 4×4 sub-regions. Gradient magnitudes and orientations are calculated

for each pixel in these sub-regions, and histograms with 45 degree bins are calculated for each sub-region. Through the use of a Gaussian weighting mask with $\sigma = 1/2$ of the descriptor window width (16 pixels for this case), points are inversely weighted proportional to their distance from the keypoint to decrease their contributions and reduce errors caused by window displacements.

To match features from a tracked objects database to features from the current scene, a matching score is calculated by the Euclidean distance between two descriptors. The matching score between a tracked object and a frame object is calculated by

$$\begin{aligned} Score(D^{obj}, D^{frm}) &= \sqrt{(d_1^{obj} - d_1^{frm})^2 + (d_2^{obj} - d_2^{frm})^2 + \dots + (d_n^{obj} - d_n^{frm})^2} \\ &= \sqrt{\sum_{i=1}^n (d_i^{obj} - d_i^{frm})^2}. \end{aligned} \tag{2.13}$$

where $D^{obj} = \{d_1^{obj}, d_2^{obj}, \dots, d_n^{obj}\}$ is the tracked object descriptor, $D^{frm} = \{d_1^{frm}, d_2^{frm}, \dots, d_n^{frm}\}$ is the frame object descriptor, and n is the length of the descriptor vector, which is 128 for this case.

The feature with the shortest Euclidean distance, i.e., the nearest neighbor, is selected as the matching feature. To remove matches that do not have a good match, a comparison is made between the nearest neighbor and the next nearest neighbor. If the ratio of the match scores between the nearest and the next nearest neighbor is

greater than 0.8, the match is rejected. Lowe found that this method rejects 90% of all incorrect keypoints and only removes 5% of the correct matches[15].

Updating the track location is based on several factors. The search region for a matching target candidate is limited to the track’s estimated bounding box, preventing erroneous associations with targets of similar appearance but at a distance away. In the event that multiple targets are located in the track bounding box, such as at a road intersection when cars become merged, SIFT features are used to select the correct target. If no target is found in the bounding box, SIFT features are matched in the bounding box region and provide a velocity measurement for a linear motion model. Tracks are propagated if no target is matched and no SIFT features are found, a typical occurrence when the target may be partially or fully occluded from view of the sensor. The propagation projects the location of the bounding box linearly into future frames based on the most recent position and velocity prior to the occlusion.

2.5 Experiment

The performance of this algorithm was evaluated with a synthetically generated data set using the DIRSIG toolset[18]. A standard mid-latitude summer model was used for the atmospheric model MODTRAN[44]. The thermal signatures for 12 vehicles were simulated with the thermal prediction software MuSES[45]. Realistic traffic patterns

were generated using the SUMO traffic simulator[46]. The video sequence consists of 600 frames of 2000×2000 pixels sampled at 20 frames per second. The ground sample distance (GSD) is 0.0635 meters and frames are co-aligned where pixels correspond geometrically between frames and registered between the spectral bands. As this is a synthetically generated data set, the locations of pixels corresponding to each vehicle is known, providing ground truth centroids of vehicles in the scene.

2.5.1 Detection Results

I now present the performance for detecting moving targets using the fusion algorithm applied to the DIRSIG data set. For the evaluation of segmented detection rates, a successful detection is a group of pixels that has a centroid with a Euclidean distance within 0.95 meters of the centroid of a ground truth object, otherwise it is considered a false alarm. False alarms are reported on a per frame basis. Targets occluded by 20% or more are not factored into the target detection score.

False alarm rates for a series of examined thresholds are shown in Table 2.4, where the optimal thresholds are highlighted. The false alarm rate is presented by the number of false alarms per frame. The optimal thresholds for image filtering were selected by choosing the lowest false alarm rate for their respective spectral combination. LWIR resulted in the highest false alarm rate at 1.30. Three of the fusion combinations have

false alarms less than 1; MWIR-LWIR, VIS-NIR-MWIR, and NIR-MWIR-LWIR. MWIR-LWIR presented the lowest false alarm rate of 0.96. In the single spectral bands, NIR had the lowest false alarm rate with 1.07.

Detection rates by segmentation for a series of examined thresholds are shown in Table 2.5, where the results for the optimal thresholds for each spectral combination are highlighted. LWIR achieved the highest detection rate of 0.94 and VIS had a detection rate of 0.93. VIS-MWIR and VIS-LWIR resulted in detection rates of 0.93, while the weighted TOT3 and VIS-LWIR3 had results of 0.91 and 0.92, respectively.

Table 2.4
Empirically calculated target false alarm rates produced for given thresholds of the foreground map.

Thr.	VIS	NIR	MWIR	LWIR	VIS NIR	VIS MWIR	VIS LWIR	NIR MWIR	NIR LWIR	MWIR LWIR	VIS NIR	VIS MWIR	VIS LWIR	NIR MWIR	NIR LWIR	TOT	TOT3	VIS LWIR3
0.0	1.14	1.07	1.15	1.32	1.17	1.14	1.38	1.18	1.45	1.41	1.17	1.42	1.39	1.45	1.42	1.42	1.38	
0.2	1.30	1.08	1.19	1.30	1.34	1.26	1.41	1.12	1.44	1.35	1.31	1.41	1.39	1.46	1.37	1.38	1.38	
0.4	1.30	1.08	1.19	1.30	1.31	1.26	1.40	1.03	1.43	1.36	1.30	1.39	1.39	1.44	1.36	1.36	1.41	
0.6	1.30	1.08	1.19	1.30	1.31	1.26	1.40	1.03	1.43	1.36	1.30	1.39	1.39	1.44	1.37	1.36	1.40	
0.8	1.39	1.26	1.77	1.65	1.26	1.27	1.46	1.11	1.42	1.46	1.26	1.49	1.49	1.43	1.49	1.50	1.44	
1.0	—	—	—	—	1.08	1.15	1.16	1.12	1.10	0.96	1.02	1.16	1.23	1.01	1.17	1.41	1.30	
1.2	—	—	—	—	1.14	1.31	1.24	1.24	1.08	1.17	0.99	1.15	1.22	0.99	1.16	1.36	1.29	
1.4	—	—	—	—	1.15	1.33	1.24	1.25	1.12	1.24	1.02	1.12	1.24	1.02	1.17	1.38	1.30	
1.6	—	—	—	—	1.12	1.36	1.17	1.34	1.16	1.34	1.11	1.16	1.27	1.05	1.18	1.37	1.30	
1.8	—	—	—	—	1.31	1.53	1.23	1.45	1.17	1.28	1.16	1.29	1.24	1.16	1.19	1.37	1.30	
2.0	—	—	—	—	—	—	—	—	—	—	1.22	1.08	1.10	1.13	1.04	1.36	1.31	
2.2	—	—	—	—	—	—	—	—	—	—	1.30	1.08	1.21	1.16	1.07	1.35	1.32	
2.4	—	—	—	—	—	—	—	—	—	—	1.37	1.11	1.24	1.09	1.09	1.40	1.56	
2.6	—	—	—	—	—	—	—	—	—	—	1.34	1.15	1.26	1.12	1.13	1.47	1.54	
2.8	—	—	—	—	—	—	—	—	—	—	1.47	1.30	1.21	1.15	1.14	1.41	1.50	
3.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.15	1.16	1.29	
3.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.17	1.20	1.18	
3.4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.16	1.18	1.14	
3.6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.24	1.21	1.14	
3.8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.25	1.21	1.29	
4.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.06	—	
4.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.08	—	
4.4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.07	—	
4.6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.06	—	
4.8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.11	—	
5.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.15	—	

‘—’ implies the threshold exceeded the maximum obtainable pixel value in the image

Total detection rates and false alarm rates are shown in graph form in Figure 2.6. The total detection rates include detections by both segmented objects and features. In the single spectral bands, LWIR resulted in a detection rate of 0.94, but suffered

Table 2.5
Empirically calculated target detection rates produced for given thresholds
of the foreground map.

Thr	VIS	NIR	MWIR	LWIR	VIS NIR	VIS MWIR	VIS LWIR	NIR MWIR	NIR LWIR	MWIR LWIR	VIS MWIR	VIS LWIR	NIR MWIR	TOT	TOT3	VIS LWIR
0.0	0.93	0.91	0.89	0.96	0.93	0.93	0.93	0.91	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
0.2	0.92	0.87	0.85	0.94	0.92	0.92	0.93	0.89	0.91	0.92	0.92	0.93	0.93	0.91	0.93	0.94
0.4	0.91	0.87	0.85	0.94	0.92	0.92	0.93	0.89	0.91	0.92	0.92	0.93	0.93	0.92	0.93	0.93
0.6	0.91	0.87	0.85	0.94	0.92	0.92	0.93	0.89	0.91	0.92	0.92	0.93	0.93	0.92	0.93	0.94
0.8	0.88	0.84	0.80	0.89	0.90	0.89	0.90	0.86	0.89	0.89	0.90	0.90	0.90	0.90	0.91	0.92
1.0	—	—	—	—	0.87	0.87	0.91	0.85	0.89	0.90	0.88	0.90	0.89	0.89	0.90	0.92
1.2	—	—	—	—	0.85	0.84	0.91	0.84	0.88	0.86	0.88	0.90	0.89	0.88	0.90	0.92
1.4	—	—	—	—	0.85	0.84	0.91	0.84	0.88	0.86	0.87	0.89	0.89	0.88	0.89	0.91
1.6	—	—	—	—	0.86	0.84	0.93	0.83	0.87	0.85	0.86	0.89	0.88	0.87	0.89	0.91
1.8	—	—	—	—	0.83	0.77	0.91	0.78	0.86	0.78	0.85	0.88	0.90	0.86	0.88	0.91
2.0	—	—	—	—	—	—	—	—	—	—	0.85	0.89	0.90	0.87	0.88	0.92
2.2	—	—	—	—	—	—	—	—	—	—	0.84	0.88	0.86	0.85	0.87	0.91
2.4	—	—	—	—	—	—	—	—	—	—	0.84	0.88	0.86	0.86	0.87	0.90
2.6	—	—	—	—	—	—	—	—	—	—	0.82	0.87	0.84	0.84	0.87	0.89
2.8	—	—	—	—	—	—	—	—	—	—	0.76	0.83	0.75	0.76	0.87	0.90
3.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.86	0.91
3.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.85	0.90
3.4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.84	0.90
3.6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.81	0.90
3.8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.73	0.90
4.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.89
4.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.89
4.4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.89
4.6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.89
4.8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.88
5.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.86

‘—’ implies the threshold exceeded the maximum obtainable pixel value in the image

from the highest false alarm rate of 1.30. The detection rate of VIS was slightly lower at 0.93, but lowered the false alarm rate to 1.14. VIS-MWIR had a detection rate of 0.94, while lowering the false alarm rate to 1.14. MWIR-LWIR produced the lowest false alarm rate at 0.96 with a detection rate of 0.91. These presented fusion results demonstrate that fusing multiple spectral bands lowers false alarms while maintaining high detection rates.

The contribution to the overall detection by segmented objects and features is shown in Figure 2.7. The black bar indicates the rate by segmented detection and the gray bar is the additional detection rate by using the SIFT features. Detection by segmentation is the primary detection mechanism and contributes to the bulk of the detection rate, whereas feature detection is secondary and has a smaller impact on

Figure 2.6: Detection and false alarm rates for detecting moving targets using the labeled spectral combinations in the fusion algorithm.

the overall detection rate. Pixel texture varies in each spectral band, providing different spatial features that are independent of one another. Extracting features from different spectral bands provides additional features for tracking and identification. Single spectral bands did not have any detections by features, which I attributed to an insignificant number of features to match between the target database and the scene. NIR-MWIR had the highest contribution for detections by features at 0.040. I attributed this to the high number of false pixels that were detected by segmentation, incorporating features that belong to the background. VIS-MWIR-LWIR had the next highest contribution of feature detections with 0.028. The overall detection contribution by features is not significant for this data set, but provides a means to track targets in difficult situations such as busy intersections or partial occlusions when they would otherwise be lost.

Figure 2.7: Detection rate of scene targets using segmented objects from the foreground detection algorithm or features in the instance that no foreground object was detected.

Algorithm performance for estimating the targets true centroid by correctly detecting pixels that belong to ground truth objects will now be discussed. A correctly detected pixel is defined as belonging to a ground truth object, otherwise is classified as a false pixel. A pixel scoring example is shown in Figure 2.8. Figure 2.8 (a) is a ground truth vehicle in MWIR and Figure 2.8 (b) are labeled foreground pixels. Blue pixels represent true positives that belong to the ground truth object, and red pixels represent pixels that were falsely detected. For this example, false pixels are attributed to the vehicle shadow.

I define the pixel detection rate for the full video sequence as

$$pixel\ detection\ rate = \frac{\sum_{i=1}^N detected\ ground\ truth\ pixels}{\sum_{i=1}^N total\ ground\ truth\ pixels}, \quad (2.14)$$

Figure 2.8: Example showing how detected pixels are scored as true positives or false alarms. (a) is an example vehicle in MWIR. (b) illustrates detected foreground pixels; blue indicates true positive and red indicates false alarm.

where N is the number of frames. High pixel detection rates result in accurate estimates of target centroids, but falsely detected pixels can negatively affect the centroid calculation, resulting in less accurate results. The false pixel rate is measured per frame and presented as

$$false\ pixel\ rate = \frac{\sum_{i=1}^N false\ pixels\ detected}{P * N}, \quad (2.15)$$

where P is the number of non-target pixels in the frame and N is the number of frames.

Pixel detection rates and false pixel rates for all spectral bands and fusion combinations are shown in Figure 2.9. LWIR produced the highest pixel detection rate of 0.95 and a false pixel rate of 0.0007. The fusion combination VIS-MWIR produced a high pixel detection rate of 0.94, but suffered the highest false pixel rate of 0.0017.

Figure 2.9: Detection rates of pixels that belong to targets vs. rate of false pixels that do not belong to any targets.

TOT3 resulted in a detection rate of 0.89 and false pixel rate of 0.0003. Seven fusion combinations resulted in false pixel rates less than 0.0005.

Displacement error between detected objects and their respective ground truth centroids is a measure of how accurately an algorithm estimates the true centroid of the target. For target tracking, centroids are input to filters that predict future targets locations, i.e., Kalman filtering, which require accurate estimates. Displacements of detected targets over all frames were measured and the RMSE was calculated as

$$RMSE = \frac{\sqrt{\sum_{t=1}^N ((x - \hat{x})^2 + (y - \hat{y})^2)}}{N}, \quad (2.16)$$

where $(\hat{x}, \hat{y})$ are the coordinates of the measured centroid, (x, y) are ground truth centroids, and N is the number of detections. The results are presented in Figure 2.10.

Figure 2.10: Mean centroid displacement of detected targets for all detected targets for all frames.

NIR-LWIR had an RMSE of 0.07 meters, which was the lowest for all spectral combinations. LWIR had the next lowest RMSE at 0.08 meters whereas the other single spectral bands had errors greater than 0.3 meters. The fusion results presented highlight the centroid accuracy improvements made by fusing spectral bands as compared to using single bands.

2.5.2 Tracking

I now present the performance of the fusion algorithm to associate targets between scenes and create a tracking profile using the foreground combination map (VIS-LWIR). This weighted combination was chosen due to the low centroid error, along

with the high detection rate, and low false alarm rate. In this evaluation, the bottom 400 rows of pixels are not considered for the tracking results due to trees obstructing the view of the imaging sensors, preventing full vehicle segmentation. Sample images from the tracking sequence are shown in Figure 2.11. The yellow box indicates the track algorithm has used a segmented object to update the track location. A teal box indicates that no segmented object matched and SIFT features were used to update the track location.

The motion of 12 vehicles was simulated for an urban traffic environment. Tracks were initiated on all 12 vehicles during the video. Of those 12 vehicle tracks, 11 were tracked though the entire video sequence with no errors. Due to being idle for extended periods of time, one track was lost, but a new track was initiated after it initiated movement. There were no instances where track identities were switched between vehicles and only one instance of a false track. Three false tracks were produced, where two false tracks are attributed to the idle vehicle. The tracking results are summarized in Table 2.6.

Table 2.6

Results summary for the proposed multi-spectral fusion algorithm tracking moving vehicles through the full motion video sequence.

Total Vehicles	Full Tracks	Lost Tracks	Switch Tracks	False Tracks
12	11	1	0	3

Figure 2.11: Examples frames from the tracking sequence (Video 1, MPEG, 14 MB). (a) - Frame - 150, (b) Frame - 300, (c) Frame - 450, (d) Frame - 600

2.6 Conclusion

In this chapter I proposed an algorithm to fuse multi-spectral data sets to increase detection accuracy of a video tracker, while maintaining a high detection rate and low false alarms per frame. Previous works consider visible and LWIR data sets[30, 31, 32, 33, 34, 35]; I extend previous work to include NIR and MWIR. In these four spectral bands, I build a GMM to detect foreground pixels by modeling the time history of the pixels intensities. Foreground pixels from all spectral bands are weighted and fused into foreground maps, and formed into targets candidates. Target candidates are tracked through the frame sequence using SIFT features to track missed detections and uniquely identify targets. My proposed algorithm was tested on synthetically generated visible, NIR, MWIR, and LWIR imagery using the DIRSIG toolset. Compared to the single spectral band cases, the fused algorithm improves detection accuracy while improving detection rates and lowering false alarm rates for most spectral combinations. The detection results provided input to a video tracker that detected the 12 moving vehicles in the scene. Of those 12 targets, 11 were tracked with no failures, one vehicle showed track-loss, but this track was re-initiated, and three false tracks occurred.

Chapter 3

Target Localization and Tracking by Fusing Doppler Differentials from Cellular Emanations with a Multi-spectral Video Tracker

3.1 Abstract

I present an algorithm for fusing data from a constellation of RF sensors detecting cellular emanations with the output of a multi-spectral video tracker to localize and

track a target with a specific cell phone. The RF sensors measure the Doppler shift caused by the moving cellular emanation and then Doppler differentials between all sensor pairs are calculated. The multi-spectral video tracker uses a Gaussian mixture model to detect foreground targets and SIFT features to track targets through the video sequence. The data is fused by associating the Doppler differential from the RF sensors with the theoretical Doppler differential computed from the multi-spectral tracker output. The absolute difference and the root-mean-square difference are computed to associate the Doppler differentials from the two sensor systems. Performance of the algorithm was evaluated using synthetically generated datasets of an urban scene with multiple moving vehicles. The presented fusion algorithm correctly associates the cellular emanation with the corresponding video target for low measurement uncertainty and in the presence of favorable motion patterns. For nearly all objects the fusion algorithm has high confidence in associating the emanation with the correct multi-spectral target from the most probable background target.

3.2 Introduction

Detection and tracking of moving targets in cluttered urban environments is an important task for local law enforcement and security forces. The ability to locate and track a single target in a large cluttered scene is difficult due to several factors based on the target and its surrounding environment. For example, it is sometimes difficult

to differentiate between targets when using a thermal infrared (IR) sensor as target signatures are similar and thermal contrast between the target and the background can be low. Radio frequency (RF) communications contain identification information about the transmitting source but lack the ability to spatially localize the target with low uncertainty [47]. The combination of data from two or more sensors, referred to as sensor fusion, exploits the advantages of multiple sensors while overcoming the disadvantages of each individual sensor [3, 4]. The goal of this work is to exploit two different sensor modalities: a constellation of RF sensors, and a multi-spectral video sensor equipped with a target tracker, to uniquely identify and track a moving vehicle carrying a specific cell phone.

RF transmissions have been instrumental for decades in detection and tracking of targets using both active and passive systems. Passive RF systems exploit existing sources of opportunity such as cellular communications or television broadcasts. These systems have been demonstrated in a number of applications such as surveillance [48, 49], geolocation [50, 51], and motion estimation [52]. A cellular phone is a device that emits an RF signal and has been of interest for surveillance by government agencies including local law enforcement and the federal bureau of investigation [53]. Cellular phones can be tracked by devices known as “stringrays” that act as a cellular tower and intercept the cellular signal to localize and track a specific target [54]. Because cellular phones contain unique identifications, they are excellent sources to identify targets with high confidence.

Because of their generally high angular resolution, and in the case of infrared sensing, night vision capability, electro-optical/infrared (EO/IR) sensors are commonly used to identify and track a variety of targets including pedestrians [55, 56, 57], vehicles [19], ships [20], and aircraft [58]. Under optimal viewing conditions EO/IR sensors can measure the location of a target with high accuracy and precision making them an important asset for security systems. Algorithms to accomplish these tasks have been demonstrated making use of background estimation [37, 38, 39], edge detection [59], and feature recognition [60]. That said, it is the case that EO/IR sensors cannot see inside most vehicles. Hence, in a crowded traffic environment where a particular cell phone is being used, associating the cell phone emanation with a particular vehicle is an important problem, which is addressed here.

A number of sensor combinations have been developed to aid in target detection and tracking applications. Noulas *et al.* fuse audio segments with a video sequence to associate the audio with its corresponding video target [61]. Kilic *et al.* track speaking targets by fusing likelihoods built from audio and visual data [62]. D’Arca *et al.* fuse audio and visual sensors to estimate a targets trajectory using separate Kalman filters that get fused into a single Kalman filter [63]. Chin *et al.* demonstrate a fusion technique using an optical tracker and Wi-Fi to track a target through obscurations [64]. I explore a fusion algorithm to localize and track a specific vehicle using two new sensor types; a constellation of RF sensors capturing cell phone emanations and a multi-spectral imaging system. To the best of my knowledge the fusion of these

sensors is unique in the literature.

I present a novel combination of passive sensor data fusion by using a constellation of RF sensors measuring a cellular emanation from a specific phone with a multi-spectral imaging sensor detecting and tracking vehicles in a target rich environment. In practice, neither signal contains enough information to allow a particular vehicle to be uniquely identified as the source of the cellular emanations. However, by fusing these two sources of data I demonstrate that a specific target can confidently be identified and tracked through a sequence of frames. From the cellular emanation I make use of the frequency difference of arrival (FDOA), also referred to as the Doppler differential (DD), which is a result of relative motion between the emitter and separated receivers [65, 66]. The multi-spectral sensor produces centroid estimates of multiple moving vehicles through a sequence of frames. Constellations of unmanned aerial vehicles (UAV) have become readily available and demonstrated for various applications [67, 68, 69]; the notional sensor configuration studied here is a UAV scenario with a multi-spectral imaging sensor located at the scene origin at an altitude of 1000 meters and RF sensors spaced on the border of the imaging sensors field of view at an altitude of 1000 meters. The sensor-scene geometry is shown in Fig. 3.1, and the sensor coordinates shown in Fig. 3.2. The geometry of the RF sensors on the border of the scene is chosen to provide diversity to the DD of received signals; other geometries would work for this application.

Figure 3.1: Geometry of the multi-spectral and RF sensors in reference to the plane in which the moving vehicles lie.

Figure 3.2: Coordinates of the multi-spectral and RF sensors in reference to the plane in which the moving vehicles lie.

The block diagram of the fusion algorithm described here is shown in Fig. 3.3. The multi-spectral video tracker fuses images for detection of moving foreground objects which are then tracked through a video sequence [70], giving two dimensional time history of (x, y) centroid locations of multiple moving targets. Radial velocity estimates are computed from the tracker outputs and used to calculate the theoretical

Figure 3.3: Flowchart of the fusion processing to associate a specific moving target between a multi-spectral video tracker and a constellation of RF sensors.

DD which would have been observed at the cell phone frequency for each tracked target. The RF receivers in the constellation each measure incoming cellular emanations and isolate a signal of interest and extract the Doppler shift. The RF process for an individual sensor is shown in Fig. 3.4. It is not in the scope of this chapter to cover the Doppler shift estimation, but I note this is a viable operation in wireless communications to isolate a single RF signal [71] and estimate its Doppler shift [72]. DD are calculated for all combinations of RF sensors and the sensor pair corresponding to the maximum DD is used to associate the RF sensors with the multi-spectral tracker output. To associate the multi-spectral image tracker with the RF sensors, the absolute difference and root mean square difference (RMSD) is calculated and the sensor pair with the minimum metric is selected as the matching target. I compare the association rate of the sensors to the correct result to evaluate performance.

This algorithm was developed and evaluated using synthetically generated datasets.

Figure 3.4: Flowchart of RF sensor processing to isolate and extract Doppler shifts.

RF sensor measurements were simulated using the known ground truth radial velocity of the emanating target to generate Doppler shifts and varying levels of random measurement uncertainty were added. The imaging sensor is multi-spectral and includes visible, near-infrared (NIR), mid-wave infrared (MWIR), and long-wave infrared (LWIR) which were simulated using the Digital Imaging and Remote Sensing Image Generation (DIRSIG) software [18].

I present results on associating the RF sensors with the corresponding target from the multi-spectral video to localize and track a specific moving target with a cell phone. Using the absolute DD, the algorithm has a high rate of correctly associating the RF emanation with the multi-spectral target at low measurement uncertainty for targets that have motion patterns that are not similar to other targets. The RMSD improves the association rate particularly for low uncertainty cases, but maintains good performance through significant uncertainty levels. The confidence of identifying the correct multi-spectral target from background targets is high for low measurement uncertainty and remains high for over half of the targets through all uncertainty levels.

The remainder of the chapter organized as follows. Section 3.3 discusses the extraction of DD from RF sensors measuring a specific cellular emanation. Section 3.4 details the multi-spectral video tracker and the calculation of the theoretical DD. Section 3.5 discusses the metrics for associating the cellular emanation measured using the RF sensors with the output of the multi-spectral video tracking to localize a specific moving target. Experimental results for matching the cellular emanation as measured from the RF sensors with the video targets is discussed in Sec. 3.6. In Sec. 3.7 conclusions are presented.

3.3 Cell Phone Emanations

In this section I present background on sensing cellular emanations from multiple RF receivers.

I start by reviewing the Doppler shift that occurs due to radial motion of the transmitter. A transmitting target has a position (x_k^t, y_k^t, z_k^t) at time instance k . RF receivers are located at positions $(x_k^\ell, y_k^\ell, z_k^\ell)$, where ℓ is the receiver label. The range $R_k^\ell(\Delta x, \Delta y, \Delta z)$ between the transmitting target and a receiver ℓ is given

$$R_k^\ell(\Delta x, \Delta y, \Delta z) = \sqrt{(x_k^\ell - x_k^t)^2 + (y_k^\ell - y_k^t)^2 + (z_k^\ell - z_k^t)^2}. \quad (3.1)$$

Figure 3.5: Range of the moving ground targets from the video sensor for the given video sequence.

Shown in Figure 3.5 are the ranges between the multi-spectral sensor ($x_k^\ell = 0, y_k^\ell = 0, z_k^\ell = 1000$) and the ground target (GT) vehicles as a function of time for the interval covered by the DIRSIG simulation. For the geometry and pattern flow of this scenario the targets enter at the edge of the scene, move towards the intersection at the center, and proceed to move towards the edge of the scene.

The derivative of the range with respect to time produces the range-rate v , also known as the radial velocity, and given by

$$\begin{aligned}
v_k^\ell(\Delta x, \Delta y, \Delta z) &= \frac{\partial}{\partial t} R_k^\ell(\Delta x, \Delta y, \Delta z) = \frac{\partial}{\partial t} (R_{k2}^\ell - R_{k1}^\ell) \\
&= \frac{1}{T} \left(\sqrt{(x_{k2}^\ell - x_{k2}^t)^2 + (y_{k2}^\ell - y_{k2}^t)^2 + (z_{k2}^\ell - z_{k2}^t)^2} - \right. \\
&\quad \left. \sqrt{(x_{k1}^\ell - x_{k1}^t)^2 + (y_{k1}^\ell - y_{k1}^t)^2 + (z_{k1}^\ell - z_{k1}^t)^2} \right), \tag{3.2}
\end{aligned}$$

where T is the sampling period between range measurements. Figure 3.6 is the range-rate for the GT in the scenario presented here. A negative range-rate indicates that the receiver and target are getting closer in range, and conversely, a positive range-rate indicates that the pair are moving away from one another. With the sensor altitude z being large compared to the (x, y) displacement and the motion being confined to the x – y plane, the large ranges and low ground velocities result in low range-rates. The similarity of the traffic patterns is due to the geometry of the scene and the traffic scenario. This introduces difficulty in distinguishing between targets since the Doppler shifts will be small in general, and when vehicles stop, for example at traffic lights, it will disappear. By using spatially separated receivers I create diversity in the Doppler shifts and create the opportunity to use the highest Doppler shifts which will least likely be associated with non-moving targets.

I now examine the effect of the radial velocity on the frequency f_c of a carrier signal for a cellular emanation. A transmitter with a radial velocity in the direction of a stationary receiver results in a shift of frequency that makes it larger, whereas radial velocity in the opposing direction of the receiver results in a frequency shift that

Figure 3.6: Range-rate of the moving ground targets from the video sensor for the given video sequence.

makes it smaller. This is a well known concept called the Doppler effect. The shift in frequency $\Delta f = f_D - f_c$ is given by

$$\Delta f = f_c \left(1 + \frac{v_k}{c} \right), \quad (3.3)$$

where c is the speed of light. Shown in Fig. 3.7 are the Doppler shifts corresponding to the radial velocities from Fig. 3.6 for a carrier frequency of 1 GHz. For the 1 GHz carrier the Doppler shifts range between -4 and 4 Hz.

Similar motion profiles resulting from using a single sensor make it difficult to distinguish between targets. The use of a constellation of spatially separated receivers produces variation in the Doppler signature for the targets. For the purpose of this

Figure 3.7: Doppler Shift at 1 GHz carrier frequency for the moving ground targets as measured from location of the video sensor for the given video sequence.

study 8 RF receivers were placed on the edge of the imaging sensors field of view and one in the center; locations are shown in Fig. 3.2. The spatial distribution of the sensors is such that some will be in the direction of travel and range will be decreasing, resulting in a negative range-rate and a negative Doppler shift. The other sensors will be opposite the direction of travel and the range will increase, resulting in a positive range-rate and positive shift in frequency.

Figure 3.8 shows the range between an example transmitting target (GT #1) and the constellation sensor locations in Fig. 3.2. As indicated, some sensors are decreasing in range while others are increasing. This is better illustrated in Fig. 3.9 and Fig. 3.10 where the range-rate and Doppler shift for a 1 GHz carrier are shown, respectively.

Figure 3.8: Example ranges between a transmitting target (GT #1) and the locations of the RF sensor constellation.

At the beginning of movement for this example, the Doppler shift for the 1 GHz carrier has a shift near -8 Hz for 3 sensors and a shift near 0 Hz for 3 sensors, giving a difference of 8 Hz. That difference decreases as the target approaches the intersection with decreasing speed and eventually comes to a stop around 6 seconds, resulting in nearly 0 Hz difference in Doppler shift.

The DD ρ_k is defined as the difference in Doppler shift between receivers $\ell 1$ and $\ell 2$ and is given by

$$\rho_k = \Delta f_k^{\ell 1} - \Delta f_k^{\ell 2}. \quad (3.4)$$

The DD varies between RF sensor pairs based on their geometry and the radial velocity of the moving target. For example, one target may be stationary and have

Figure 3.9: Example range-rate between a transmitting target (GT #1) and the locations of the RF sensor constellation.

Figure 3.10: Example Doppler shifts for 1 GHz carrier frequency between a transmitting target (GT #1) and the locations of the RF sensor constellation.

Figure 3.11: Maximum Doppler differential for GT #1 using all RF sensor combinations (left Y-axis) vs. its corresponding ground speed (right Y-axis).

no DD, whereas another target may be moving radially with respect to the sensors which results in a near zero DD (but may be non-zero to other sensor pairs). The maximum DD selects the sensors that are positioned orthogonal to the targets motion. An example of the maximum DD is shown in Figure 3.11 for GT #1 with the corresponding ground speed. The target starts out with its highest DD when it first enters the scene and decelerates as it moves towards the intersection at the center of the scene and reaches a DD of 1 Hz. After reaching the intersection the target increases in speed and the DD increases to 8.3 Hz.

3.4 Multi-spectral Video Tracker

I present an overview of the detection and tracking algorithm for moving targets in the multi-spectral video sequence to produce track sequences in which the theoretical DD is calculated. This algorithm has been used successfully to detect and track targets in cluttered urban environments [70]. Spectral bands used were visible, NIR, MWIR, and LWIR and was developed using a synthetically generated dataset from the DIRSIG toolset [18]. Figure 2.2 shows a frame of each spectral band with a frame size of 2000×2000 and a ground sample distance of 0.0635 meters, resulting in a field of view covering 137×137 meters. A normal traffic scenario was simulated using the open source tool Simulation of Urban MObility (SUMO) to provide realistic traffic maneuvers [46]. By visual inspection, the appearance of target vehicles varies between the spectral bands, providing different intensity information.

Pixel intensities fluctuate due to noise, changes in illumination, and movement from both background and target objects. This does not allow a single value to characterize the time history of the intensity of a single pixel for a given video sequence. To compensate for these changes, background modeling techniques are used to describe the probability distribution of the pixels intensity by empirically deriving the parameters from the video sequence. The Gaussian mixture model (GMM) has been successfully demonstrated to compensate for the fluctuations in pixel intensities [37, 38, 39]. In a

Figure 3.12: Example multi-spectral frames from synthetic DIRSIG data set for (a) Visible, (b) NIR, (c) MWIR, and (d) LWIR spectral bands.

scene where the sensor is fixed, keeping the viewpoint stationary, I use statistical information extracted from the time history of the intensity fluctuations to understand the probability distribution of intensity at each pixel, and use these distributions to make hypotheses about the label of each pixel. Each pixel in the scene is classified as a foreground or background pixel, and I update the parameters of the GMM during

each frame.

The association of targets involves relating a track sequence from prior frames with target candidates detected in the current frame. This task is trivial in the case where targets stay separated and no occlusions exist. However, in actual practice and in this data set, targets can be merged, or occluded by trees, etc., making distinguishing between targets and maintaining the correct association of tracker data and target difficult. I have chosen to use SIFT features for identification due to their robustness with respect to changes in rotation and scale, and their invariance to change in camera viewpoints and illumination changes [15, 73]. SIFT features are composed of a keypoint that gives sub-pixel location and orientation of the feature, along with a descriptor that is calculated based on local pixel texture.

Results from the multi-spectral imaging system are a collection of tracked targets that have correlated (x, y) centroid measurements in the imaging plane. The root mean-square error (RMSE) between the true and estimated centroid locations for the video tracked objects are shown in Figure 3.13. Video targets 8 and 10 have the largest RMSE where the Y -error is over 0.3 meters and the X -error is over 0.15 meters. The other targets have a considerably lower RMSE in both X and Y .

Theoretical DD for the output targets of the multi-spectral tracker are calculated using Eqns.(3.1)-(3.4) where target positions x_k^t and y_k^t are ground coordinates from the measured video frames and z_k^ℓ is the ground height which is 0 meters for this

Figure 3.13: RMSE of the centroid estimate using the multi-spectral video tracker for the ground-truth video targets.

scenario. Sensor coordinates $(x_k^\ell, y_k^\ell, z_k^\ell)$ are the locations of the RF sensors.

3.5 Data Association

The data fusion process associates a specific cellular emanation measured by the constellation of RF sensors with the corresponding target in the output of the multi-spectral video tracker. A flowchart of the process to associate the two sensors at frame k is shown in Fig. 3.14. The multi-spectral video tracker produces estimates of the radial velocity computed from two dimensional tracker output which is used to calculate the theoretical DD between the tracked target and locations of the RF sensors. The RF constellation produces measurements of the DD from all combinations of RF receivers and the pair that maximizes the magnitude of the DD is selected. The difference in DD between the RF sensors and the multi-spectral video sensor is

calculated for all video tracked targets. I propose two metrics to associate the max DD from the RF sensor measurements with the theoretical DD from multi-spectral video tracker output; 1) the absolute difference ΔI_k and 2) the root-mean-square difference $RMSD_k$. In the case of metric 2) the difference in DD is stored in a database for the time averaged calculation. The absolute difference ΔI_k at frame k between the DD of RF sensors ρ_k^{RF} and the video tracker output ρ_k^{vid} is given

$$\Delta I_k = |\rho_k^{RF} - \rho_k^{vid}|. \quad (3.5)$$

This absolute difference is particularly well suited for systems that do not have adequate samples to produce a statistical average due to sparse measurements by either the RF or video sensors. The root-mean-square difference $RMSD_k$ is a time averaged difference in DD between the RF sensors and the video sensor given by

$$RMSD_k = \sqrt{\frac{\sum_{n=1}^N (\rho_n^{RF} - \rho_n^{vid})^2}{N}}, \quad (3.6)$$

where N is the number of averaged measurements. The $RMSD_k$ metric lowers false associations that are attributed to sporadic measurement error which may occur using ΔI_k .

Figure 3.14: The process of associating the DD at a given frame measured from a constellation of RF sensors with the output of a multi-spectral video tracker using two statistical metrics; 1) absolute difference and 2) root-mean-square difference.

3.6 Experiments

I present the performance of the fusion algorithm for associating a specific cellular emanation detected from a constellation of RF receivers with the corresponding target from the multi-spectral video tracker. An association is classified as correct when the cellular emanation from the RF sensors is correctly matched with the corresponding video target. Performance of the algorithm was evaluated using Doppler measurements from the RF receivers with added measurement uncertainty and (x, y) centroid locations from a multi-spectral video tracker. Doppler shifts for a cellular frequency were generated from the ground truth radial velocity with varying RMS levels of white Gaussian noise added to the Doppler shifts to model measurement

uncertainty. Centroid location information was extracted from a multi-spectral video set that was developed with the DIRSIG software tool [18] with the detection and tracking algorithm presented in [70]. The results are the correct association rates over the video sequence for 100 Monte Carlo simulations.

Figures 3.15 and 3.16 show the association rates using the absolute difference metric. With no uncertainty in the Doppler measurement targets 1, 9, and 11 have association rates above 0.96. Target 1 maintains a rate of 0.92 through 1 Hz of RMS uncertainty and has a gradual decrease to 0.52 at 10 Hz RMS uncertainty. The association rate for target 11 drops to 0.84 with 1 Hz RMS uncertainty, and has a slow decrease through 10 Hz which has a rate of 0.51. Target 2 has a steeper decrease in rate and drops to 0.7 with 1 Hz RMS uncertainty and decreases to 0.40 with 10 Hz. Remaining targets have an association rate greater than 0.60 with no uncertainty. Targets 4 and 6-8 have the poorest performance as they drop below 0.5 with 0.2 Hz RMS uncertainty. Overall the results with no Doppler uncertainty are promising, but with added uncertainty there is a significant decrease in performance.

Figures 3.17 and 3.18 show the association rates using the RMSD. With no added Doppler uncertainty all targets start off with an association rate above 0.82 excluding target 4. Targets 1, 3, 5, 9, and 11 have perfect association rates with 7 of the 11 targets being 0.92 or greater. Targets 1 and 11 have a gradual decrease in association rate and maintain a rate greater than 0.83 through 10 Hz of RMS uncertainty. Target

Figure 3.15: Correct association rate for matching specific cellular emanation measured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the absolute DD.

Figure 3.16: Correct association rate for matching specific cellular emanation measured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the absolute DD.

5 has an association rate greater than 0.9 through 4 Hz and gradually decreases to 0.59 at 10 Hz. Target 2 has a rate above 0.92 through 0.5 Hz of RMS uncertainty and gradually decreases to below 0.51 at 5 Hz. Targets 6 and 7 drop below 0.62 at 0.4 Hz of RMS uncertainty due to similar Doppler signatures. Target 10 has a rate above 0.77 through 0.7 Hz RMS uncertainty but begins to decrease significantly. This target has poor performance in the video tracker and a similar DD signature as target 2. Target 8 has the most immediate decrease in performance by dropping below 0.44 at 0.3 Hz due to a similar Doppler signature as target 4. There is an improvement in performance for some targets as uncertainty is increased, particularly target 4 when increasing from no uncertainty through 0.3 Hz. This is attributed to similar Doppler signatures for a particular sensor configuration. With added uncertainty a different sensor pair produces a higher DD, and that sensor pair proves to have better performance with the multi-spectral video tracker. For example, with no RMS uncertainty target 4 has multiple sensor pairs with nearly identical DD, with a slightly higher DD for Sensor 1 and Sensor 9. As RMS uncertainty is added, this optimal sensor pair begins to switch between different pairs (i.e. Sensor 1 & Sensor 9, Sensor 4 & Sensors 6, Sensor 6 & Sensor 7). These new sensor pairs produce DD for target 2 that are not similar to target 4, reducing the incorrect associations. Overall, these results are improved in comparison to the absolute difference and provide robustness to measurement uncertainty.

Figure 3.17: Correct association rate for matching specific cellular emanation measured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the RMSD DD.

Figure 3.18: Correct association rate for matching specific cellular emanation measured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the RMSD DD.

The confidence ratio CR is a performance measure of confidence for correctly identifying the true video target from the most probable incorrect target after all frames measurements have been made. It is defined as the normalized difference ratio between the number of correct associations $Corr$ and the number of associations for the highest detected incorrect target Bkg and given by

$$CR = \frac{Corr - Bkg}{Corr + Bkg}. \quad (3.7)$$

A positive CR detects the correct target more than an incorrect target, where a maximum value of 1.0 indicates that the correct association was made for all frames. A negative CR detects an incorrect target more than the correct target, where a value of -1.0 indicates that the incorrect association was made for all frames. This metric gives us a value on how likely the algorithm is to differentiate the true target from the most probable incorrect target.

Figures 3.19 and 3.20 show the confidence ratios. Targets 1, 3, 5, 9, and 11 have an CR of 1.0 with no Doppler uncertainty. Targets 1, 5, and 11 maintain a CR greater than 0.75 through 10 Hz RMS uncertainty while target 3 maintains a ratio greater than 0.6. The CR for target 6 and 7 becomes negative at 0.5 Hz of RMS uncertainty, but becomes positive again at 3 Hz. The increase in CR is attributed to Bkg becoming distributed between multiple incorrect targets, resulting in a lower Bkg for the single target. Due to the similarity with another target at the beginning of

Figure 3.19: Confidence ratio for matching specific cellular emanation measured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the RMSD DD.

the time interval, the CR for target 4 begins low but after 0.1 Hz of RMS uncertainty it increases above 0.66 for all levels of uncertainty. Target 10 has the lowest CR at 1 Hz RMS uncertainty and remains low for all levels. Target 10 has the highest RMSE in the video tracker and has a similar DD signature to target 2. For targets 1-5 and 11 the confidence is positive for all uncertainty levels, indicating the fusion algorithm can successfully associate the RF emanations with the corresponding target from the multi-spectral video tracker through all uncertainty levels.

Figure 3.20: Confidence ratio for matching specific cellular emanation measured by the constellation of RF receivers with the corresponding target from the multi-spectral video tracker using the RMSD DD.

3.7 Conclusion

Detection and tracking of moving targets is an important task to local and national security forces to monitor cluttered urban environments. Individual sensors are limited in their ability to confidently identify and localize a specific target in cluttered environments. The fusion of disparate sensors has been proposed to overcome limitations that individual sensors have when operating independently. In this chapter, I proposed an algorithm to fuse data from a constellation of RF receivers measuring Doppler shifts from a specific cell phone with the output from a multi-spectral video tracker. This work is unique from other fusion literature in that it fuses a new set

of sensors to localize and track a specific moving target. The constellation of RF sensors measure the Doppler shift from a cellular emanation in a specific moving vehicle and the DD between all sensor pairs is calculated. The multi-spectral video tracker uses a GMM to detect foreground objects and SIFT features to track them, and produces (x, y) centroid locations of detected vehicles. The specific target is localized by associating the DD from the RF sensors with the theoretical DD calculated from the multi-spectral video tracker by comparing the DD using two metrics; the absolute and RMSD DD. Using synthetically generated data, results demonstrate the fusion algorithm successfully associates cellular emanations with their corresponding target in a multi-spectral video tracker, but measurement uncertainty and motions patterns affect the correct association rate. The confidence of identifying the correct multi-spectral target from the most probable background target is high for low measurement uncertainty and remains high for over half of the targets through all uncertainty levels.

Chapter 4

Multi-spectral image

shift-estimation error calculations

using simulated phenomenology

4.1 Abstract

Registration of multi-spectral imagery is a critical pre-processing step for applications such as image fusion, but phenomenological differences between spectral bands can lead to significant estimation errors. To develop credible requirements for multi-spectral imaging systems, it's critical to characterize errors, both algorithmic and

fundamental, associated with estimating registration parameters; however, attempting to quantify error using archival data sets poses a number of problems. In this chapter, I demonstrate the use of commercially-available graphics software and available optical properties measurements to create fully-synthetic, multi-spectral imagery with high-fidelity representations of emissive and reflective phenomenology. I discuss and demonstrate techniques needed to quantify error for both area- and feature-based algorithms. I further show that such synthetic data sets can be used to quantify both the Fisher Information and sample errors associated with estimation of the shift between images acquired in different spectral bands; and, by extension, estimation of registration model parameters. With the flexibility offered by synthetic data, such characterization can be obtained for robust domains of image brightness, sensor parameters, and differences in image phenomenology.

4.2 Introduction

Registration is a fundamental image processing operation, critical for such applications as super-resolution [74], astronomy [75], and medical imaging [76]. Image registration is the process of estimating parameters for the geometric transformation between the coordinate systems of two images. The model for the transformation may include translation, rotation, distortion, and scaling. Mis-registration is caused by factors that include non-common aberrations or perspective differences between

sensors, jitter, or misalignment.

Algorithms for image registration can be grouped into two general categories; feature- and area-based methods. Feature-based algorithms allow high-order transformations to be estimated by matching edges, corners, and other points of contrast according to parametric “descriptors.” Harris corners [77] and scale-invariant feature transform (SIFT) [15] features are examples often used for image registration. Area-based algorithms maximize a similarity measure between a reference and shifted image to estimate some function of the transformation parameter. Example similarity measures include sum-of-square difference [78], correlation [79], and mutual information [80]. I note that deep learning techniques are beginning to emerge for use in image registration [81, 82]; due to the requirement of potentially massive amounts of training data – and, for my purposes, multi-spectral training data – I do not consider them here.

Imaging in multiple spectral bands generally yields more information about an object than a single image: For example, images in the visible/near-infrared (VNIR) spectrum ($0.4\ \mu m$ - $1.0\ \mu m$) encode information about the spectral reflectivity of features in the field of view. Without a source of illumination, however, VNIR image contrast is significantly reduced or non-existent. Images formed in long-wave infrared (LWIR) bands ($8\ \mu m$ - $14\ \mu m$), however, encode information about the emissivity properties of the object, independent of illumination.

These very band-to-band differences, unfortunately, make multi-spectral image fusion [1, 83, 84, 85] an especially demanding application of registration. The same objects or features in a given field of view generally have different intensities in different spectral bands. Contrast can even be inverted between images when one image maps primarily reflected phenomena and the other maps primarily emissive phenomena. For this reason, registration of multi-spectral and multi-mode (e.g., electro-optical and synthetic aperture radar) images remains an area of active research.

Images of a particular field recorded in different spectral bands will have different distributions of intensity and contrast, introducing bias (often significant) in both area and feature methods. Various approaches have been developed to improve area similarity measures between multi-spectral images such as correlation normalization [79] and mutual information [80]. In multiple studies, the original SIFT descriptor presented by Lowe [15] yielded incorrect feature matches for multi-spectral images [86, 87, 88]. Modified descriptors based on the orientation of edge contours have been developed to improve performance for multi-spectral imagery [87, 89, 90].

To develop imaging system requirements and understand registration-imposed limits to the information content of fused multi-spectral images, development of quantitative evaluation methods is as important as development of the algorithms themselves. To address this need, I have exploited advances in physics-based computer animation and graphics, along with parallel computing, to develop algorithm evaluation methods for

a variety of image processing applications.

Estimation bias and variance for registration model parameters should be compared not only among various approaches, but also against algorithm-independent fundamental limits. The Cramer-Rao lower bound (CRLB) is a widely-used tool in engineering and estimation theory and is the theoretical minimum variance for unbiased estimation given noisy data [91]. The CRLB is a useful tool for system development, quantifying achievable performance independent of any algorithm. For a vector parameter, the unbiased CRLB (UCRLB) for each element is given by the associated diagonal elements of the inverse of the Fisher information matrix (FIM).

While registration models can include several parameters, the simplest model - an isometric, non-rotational (or “rigid”) motion between images along arbitrary but orthogonal axes - is easily generalized to higher-order models by dividing the images to be registered into relatively small sub-fields and estimating the shift parameters between related sub-fields. With the shift between each pair of sub-fields representing a degree of freedom in a parametric model, a number of shift estimates can be used to estimate a higher-order transformation by generating an interpolation function [92, 93].

Robinson and Milanfar [94] developed a CRLB for shift estimation with images corrupted by additive noise. Tyler and Dank [95] developed and characterized a shift-estimate CRLB generalized to include Poisson noise and the noise-correlating effects

of using a linear filter on images. Tyler [96] further generalized the FIM developed in [95] to account for intrinsic bias, or the shift-estimation bias due to the above-mentioned intensity and contrast differences between multi-spectral and polarization image pairs. Intrinsic bias is independent of the estimation algorithm, and, like Fisher information, a property of the image data.

In this chapter, I exploit advances in synthetic phenomenology and computer graphics to demonstrate methods for quantitative comparison of various feature- and area-based image registration algorithms. Using synthetic imagery, among other advantages, means the true values of the shift parameters are controlled, allowing comparison methods to be framed in terms of estimation-theoretical metrics; including estimator bias and variance, bias gradient, and error cumulative distribution functions (CDFs). Because noise is represented formally in Fisher information, use of synthetic imagery also allows me to simulate noise-free images to compare algorithm errors to variances derived from relevant FIMs. To the best of my knowledge, this work is the first demonstration of the following:

- a)* The use of synthetic architecture and measured optical properties to develop test data sets for multi-spectral algorithm evaluation,
- b)* Comparison of ensemble-estimates of shift-estimation error to an algorithm-independent measure calculated using a unique Fisher information approach,

- c)* Comparison of shift-estimation algorithms using the bias gradient, including a technique for calculating an estimate of the bias gradient, and
- d)* The exploitation of parallel computing to develop “error maps” for both area- and feature-based algorithms, allowing intuitive association between estimation error and spatial features and phenomenology in the image.

The remainder of the chapter is organized as follows. Section 4.3 discusses feature- and area-based image registration techniques. Section 4.4 presents the error bounds on image registration. Section 4.5 I present a set of metrics for evaluating area- and feature-based shift estimation algorithms. Section 4.6 discusses generation of multi-spectral synthetic imagery used for algorithm evaluation. Numerical calculations are presented in Sec. 4.7. In Sec. 4.8 I present conclusions.

4.3 Image Registration Algorithms

4.3.1 Area-based

Area-based algorithms estimate shift by maximizing a similarity measure between a template and reference image using information in the spatial or frequency domains. Several area-based algorithms use a functional of the cross-correlation between two

images as the similarity metric. An efficient implementation of a cross correlation method using the discrete-Fourier transform (DFT) was presented by Guizar-Sicairos *et al.* [16]; an estimate of the cross-correlation peak is performed with an up-sampled image and refined using a non-linear optimization conjugate-gradient routine. Increased performance is accomplished using the DFT performed (over the FFT) over a small region of interest. Evangelidis and Psarakis presented a method to mitigate photometric distortions for template registration using an enhanced correlation coefficient (ECC) [97]. The presented similarity measure is an L_2 norm where intensities in both the reference and template image are normalized.

Mutual information (MI) is a similarity metric that matches the statistical dependence between the template and reference image [98, 99, 100]. MI is not directly dependent on the image intensities, making it useful for multi-spectral image registration. Mattes *et al.* [99] present a method for sub-pixel registration using MI; histograms of the reference and shifted image are represented as continuous functions using B-splines and a joint probability distribution function between the two images is estimated.

Area-based algorithms have been explored for registration of multi-modal imagery. Maes *et al.* used MI to register multi-modal images for medical applications that include MRI, CT, and PET images [98]. The authors present sub-voxel accuracy for rigid-shift estimation. Yun-hui registered optical and synthetic aperture radar (SAR) imagery using a normalized cross correlation on gradient imagery; registration

precision for the presented algorithm falls within 2-5 pixels of the true shift [101].

4.3.2 Feature-based

Image registration using feature-based methods has been demonstrated to yield sub-pixel accuracy for same-band imagery [15]. Feature-based registration algorithms detect points of interest, referred to variously as “keypoints,” “tie-points,” or “control points,” through various algorithms and are generally associated with relatively high-contrast features (e.g., peaks, edges, corners, etc.). A popular keypoint is the scale-invariant feature transform (SIFT) , which is calculated using difference-of-Gaussian filters at multiple scales to find regions of high gradient. Once a keypoint is detected, an identifier for associating keypoints between images is formed using surrounding pixel information, referred to as the descriptor. A similarity metric is used to match feature descriptors between images, resulting in matched locations used in high-order transformation models.

In the SIFT descriptor proposed by Lowe, a histogram is created using a 4x4 bin region surrounding the keypoint which is further divided into 8 sub-bins composed of the magnitude of the gradient. The 128 histogram values are merged into the descriptor vector. The original Lowe descriptor has exhibited poor performance in

multi-spectral imagery due to incorrect feature matching [86, 87, 88]. Modified descriptors have been explored to reduce error of keypoint matching: among these are Edge Histogram Descriptor (EHD) [87], Phase Congruency Edge Histogram Descriptor (PCEHD) [89], and Log-Gabor Histogram Descriptor (LGHD) [90]. Rather than constructing a histogram based on the gradient magnitude, these approaches use the orientation of edge contours to describe a keypoint, where each algorithm differs in the method to calculate the edge contour; EHD uses a Canny edge detector, PCEHD uses phase congruency, and LGHD uses Log-Gabor filters.

Feature-based algorithms have been demonstrated on additional multi-modal data types. Ma *et al.* presented registration of facial imagery in visible and thermal bands by matching edge-based features with regularized Gaussian field criterion [102]. The authors demonstrate the proposed algorithm registers non-rigid transformations of facial structures to under 3 pixels for a number of image sets. To overcome intensity variations in multi-modal retina images, Ma *et al.* present an approach to match edge-based features using a feature guided Gaussian mixture model and semi-supervised expectation maximization [103]. Toews *et al.* presented a modified SIFT descriptor based on locally inverted intensity information using MRI and CT images [104]. The authors present results that effectively register a multi-modal brain dataset. Chen *et al.* developed a feature-based method for registration of SAR and optical imagery [105]. Keypoints are detected based on phase congruency and the descriptor is a histogram of oriented gradients calculated from a Gaussian-Gamma-shaped bi-window

gradient operation. Matching of feature points using the proposed framework outperforms a number of existing features including SIFT.

4.4 Fisher Information Error

4.4.1 Cramer-Rao Lower Bound

The mean squared error (MSE) for estimation of a scalar parameter θ is given

$$\begin{aligned}
 MSE(\tilde{\theta}) &= E \left[\left(\tilde{\theta} - \theta \right)^2 \right] \\
 &= E \left[\left(\tilde{\theta} - E \left[\tilde{\theta} \right] \right)^2 \right] + \left(E \left[\tilde{\theta} \right] - \theta \right)^2 \\
 &= \sigma_{\tilde{\theta}}^2 + b_{\tilde{\theta}}^2
 \end{aligned} \tag{4.1}$$

where $\tilde{\theta}$ is the estimated parameter, $\sigma_{\tilde{\theta}}^2$ is the variance, and $b_{\tilde{\theta}}$ is the bias.

Estimator bias is the difference between the mean of the estimate and the true parameter value, defined as

$$b_{\tilde{\theta}} = E \left\{ \tilde{\theta} \right\} - \theta. \tag{4.2}$$

A discussion of bias sources in gradient-based registration algorithms was presented by Robinson and Milanfar [94]. The authors present derivations for bias due to Taylor

series truncation and approximations of the image gradient.

Any practical estimator is biased. If the bias for any particular algorithm depends on the true value of the (unknown and random) parameter to be estimated, this dependence adds to the variance according to the inequality [106]

$$\mathbf{C}_b(\tilde{\boldsymbol{\theta}}) \geq (\mathbf{I} + \nabla \mathbf{b}_{\tilde{\boldsymbol{\theta}}}) \mathbf{F}^{-1} (\mathbf{I} + \nabla \mathbf{b}_{\tilde{\boldsymbol{\theta}}})^T, \quad (4.3)$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{F}$ is the FIM, and $\nabla \mathbf{b}_{\tilde{\boldsymbol{\theta}}}$ is the bias gradient matrix, with elements

$$\nabla b_{ij} = \frac{\partial}{\partial \theta_j} b_i = \frac{\partial}{\partial \theta_j} \left[E \left\{ \tilde{\theta}_i \right\} - \theta_i \right]. \quad (4.4)$$

Note the biased CRLB degenerates to the inverse of the FIM if the bias is constant, independent of the actual value of the parameters to be estimated. In this case,

$$\mathbf{C}(\tilde{\boldsymbol{\theta}}) \geq \mathbf{F}^{-1}. \quad (4.5)$$

This makes intuitive sense, because if the bias is constant, it can be calibrated experimentally and removed.

The bias gradient is generally algorithm dependent, therefore it must be calculated for each algorithm. For two-parameter estimation, such as shift estimation, the bias

gradient matrix $\nabla \mathbf{b}$ is defined

$$\nabla \mathbf{b} = \begin{pmatrix} \nabla b_{ii} & \nabla b_{ij} \\ \nabla b_{ji} & \nabla b_{jj} \end{pmatrix}. \quad (4.6)$$

Formally, the FIM is the expected value of the curvature of the relevant log-likelihood function

$$\begin{aligned} \mathbf{F}(\Delta \mathbf{r})_{xy} &= -E \left[\frac{\partial \ln \mathbf{P}(\mathbf{s}, \Delta \mathbf{r})}{\partial \Delta \mathbf{r}_x} \frac{\partial \ln \mathbf{P}(\mathbf{s}, \Delta \mathbf{r})}{\partial \Delta \mathbf{r}_y} \right] \\ &= -E \left[\frac{\partial^2 \ln \mathbf{P}(\mathbf{s}; \Delta \mathbf{r})}{\partial \Delta \mathbf{r}_x \partial \Delta \mathbf{r}_y} \right]. \end{aligned} \quad (4.7)$$

where $\mathbf{s}$ is the reference image, $\Delta \mathbf{r}_{\mathbf{x}(\mathbf{y})}$ are random shift parameters, and $\Delta \mathbf{r}$ is the random shift vector. The curvature is with respect to the parameter to be estimated. If the expected curvature of the log-likelihood is high, relatively few values of the parameter are consistent with any measurements, and the information associated with those measurements is also relatively high. Conversely, relatively low values of the expected curvature indicate measurements will have relative low information since many values of the parameter to be estimated will be consistent with data. The

likelihood function for shifted images with normally distributed noise is given by [95]

$$\begin{aligned} \mathbf{P}(\mathbf{s}, \Delta\mathbf{r}) &= \frac{1}{2\pi[\det\{\mathbf{C}(\mathbf{s}; \mathbf{s}'(\Delta\mathbf{r}))\}]^{1/2}} \\ &\times \exp\left\{-\frac{1}{2}[\mathbf{s} - \mathbf{s}'(\Delta\mathbf{r})]^T \mathbf{C}^{-1}(\mathbf{s}; \mathbf{s}'(\Delta\mathbf{r})) [\mathbf{s} - \mathbf{s}'(\Delta\mathbf{r})]\right\} \end{aligned} \quad (4.8)$$

where $\mathbf{s}'$ is the measured shifted image and $\mathbf{C}$ is the covariance matrix of the measured data. The general FIM for a normally distributed random variable was derived by Kay [91] and given

$$\begin{aligned} \mathbf{F}(\Delta\mathbf{r})_{xy} &= \left[\frac{\partial\mathbf{s}}{\partial\Delta\mathbf{r}_x}\right] \mathbf{C}^{-1} \left[\frac{\partial\mathbf{s}}{\partial\Delta\mathbf{r}_y}\right] \\ &+ \frac{1}{2}\text{tr}\left\{\mathbf{C}^{-1}\frac{\partial\mathbf{C}}{\partial\Delta\mathbf{r}_x}\mathbf{C}^{-1}\frac{\partial\mathbf{C}}{\partial\Delta\mathbf{r}_y}\right\}. \end{aligned} \quad (4.9)$$

where “tr” is the trace operator.

4.4.2 Shift-estimation Error for Multi-spectral Imagery

Shift estimation for multi-spectral imagery is considerably more difficult due to phenomenological differences in the data not attributed to geometric transformation, referred to as intrinsic bias. The single band Fisher information must be extended to account for these differences. Tyler [96] generalized the tie-point shift estimation (TPSE) FIM in [95] to treat images that differ due to the phenomenology of sources

in the field of view, such as color or polarization. Tyler referred to phenomenology-related contrast differences between images as “intrinsic bias” to distinguish this source of error from algorithm bias. The modified FIM is the difference between the mean of FIMs of the images to be registered and the FIM of the difference between the two images. Tyler shows this FIM has the expected and desired properties, including

† If the two images are identical, $\hat{\mathbf{F}}$ reduces to the conventional FIM; and

† If one image has no features, or if the two images have no spatial derivatives at common coordinates (the derivative images are ”orthogonal”), all elements of the resulting $\hat{\mathbf{F}}$ are negative definite.

Note that since information is a non-negative quantity, the second property means for these limiting scenarios there is no information with which to register the images, as expected. A practical consequence of this property, as noted in Sec. 4.7, is that FIM elements can be negative when the contrast between images is sufficiently different. The modified FIM, which exploits the mean of the FIM from each image and the Fisher information in a given difference image, is given by

$$\hat{\mathbf{F}} = \frac{1}{2} [\mathbf{F}_1 + \mathbf{F}_2] - \mathbf{F}_d, \quad (4.10)$$

or

$$\begin{aligned}
\widehat{F}_{ij} = & \frac{1}{2} \left(\left[\frac{\partial \mathbf{s}_1}{\partial r_i} \right]^T \mathbf{C}_1^{-1} \left[\frac{\partial \mathbf{s}_1}{\partial r_j} \right] + \frac{1}{2} \text{tr} \left\{ \mathbf{C}_1^{-1} \frac{\partial \mathbf{C}_1}{\partial \Delta r_i} \mathbf{C}_1^{-1} \frac{\partial \mathbf{C}_1}{\partial \Delta r_j} \right\} \right) \\
& + \frac{1}{2} \left(\left[\frac{\partial \mathbf{s}_2}{\partial r_i} \right]^T \mathbf{C}_2^{-1} \left[\frac{\partial \mathbf{s}_2}{\partial r_j} \right] + \frac{1}{2} \text{tr} \left\{ \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \Delta r_i} \mathbf{C}_2^{-1} \frac{\partial \mathbf{C}_2}{\partial \Delta r_j} \right\} \right) \\
& - \left[\frac{\partial(\mathbf{s}_1 - \mathbf{s}_2)}{\partial r_i} \right]^T \mathbf{C}_d^{-1} \left[\frac{\partial(\mathbf{s}_1 - \mathbf{s}_2)}{\partial r_j} \right] \\
& - \frac{1}{2} \text{tr} \left\{ \mathbf{C}_d^{-1} \frac{\partial}{\partial \Delta r_i} [\mathbf{C}_1 - \mathbf{C}_2] \mathbf{C}_d^{-1} \frac{\partial}{\partial \Delta r_j} [\mathbf{C}_1 - \mathbf{C}_2] \right\}, \tag{4.11}
\end{aligned}$$

where $r_{i(j)}$ are shift directions, $\mathbf{s}_{1(2)}$ are the mean signal vectors (the first or second image in the pair), $\mathbf{C}_{1(2)}$ are image covariance matrices, and $\mathbf{C}_d$ is the covariance matrix for the difference of the two images; that is, $\mathbf{C}_d = \mathbf{C}_1 + \mathbf{C}_2$.

Tyler did not show that the diagonal elements of $\widehat{\mathbf{F}}^{-1}$ are, in fact, a lower bound for the case of multi-spectral data; however, the resulting variances can be used as an error metric and a measure of intrinsic bias. In Sec. 4.7, I quantify the biased and unbiased Fisher information error (FIE) as the root of the trace (RT) of the 2×2 FIM inverse.

4.5 Evaluation Metrics

4.5.1 Tie-point Shift Estimation Error

TPSE error is calculated for a given sub-field by performing the algorithm over a statistically-significant number of independent noise realizations. From these runs, the elements of the diagonal bias matrix $\mathbf{B}$ are given by

$$B_{ii} = \frac{1}{N} \sum_{n=1}^N \left(\hat{\theta}_i - \theta_i \right) \left(\hat{\theta}_i - \theta_i \right)^T \quad (4.12)$$

and covariance $\mathbf{C}$, with elements

$$C_{ij} = \frac{1}{N} \sum_{n=1}^N \left((\hat{\theta}_i - \mu_i)(\hat{\theta}_j - \mu_j) \right) \quad (4.13)$$

where N is the number of independent algorithm trials. The bias in X and Y are combined to form the root-sum of squares bias (RSSB) given

$$RSSB = \sqrt{\text{tr} \{ \mathbf{B} \}}. \quad (4.14)$$

The variances combine into the root-sum variance (RSV) given

$$RSV = \sqrt{\text{tr}\{\mathbf{C}\}}. \quad (4.15)$$

These two metrics are combined into the root-mean square error (RMSE) given

$$RMSE = \sqrt{\text{tr}\{\mathbf{B} + \mathbf{C}\}}. \quad (4.16)$$

4.5.2 Bias Gradient

As shown in Eq. 4.3 and discussed in Sec. 4.4.1, bias dependence on the actual shift – bias gradient – always increases the TPSE FIE and will generally increase estimator variance; thus, the bias gradient is a valuable algorithm comparison metric. The gradient is algorithm-specific and usually impractical to express formally. For the calculations below, I approximate the bias gradient by calculating the bias over ensembles of noisy image pairs. For each image pair in the ensemble, algorithm shift estimates are generated for a symmetric array of shift values. From these, true shift is subtracted to form an array of sampled bias values $\mathbf{b}$, where

$$\mathbf{b} = \begin{bmatrix} b_{ij}(\Delta s_{i-1}, \Delta s_{j-1}) & b_{ij}(\Delta s_i, \Delta s_{j-1}) & b_{ij}(\Delta s_{i+1}, \Delta s_{j-1}) \\ b_{ij}(\Delta s_{i-1}, \Delta s_j) & b_{ij}(\Delta s_i, \Delta s_j) & b_{ij}(\Delta s_{i+1}, \Delta s_j) \\ b_{ij}(\Delta s_{i-1}, \Delta s_{j+1}) & b_{ij}(\Delta s_i, \Delta s_{j+1}) & b_{ij}(\Delta s_{i+1}, \Delta s_{j+1}) \end{bmatrix}, \quad (4.17)$$

where $(\Delta s_i, \Delta s_j)$ indicate the shift for which the gradient is to be calculated. The gradient is calculated by operating on the bias array with a Prewitt filter, defined for the X -direction as

$$G_X = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad (4.18)$$

and $G_Y = G_X^T$ for Y -direction. The bias gradient matrix in Eqn. (4.4) is calculated for any particular shift by applying the gradient operator to the bias matrix at that shift.

I also calculate the increase in the FIE from the bias gradient. The increase in variance due to bias gradient is the difference between the trace of the biased FIE (BFIE) and unbiased FIE (UFIE):

$$C_{BG} = \text{tr} \left\{ \widehat{\mathbf{C}}_b(\tilde{\boldsymbol{\theta}}) \right\} - \text{tr} \left\{ \widehat{\mathbf{C}}(\tilde{\boldsymbol{\theta}}) \right\}, \quad (4.19)$$

where $\widehat{\mathbf{C}}$ is $\widehat{\mathbf{F}}^{-1}$ from Eq. 4.10 and

$$\hat{\mathbf{C}}_b(\tilde{\boldsymbol{\theta}}) = (\mathbf{I} + \nabla \mathbf{b}_{\tilde{\boldsymbol{\theta}}}) \hat{\mathbf{F}}^{-1} (\mathbf{I} + \nabla \mathbf{b}_{\tilde{\boldsymbol{\theta}}})^T, \quad (4.20)$$

4.5.3 Cumulative Distribution Function

Feature-based algorithms estimate shift between image fields from the difference in coordinates of features determined by the algorithm to be common to both fields, or “matched.” With noise-corrupted data, more matched features generally means reduced variance; however, bias can be significant in multi-spectral shift-estimation, even with a large number of detected features. The cumulative distributive function (CDF) can be a useful as a measure of the feature coordinate error. The CDF $H(x)$ for a discrete random variable X with probability density $h(x)$ is given by

$$H(x) = P(X \leq x) = \sum_{t \leq x} h(t). \quad (4.21)$$

I construct a CDF for featured-based algorithms by executing the algorithm over an ensemble of noisy image pairs with zero shift and accumulating the difference between estimated coordinates for all algorithm-matched features.

I construct a CDF for area-based algorithms using the sample-based RMSE for all sub-fields in the image frame. The CDF is useful for determining the number of tie-points that exceed a given error threshold. Further, while the CDF does not allow one

to associate bias or variance with spatial features, it provides an easily interpreted measure of error over the full field of view.

4.6 Synthetic Multi-spectral Imagery

Synthetic imagery is a valuable asset for evaluating algorithm performance for any image processing task, including for reasons beyond that given in the Introduction: First, noise is represented formally in the FIM calculations, so only synthetic imagery or very-high SNR archival data can be used. Second, removal of optical blurring and sampling effects of archival imagery results in undesirable artifacts. Finally, objects in the archival image fields may be undesirable and impossible to remove, such as shadows inconsistent with desired model illumination conditions. Following the approach in Tyler and Dank [95], I generate synthetic multi-spectral imagery in the visible and LWIR bands using commercially-available animation software, the extensively-validated DIRSIG model [18], and a custom optics and detector effects model.

To simulate a variety of images with realistic reflective and emissive phenomenology, I have developed methods and software to associate spectral emissivity and bi-directional reflectivity distribution function (BRDF) measurements in available

databases (e.g., archived MODIS and ASTER products [107] and the nonconventional exploitation factors database system [108]) with individual facets of 3D mesh object representations in “sets” created using modern computer graphics software. Since most such synthetic constructs are built for animated graphics in games or movies, the term “set” was borrowed by computer animators from the movie industry, and that convention was adopted. A detailed description of the approach requires considerable background in computer graphics and is outside the scope of this article, but the key concept is that individual set facets are attributed with material characteristics appropriate for the relevant phenomenology (reflective or emissive).

DIRSIG is a physics-based radiation propagation tool that generates high-fidelity radiance maps for wavelengths from the visible to far-infrared wavelengths (0.4 to 20.0 microns) [18]. DIRSIG has been applied to applications such as sensor trade studies [109, 110, 111] and training both humans [112] and neural networks [113]. DIRSIG calculates aperture-reaching radiance for user inputs describing the set, atmosphere, aperture, and tasking (including sensor motion). For reflective wavelengths, DIRSIG accesses reflectivity and BRDF information for each set facet. For emissive bands, DIRSIG accesses conductivity, emissivity, and specific heat information used by the DIRSIG thermal solver (THERM). If a more rigorous temperature solver than THERM is needed, external radiative transport codes, such as MuSES [114], can be used to directly attribute facets with temperature distributions. Atmosphere variables define the aerosol model and scattering for a local climate. In

this work I utilized the mid-latitude summer atmospheric model. A physical time for a simulation is set in the tasking block and defines the solar incidence angle and time interval over which local weather data is accessed. The weather file contains information used by the thermal solver and includes air temperature, wind speed, and relative humidity. DIRSIG uses MODTRAN [115] for solar radiance calculations and transmittance through the atmosphere. Unless temperature distributions are imported, weather history is used by THERM to calculate facet temperatures at the time of collection. A self-developed software toolkit, referred to as “BART,” was utilized to generate the over-sampled field angle by dividing it into a number of smaller field angles that were processed in parallel, greatly reducing computation time of the radiance map generation.

The process of converting DIRSIG radiance maps to simulated image digital counts is shown in Fig. 4.1. Note DIRSIG radiance maps are typically sampled more finely than the size of simulated detector “pixels” would require: Such oversampling ensures enough rays are cast at the object to avoid spatial aliasing of radiance variations, and also allows random shifts (jitter and motion blur) to be applied in increments smaller than the detector pixels. Discrete narrow-band radiance maps produced by DIRSIG are converted to photon maps using a specified f-number, integration time, and pixel size. System weighting due to optical transmission losses and quantum efficiency is applied to produce a spectrally weighted map of photo-detection events. Effects of blurring due to a poly-chromatic point spread function or field-dependent blur are

applied to account for diffraction and aberrations in the collecting optics and the atmosphere. Jitter and vibrations are included by summing an appropriate number of randomly-shifted versions of the image, which may be drawn from a relevant power spectral density. After downsampling to the pixel size of the detector to be simulated, the resulting image is corrupted with signal-dependent (Poisson) noise, to which samples from additive noise sources, such as read noise, are added. The coefficients of the response curve, which can be linear or cubic, are user-selectable, as is the electronics/count gain.

Figure 4.1: Process of converting DIRSIG radiance maps to imagery using the ODEM.

The scene developed for this work is an airfield set that includes a hangar, planes, a maintenance stand, and truck. VIS and LWIR images are shown in Fig. 4.2. Images are 512x512 pixels with a ground sample distance of 0.056 meters. The visible band is centered at 600 nm, with a full-width, half-maximum (FWHM) of 300 nm; the LWIR band is centered at 10 μm , with a FWHM of 4 μm . In DIRSIG, the VIS band is divided into 6 discrete spectral bins, and the LWIR is divided into 8 bins. The imaging system for both VIS and LWIR systems were modeled with a pixel size of

16 microns, focal length of 0.4 meters, and aperture of 0.05 meters. The integration time for VIS was 0.16 milliseconds and 0.0067 milliseconds for LWIR.

The DIRSIG simulation time was 0915 local and the atmosphere was mid-latitude summer. The engine of the truck was simulated in the “running” mode with a peak temperature of 120°F, which creates a phenomenological difference between the two images due to thermal heating from the warm engine; the warm hood of a running truck emits thermal radiation for the LWIR not observable in the visible band. Contrast inversions exist between the two images; e.g, the white painted lines have a high reflectivity (high radiance) in the visible band which results in a lower temperature (low radiance) in LWIR. Note also the solar glints from the visible-image aircraft fuselage and wings not present in the LWIR image; also, note the “smearing” of shadows in the direction away from the rising sun: These are “thermal scars,” or areas of the ramp that were recently in shadow, but are now exposed to thermal loading from the sun. Another example of the impact of solar thermal loading is shown by the tires of the truck: The temperature of the front tire is relatively cool due to sufficient time having passed since exposure to the sun. The rear tires, still exposed to the sun and on a surface also heated by the sun, are relatively warm.

A radiance map with typical DIRSIG sampling is shown in Fig. 4.3. An artifact of the temporal sampling in DIRSIG’s THERM solver is apparent in the LWIR shadow in the form of discrete thermal scarring due to the hangar. DIRSIG performs the thermal

solver at discrete time intervals, where the solar position between these intervals changes and effects the material temperature. This artifact is visually apparent in the over-sampled noise-free radiance map, but PSF blurring and noise typically render the artifacts imperceptible in the final image.

Figure 4.2: Synthetically generated imagery of an airfield set of (a) VIS and (b) LWIR bands. Images are 512×512 pixels with a ground sample distance of 0.056 meters. For reference, the image color scale in digital units is shown in the upper-left and a 32×32 sub-field is drawn in the lower-right of each image.

Figure 4.3: DIRSIG radiance map in (a) VIS and (b) LWIR displaying discrete thermal scarring as a result of the temporal sampling in the THERM solver. Note these radiance maps are free of an optical or detector effects to enable thermal scarring to be visualized.

4.7 Numerical Calculations of Error Metrics

As discussed in the Introduction, one of the advantages of using synthetic imaging targets is the ability to associate error with spatial features. To facilitate such association graphically, I generate “error maps” to show the spatial variation of the error over the field. The process of creating error maps using multi-spectral imagery is shown in Fig. 4.4. Exploiting parallel processing on a 48-CPU server, I have plotted the various shift-estimation metrics for all possible 32×32 -pixel sub-fields of the VIS images and between all identical sub-fields in the VIS and LWIR images, respectively. Metrics for the LWIR images are representative with only slight deviations of the metrics produced by VIS images; hence, their plots were not included. The color in each pixel of these figures is mapped to the error for the sub-field centered at the associated field angle. Note a 15-pixel (white) guard-band around each error map, indicating pixels that can’t be in the center of 32×32 -pixel sub-field.

Some consideration is required to create sub-field error maps for feature-based algorithms: A feature, including the pixels required for the descriptor, is typically quite small relative to the full image field-of-view, and the information used to create a feature match is accordingly localized; by contrast, area algorithms generally exploit the information over a larger angle than that subtended by a feature. As previously mentioned, area algorithms can be used to estimate shifts between multiple,

Figure 4.4: Process of generating error maps from shift-estimation error metrics for multi-spectral imagery. For each pixel, sub-fields surrounding the pixel are extracted from the image pair, and error metrics are calculated from an ensemble of noise-corrupted sub-field pairs. The resulting metric is inserted into the error map at the pixel location.

relatively small sub-fields, functioning effectively as matched features; equivalently, multiple feature matches can be (and typically are) used over a relatively large field of view, providing what amounts to a similarity measure to estimate a multi-parameter transformation. Even with these similarities, the difference in information localization means the error associated with a matched pair of features can't be arbitrarily associated with the sub-field in which the features are found. Further, the algorithm-calculated location of a feature will generally vary between noise samples, so while one may refer to a particular sub-field by referencing its coordinates, any reference to the coordinates of “a feature” would be to its ensemble average location.

Possibilities to resolve the inconsistency include calculating error metrics for a given sub-field from the weighted average of shift estimates produced by all feature matches

in that sub-field; however the resulting errors would depend not only on the feature-matching algorithm and descriptors used, but also the measure used to weight the average of estimates. My approach is to exploit the fact that image features are matched based on the quantitative similarity of their associated descriptors (the “score” of the match). Over an ensemble of noise sample functions, I calculate sample error statistics for the shift estimate from the highest-scoring feature match in the sub-field. Because each feature match is then associated with a sub-field, I can also compare errors with the FIE for the sub-field.

While the error maps are an excellent way to visualize the spatial distribution of error for a single algorithm, other methods are more practical for comparing multiple algorithms over the full image field or for a single sub-field. As examples, I have generated a.) CDFs for area- and feature-based algorithm VIS-LWIR RMSEs over the full field and b.) simple bar chart metric comparisons for selected sub-fields.

In Fig. 4.5 (a-c), the error metric is the UFIE. Note in Fig. 4.5 (a) and (b), there are broad, diagonal bands for which the UFIE is greater than 1 pixel. These bands run from approximately (0,280) to (256,0) in both images, and indicate low contrast in the image of the front of the hangar (on which small, external lighting fixtures are mounted). Straddling this band in both figures are parallel bands with lower UFIE; these bands indicate sub-fields containing diagonal contrast edges caused by the hangar roof-to-front and front-to-shadow transitions. Note also the expected

correspondence between VIS and LWIR image contrast and UFIE error in Fig. 4.5 (a) and (b); the maintenance stand and truck hood are excellent examples.

In Fig. 4.5 (c), “black” represents regions where the FIM has negative elements due to excessive contrast difference between the VIS and LWIR image features. Note nearly 70% of the FIMs calculated in Fig. 4.5(c) have negative elements. UFIE for VIS-LWIR registration is lowest for sub-fields centered on edges of the plane, truck, and maintenance stand with a value near 0.1 pixels. The edge running along the hangar shadow has a UFIE near 0.5 pixels. In all three figures, multiple regions exist where the UFIE is near 10 pixels, and corresponding to sub-fields with relatively low contrast. Note the 15-pixel (white) guard-band around each figure, indicating pixels that can’t be in the center of a 32×32 -pixel sub-field.

Figure 4.5: RT of shift-estimate UFIE for the VIS (a), LWIR (b), and VIS registered to LWIR (c) sub-fields centered at the indicated field angles. In (c), black pixels indicate that the associated FIM has negative elements.

4.7.1 Area-based Registration Algorithm Errors Maps

4.7.1.1 DFT

Shown in Fig. 4.6 are the DFT algorithm shift-estimation RSSB, RSV, and RMSE for (a-c) VIS-LWIR and VIS-VIS (d-f) registration for all possible sub-fields as described above. RSSB and RSV in Fig. 4.6 (a) and (b) is 0.1 pixels where shadows are cast by the truck, plane, and maintenance stand. These shadow regions offer edges with dark-to-light transitions having the same direction (that is, edges without contrast reversal) in both VIS and LWIR images. The diagonal band of the hangar shadow near the right-most shadow edge has an RSSB near 0.3 pixels where the shadow does not intersect the truck or either aircraft. RSSB for regions inside the supports of the two aircraft and truck, and along the painted lines in the lower right are greater than 10 pixels, likely due to contrast reversals between VIS and LWIR, including those apparent in the two images of the aircraft camouflage. The RSV for most of the upper-left region above the blue diagonal in Fig. 4.6 (b) is greater than 10 pixels; this region corresponds well with regions in Fig. 4.5 (c) where the FIM is undefined.

As an easily-interpreted example of the effect of contrast reversal on VIS-LWIR registration, it's worth comparing Fig. 4.6 (a-c) with Fig. 4.6 (d-f) for the previously-mentioned hangar roof-to-front, hangar front-to-shadow, and shadow edge bands. For

VIS-VIS registration, the RSSB, RSV, and RMSE are below 0.2 pixels for both the roof-to-front and front-to-shadow transition bands and the shadow edge. For VIS-LWIR, the RSSB is near 1 pixel and the RSV is greater than 10 pixels for both hangar bands; however, the error for sub-fields subtending the hangar shadow edge is significantly lower, with RSSB is less than 0.5 pixel and the RSV less than 0.1 pixel. The difference in error is due to the gradient direction: The gradients at both hangar transition bands are inverted between the VIS and LWIR bands, whereas for the shadow edge, the gradient is in the same direction in both bands. These examples are consistent with relatively high error – and undefined FIMS, as shown in Fig. 4.5 (c) – for the painted lines in the lower right corner, where contrast and gradient directions are also reversed in the VIS and LWIR images.

Figure 4.6: Samples VIS-LWIR (a-c) and VIS-VIS (d-f) DFT algorithm shift-estimation errors for 32×32 -pixel sub-fields centered at the indicated field angles.

4.7.1.2 ECC

Shown in Fig. 4.7 are the ECC algorithm shift-estimation RSSB, RSV, and RMSE for (a-c) VIS-LWIR and VIS-VIS (d-f) registration for all possible sub-fields as described above. Significantly, and in marked contrast to the DFT results, the RSV for nearly half of the possible sub-fields is below 1.0 pixel. Most regions where the RSV is high correspond to low-contrast sub-fields, such as the upper-left region that corresponds to the hangar roof. Because the similarity metric is designed to be robust with respect to gain and contrast differences, the ECC algorithm performed well on edges where the contrast was inverted between VIS and LWIR. For example, the painted lines in the lower right corner had an RSSB of under 0.1 pixels, resulting in an RMSE for those sub-fields nominally $100\times$ less than the corresponding DFT RMSE. The fact that the ECC estimator tends to be less affected by contrast reversal is also readily apparent in the hangar roof-to-front and front-to-shadow edges: It can be observed that distinct bands of relatively low bias are associated with these edges in Fig. 4.7 (a), unlike the unstructured, high-error, orange-and-red mottling in Fig. 4.6 (a), but the somewhat weaker front-to-shadow edge in the LWIR results in higher RSSB relative to the RSSB in the sub-fields containing the roof-to-front edge.

In light of the above, it's interesting to note that for edge regions without contrast reversal, the DFT algorithm yields lower RSV and RMSE. For example, while the

DFT algorithm yields 10-pixel RMSE for the hangar roof-to-front and front-to-shadow edge sub-fields, the band of shadow edge sub-fields in Fig. 4.6(b-c) indicate much lower RSV and RMSE for the DFT algorithm than the ECC. The DFT can also be seen to outperform ECC in the same-contrast regions associated with the maintenance stand stairs, the shadow cast by the empennage of the airplane facing left, the shadow case by the fuselage of the airplane facing right, and the shadow case by the truck.

Figure 4.7: Sample VIS-LWIR (a-c) and VIS-VIS (d-f) ECC algorithm shift-estimation errors for 32×32 -pixel sub-fields centered at the indicated field angles.

4.7.1.3 MI

Shown in Fig. 4.8 are the MI algorithm shift-estimation RSSB, RSV, and RMSE for (a-c) VIS-LWIR and VIS-VIS (d-f) registration for all possible sub-fields as described

above. Using the error maps (a-c), it's very apparent the MI algorithm multi-spectral shift-estimation error is significantly worse than for either DFT or ECC algorithms. For the majority of sub-fields, RMSE is on the order of 10 pixels. Notably, MI error is relatively high for even sub-fields where the other area-based algorithms performed well. Even VIS-VIS registration error in Fig. 4.8 (d-f) is significantly higher for many sub-fields when compared to DFT and ECC results.

Figure 4.8: Sample VIS-LWIR (a-c) and VIS-VIS (d-f) MI algorithm shift-estimation errors for 32×32 -pixel sub-fields centered at the indicated field angles.

4.7.1.4 Area-based Algorithm Comparison Using the CDF

CDF plots of the shift-estimation RMSE for area-based algorithms are shown in Fig. 4.9 and summarized in Table 4.1. For each sub-field, the RMSE is calculated over

200 noise samples, and the CDF is over the RMSE for all possible sub-fields. From the CDF plot, it's easy to see the full-field RMSE of the DFT and ECC algorithms are roughly equivalent between 0.2 and 1 pixel, with the number of DFT estimates with RMSE less than 0.5 pixels exceeding equivalent ECC estimates by at most a factor of just over 3. As in Figs. 4.6 and 4.7, the MI RMSE is again substantially worse than for either DFT or ECC, with a.) the number of MI estimates with RMSE less than 0.5 pixels over $10\times$ lower than for DFT and $4\times$ lower than for ECC, and b.) the number of MI estimates with RMSE less than 1 pixel nearly $5\times$ lower than DFT or ECC

Figure 4.9: CDF plot of the RMSE for shift-estimation of 32×32 -pixel sub-fields using area-based algorithms.

Table 4.1

Summary of metrics for shift-estimation of 32×32 -pixel sub-fields using area-based registration with VIS and LWIR imagery.

	Min. RMSE	< 0.1-pixel Percent	< 1.0-pixel Percent
DFT	0.0465	0.33	14.65
ECC	0.0165	0.86	13.59
MI	0.0389	0.48	2.58

4.7.2 Feature-based Shift-estimation Error Maps

Shown in Fig. 4.10 (a-d) are SIFT keypoints for VIS (left - ‘x’) and LWIR (right - ‘o’) with colored lines connecting matched keypoints using (a) Lowe, (b) EHD, (c) PCEHD, and (d) LGHD descriptors; a blue line represents a matching feature pair with an error displacement less than 5 pixels and a red line represents a matching feature pair with an error displacement greater than 5 pixels. For the original SIFT descriptor in Fig. 4.10 (a), nearly all lines are red and only three matches have blue lines indicating match displacement error less than 5 pixels. The modified descriptors in Fig. 4.10 (b-d) have an increased number of blue lines connecting similar locations in both images, displaying improvements to the displacement error between matched feature pairs.

The quantitative error calculations shown in the error maps below were performed as follows: An ensemble of 4000 noise-corrupted image pairs were created (VIS and LWIR). For each pair of noise samples, the shift-estimation algorithms were run using

Figure 4.10: SIFT keypoints VIS (left) and LWIR (right) images with lines connecting matched features using the (a) Lowe, (b) EHD, (c) PCEHD, and (d) LGHD descriptors. Blue lines connecting features indicate an error displacement of less than 5 pixels and red lines connect features indicate an error displacement greater than 5 pixels.

the full field of view. Colors are then mapped to each sub-field in which feature matches were found for at least 200 noise samples.

4.7.2.1 Lowe

Shown in Fig. 4.11 are the VIS-LWIR shift-estimation (a) RSSB, (b) RSV, and (c) RMSE for 32×32 -pixel sub-fields with at least 200 feature matches using the Lowe descriptor. Fewer than 3% of all sub-fields exceed this threshold. Most of the sub-fields with more than 200 matches are in shadows cast by the aircraft, truck, and maintenance stand. The lowest RMSE for the Lowe descriptor was 2 pixels.

Figure 4.11: Sample shift-estimates errors using the highest weighted feature match from the Lowe descriptor for a 32×32 -pixel sub-field centered at the indicated field angle. “Black” pixels indicate sub-fields where a minimum of 200 feature point were not matched.

4.7.2.2 EHD

Shown in Fig. 4.12 are the VIS-LWIR shift-estimation (a) RSSB, (b) RSV, and (c) RMSE for 32×32 -pixel sub-fields with at least 200 feature matches using the EHD descriptor. Relative to the Lowe descriptor, there are over $4 \times$ more sub-fields that exceeded the 200 match requirement. These estimates are produced on the fuselage of the aircrafts, body of the truck, the maintenance stand, and the painted lines. The RSSB, RSV, and RMSE in Fig. 4.12 (a-c) are below 1.0 pixel for regions that correspond to the painted lines in the lower right; this region also produced low shift-estimate error for the ECC algorithm in Fig. 4.7 (a-c). Regions surrounding the fuselage of the aircraft had an RMSE between 1 and 9 pixels. The RSV was generally low for all sub-fields and the RSSB was the primary source of error.

Figure 4.12: Sample shift-estimates errors using the highest weighted feature match from the EHD descriptor for a 32×32 -pixel sub-field centered at the indicated field angle. “Black” pixels indicate sub-fields where a minimum of 200 feature point were not matched.

4.7.2.3 PCEHD

Shown in Fig. 4.13 are the VIS-LWIR shift-estimation (a) RSSB, (b) RSV, and (c) RMSE for 32×32 -pixel sub-fields with at least 200 feature matches using the PCEHD descriptor. The number of sub-fields was nearly the same to that produced by the EHD, with a slight change in their relative locations. The painted line region produced shift-estimation errors below 1 pixel; this was also true for the EHD descriptor.

Figure 4.13: Sample shift-estimates errors using the highest weighted feature match from the PCEHD descriptor for a 32×32 -pixel sub-field centered at the indicated field angle. “Black” pixels indicate sub-fields where a minimum of 200 feature point were not matched.

4.7.2.4 LGHD

Shown in Fig. 4.14 are the VIS-LWIR shift-estimation (a) RSSB, (b) RSV, and (c) RMSE for 32×32 -pixel sub-fields with at least 200 feature matches using the LGHD descriptor. Relative to the other three descriptors, LGHD produces the most sub-field shift-estimates. A large number of the additional estimate were around the fuselage of the bottom plane, the rear of the top plane, and the body of the truck. As was true for the EHD and PCEHD descriptors, the painted line region produced shift-estimate errors below 1 pixel.

Figure 4.14: Sample shift-estimates errors using the highest weighted feature match from the LGHD descriptor for a 32×32 -pixel sub-field centered at the indicated field angle. “Black” pixels indicate sub-fields where a minimum of 200 feature point were not matched.

4.7.2.5 Feature-based Algorithm Comparison Using the CDF

CDF plots of the VIS-LWIR shift-estimation RMSE for feature-based algorithms are shown in Fig. 4.15 and summarized in Table 4.2. The feature-based matching

Table 4.2

Summarized RMSE metrics for 32×32 -pixel sub-fields using four different feature descriptors. Note the Lowe descriptor did not have any sub-fields with a sub-pixel RMSE.

	Sub-field Matches	Min. RMSE	< 0.5-pixel Percent	< 1.0-pixel Percent
Lowe	5911	1.990	-	-
EHD	26311	0.129	7.25	11.7
PCEHD	25352	0.126	7.55	14.0
LGHD	45306	0.128	3.87	7.0

algorithms were performed for 4000 samples of the relevant noise processes. The CDF is over the RMSE for the ensemble of highest-scoring matches in each sub-field with at least 200 matches. The Lowe descriptor resulted in the lowest number of sub-field matches and did not have any sub-fields with an RMSE less than 1 pixel. From the CDF plot, the RMSE for EHD and PCEHD are similar up though 0.5 pixels and $2\times$ higher than the LGHD. The PCEHD has 14% of sub-fields with an RMSE less than 1 pixel which outperforms the EHD with 11.7% and LGHD at 7.0%. The LGHD had the highest number of sub-field matches but had the lowest percent between the modified descriptors of sub-fields with an RMSE below 0.5 and 1.0 pixel.

The CDF shows at a glance that the modified SIFT descriptors significantly improved feature matching compared to the original Lowe descriptor; there was a both an increase in the number of matched features and the match error was decreased.

Figure 4.15: Cumulative distribution plot of RMSE for 32×32 -pixel sub-fields using different descriptors for feature matching on the synthetic VIS and LWIR imagery.

4.7.3 Summary of Algorithm Performance for Selected Sub-fields

In Fig. 4.16, I show four sub-fields used to demonstrate comparison of multiple algorithms. The results are shown in the bar charts in Fig. 4.17. Sub-fields 1 and 4 are found on shadow regions and sub-fields 2 and 3 are located on the fuselage of the lower aircraft, but sub-fields can be selected arbitrarily according to the need or interest. The individual bar charts in Fig. 4.16 show (a) the BFIE from Eq. 4.20, (b) the root-sum variance increase due to the bias gradient (Eq. 4.19), (c) the RSSB, (d) the RSV, and (e) the RMSE. The bar charts make quite apparent the relatively large errors associated with the MI algorithm. The charts also illustrate a point made in

Sec. 4.7.1.2: While the ECC algorithm has lower error in regions with the contrast reversal typical in VIS-LWIR shift estimation, the DFT algorithm can exhibit lower – for some sub-fields, dramatically lower – error than the ECC algorithm.

Figure 4.16: Sub-field locations for comparison of area- and feature-based algorithm metrics.

4.8 Conclusion

I have demonstrated the utility of synthetic phenomenology and simulated imagery for quantitative analysis of algorithm error in estimation of isometric shifts between images in multi-spectral image data sets. Shift estimation can be extended to estimate parameters in more general registration transformation models. In addition to the error metrics used, I discussed the techniques and methods for creating synthetic visible-wavelength and long-wave infrared imagery, including radiance calculations

Figure 4.17: Performance comparison of area and feature-based techniques showing (a) BFIE, (b) the variance increase due to the bias gradient, (c) RSSB, (d) RSV, and (e) RMSE for four 32×32-pixel sub-fields.

with DIRSIG and the simulation of optics and detector phenomena. Because I can create simulated images without noise, I was able to compare empirical algorithm error to error derived from relevant Fisher information matrices, including the uncertainty added by algorithm bias gradient. I discussed the difference in error computation methods for area- and feature-based algorithms. I also demonstrated how synthetic data and parallel computing can be used to generate a.) “error maps” to associate error with spatial image features for a single algorithm, and b.) CDFs and

bar chart plots to evaluate multiple algorithms in some number of selected sub-fields.

Incidentally to the main thrust of the chapter, I make some observations regarding the evaluated algorithms: For area-based algorithms and the synthetic image pair, the MI approach had significantly higher error and computational cost than either the DFT or ECC algorithms. Also, over field angles where the difference between reflective and emissive phenomenology resulted in contrast reversals at edges, the ECC algorithm performed noticeably better than DFT; however, where edges did not have contrast reversal between VIS and LWIR, such as shadow edges, the DFT algorithm tended to outperform ECC.

For feature-based algorithms and the synthetic image pair, the Lowe SIFT descriptor produced the worst results, with no sub-fields having an RMSE less than 1 pixel. The PCEHD descriptor had the highest percentage of errors between 0.5 and 1 pixel, and the LGHD descriptor had feature matches in the largest number of sub-fields.

Finally, noise-free data can be generated as well as images corrupted by additive and signal-dependent noise, I was able to calculate variances derived from relevant Fisher Information matrices for both unbiased estimation and non-zero algorithm bias gradient. Over 4 sub-fields selected from the image pair, the feature-based algorithms appeared to depend less on the true shift (lower bias gradient), resulting in relatively smaller variance added to the unbiased Fisher Information error.

In future work, I anticipate characterizing algorithm performance of multi-spectral image registration more generally, including additional spectral bands, various scenes, and band-to-band differences in pixel sampling and optical effects. I also expect to characterize algorithms that directly estimate higher-order transformations, such as rotation and scaling.

Chapter 5

Conclusion

In this dissertation, approaches to fusing data from multiple sensors, including multi-spectral and radio-frequency sensors, were presented. Chapter 2 presented an algorithm to fuse multi-spectral imagery to improve target detection rate, lower false alarm rate, and increase the performance of tracking targets through occlusions and target mergers. A Gaussian mixture model was used to detect foreground pixels in all spectral bands, which were then fused into a single foreground map and a segmentation process extracted target candidates. Targets were tracked through occlusions and target mergers using SIFT features. This algorithm was evaluated using synthetic imagery in the visible, NIR, MWIR, and LWIR spectral bands. Chapter 3 presented methods to localize a moving target by fusing a multi-spectral tracker with Doppler differential (DD) measurements from a constellation of RF sensors. The data was

fused by comparing the maximum DD produced by the the RF sensors with the theoretical DD computed from the multi-spectral tracker using the absolute difference at each time instance and the root-mean-square difference. Chapter 4 presented a framework to evaluate and compare performance of shift-estimation of visible and long-wave infrared imagery using area- and feature-based algorithms. Synthetically generated imagery was created by using 3D computer animation software to create sets of models attributed with reflective and emissive properties. The DIRSIG tool-set was utilized to generate radiation maps of the sets, and then processed by an optics and detector effects model to create imagery that is representative of that produced by a real sensor. Performance of shift-estimation using area- and feature-based algorithms were evaluated using the synthetic imagery and quantitative metrics were presented for algorithm comparison.

5.1 Summary of Key Results

The multi-spectral fusion algorithm was evaluated using synthetically generated data in visible, near infrared, mid-wave infrared, and long-wave infrared spectral bands. The fusion of the multi-spectral foreground maps improved the accuracy of target centroid estimation while increasing the detection rates and lowering the false alarm rates for most spectral bands. LWIR produced the highest detection rate, but produced the highest false alarm rate. The spectral combination VIS-MWIR had only a slightly

lower detection rate relative to LWIR, but significantly reduced the false alarm rate. Using features as the tracking mechanism, NIR-MWIR had the largest percentage of tracks attributed to features. For centroid location accuracy, NIR-LWIR produced the lowest displacement of less than 0.1 meters. The tracking algorithm successfully tracked all 12 moving targets, with only one instance of a lost track that was later recovered.

Track output from the multi-spectral fusion algorithm was fused with DD of a moving cellular emanation measured from an RF constellation. For low measurement uncertainty and favorable motion patterns, the absolute DD correctly associates the cellular emanation with the corresponding video target. Compared to the absolute DD, the association rate using the RMSD increases through higher levels of uncertainty. For nearly all objects the algorithm has high confidence in associating the emanation with the correct multi-spectral target from the most probable background target.

Finally, synthetically generated data was utilized to evaluate and compare area- and feature-based image registration algorithms; metrics include bias, variance, and the bias gradient variance. For the area-based algorithms and the synthetic image pair, the MI algorithm performed much worse when compared to the DFT and ECC algorithms. The ECC performed best in regions where the reflective and emissive phenomenology results in contrast results of edges, such as the paint lined. For edges

where there was no contrast reversal, such as edges, the DFT algorithm performed best. For feature-based algorithms and synthetic image pair, the Lowe SIFT descriptor produced no sub-fields having an RMSE less than 1 pixel. The PCEHD descriptor had the highest percentage of errors between 0.5 and 1 pixel, and the LGHD descriptor had feature matches in the largest number of sub-fields. Over 4 sub-fields selected from the image pair, the feature-based algorithms appeared to depend less on the true shift (lower bias gradient), resulting in relatively smaller variance added to the unbiased Fisher information error.

5.2 Suggested Future Work

This work considered image fusion at the feature-level by fusing foreground maps from different spectral bands. Pixel-level fusion is generally more complex as it combines data from different phenomenological sources. This dissertation examined errors for registering these multi-spectral images and I would like to pursue pixel-level fusion of multi-spectral imagery to create a single fused image.

Image fusion requires precise pixel-to-pixel alignment that is difficult to accomplish in multi-spectral imagery where phenomenological differences result in estimator errors. Phenomenological differences effect the performance of algorithms differently, and can change significantly over the course of a day. I would like to perform a more thorough

study of algorithm performance using different spectral bands over a wide variety of scenarios including different sets, ground sample distance, and collection times.

Synthetic imagery was utilized in this dissertation for algorithm development and evaluation. While synthetic data can be representative of a real sensor, there are effects that are difficult to model and implement in synthetic imagery. Therefore, evaluation of the presented algorithms in this work using real sensors data is strongly desired.

References

- [1] Demars, C. D.; Roggemann, M. C.; Havens, T. C. *Optical Engineering* **2015**, *54*(12), 123106–123106.
- [2] Demars, C. D.; Tyler, D. W.; Allen, D.; Dank, J.; Roggemann, M. C. *Applied Optics* **2018**, *57*(32).
- [3] Hall, D. L.; Llinas, J. *Proceedings of the IEEE* **1997**, *85*(1), 6–23.
- [4] Atrey, P. K.; Hossain, M. A.; El Saddik, A.; Kankanhalli, M. S. *Multimedia systems* **2010**, *16*(6), 345–379.
- [5] Hall, D. L.; Garga, A. K. In *Proceedings of the Second International Conference on Information Fusion (Fusion99)*, Vol. 1, pages 429–436, 1999.
- [6] Esteban, J.; Starr, A.; Willetts, R.; Hannah, P.; Bryanston-Cross, P. *Neural Computing & Applications* **2005**, *14*(4), 273–281.
- [7] Khaleghi, B.; Khamis, A.; Karray, F. O.; Razavi, S. N. *Information Fusion* **2013**, *14*(1), 28–44.

- [8] Fasano, G.; Accardo, D.; Tirri, A. E.; Moccia, A.; De Lellis, E. *Aerospace Science and Technology* **2015**, *46*, 436–450.
- [9] Hollinger, J.; Kutscher, B.; Close, R. In *Unmanned Systems Technology XVII*, Vol. 9468, page 946806. International Society for Optics and Photonics, 2015.
- [10] James, A. P.; Dasarathy, B. V. *Information Fusion* **2014**, *19*, 4–19.
- [11] Moghaddam, M.; Dungan, J. L.; Acker, S. *IEEE Transactions on Geoscience and Remote Sensing* **2002**, *40*(10), 2176–2187.
- [12] Prakash, R.; Singh, D.; Pathak, N. P. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2012**, *5*(1), 196–206.
- [13] Ghassemian, H. *Information Fusion* **2016**, *32*, 75–89.
- [14] Torabi, A.; Massé, G.; Bilodeau, G.-A. *Computer Vision and Image Understanding* **2012**, *116*(2), 210–221.
- [15] Lowe, D. G. *International journal of computer vision* **2004**, *60*(2), 91–110.
- [16] Guizar-Sicairos, M.; Thurman, S. T.; Fienup, J. R. *Optics letters* **2008**, *33*(2), 156–158.
- [17] Ros, G.; Sellart, L.; Materzynska, J.; Vazquez, D.; Lopez, A. M. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3234–3243, 2016.

- [18] Sanders, J. S.; Brown, S. D. In *Proc. SPIE*, Vol. 4029, page 278, 2000.
- [19] Heintz, F.; Rudol, P.; Doherty, P. In *Information Fusion, 2007 10th International Conference on*, pages 1–8. IEEE, 2007.
- [20] Meng, L.; Kerekes, J. P. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2012**, 5(1), 146–152.
- [21] Liang, P.; Shen, D.; Blasch, E.; Pham, K.; Wang, Z.; Chen, G.; Ling, H. In *SPIE Defense, Security, and Sensing*, pages 875105–875105. International Society for Optics and Photonics, 2013.
- [22] Kumar, K. S.; Kavitha, G.; Subramanian, R.; Ramesh, G. *Edited by Clara M. Ionescu* **2011**, page 307.
- [23] Choi, H.; Kim, Y. *Control Engineering Practice* **2014**, 22, 10–19.
- [24] Beier, K.; Gemperlein, H. *Aerospace Science and Technology* **2004**, 8(1), 63–71.
- [25] Li, H.; Manjunath, B.; Mitra, S. K. *Graphical models and image processing* **1995**, 57(3), 235–245.
- [26] Ardeshir Goshtasby, A.; Nikolov, S. *Information Fusion* **2007**, 8(2), 114–118.
- [27] Nunez, J.; Otazu, X.; Fors, O.; Prades, A.; Pala, V.; Arbiol, R. *Geoscience and Remote Sensing, IEEE Transactions on* **1999**, 37(3), 1204–1211.
- [28] Li, S.; Yang, B.; Hu, J. *Information Fusion* **2011**, 12(2), 74–84.

- [29] Miao, Q.-g.; Shi, C.; Xu, P.-f.; Yang, M.; Shi, Y.-b. *Optics Communications* **2011**, *284*(6), 1540–1547.
- [30] Davis, J. W.; Sharma, V. *Computer Vision and Image Understanding* **2007**, *106*(2), 162–182.
- [31] Chen, C.-Y.; Wolf, W. In *Proceedings of the 4th ACM international Workshop on Video Surveillance and Sensor Networks*, pages 27–34. ACM, 2006.
- [32] Leykin, A.; Hammoud, R. *Machine Vision and Applications* **2010**, *21*(4), 587–595.
- [33] Torresan, H.; Turgeon, B.; Ibarra-Castanedo, C.; Hebert, P.; Maldague, X. P. In *Defense and Security*, pages 506–515. International Society for Optics and Photonics, 2004.
- [34] Feng, Z.-J.; Zhang, X.-L.; Yuan, L.-Y.; Wang, J.-N. *Mathematical Problems in Engineering* **2013**, *2013*.
- [35] Xiaoqi, L.; Baohua, Z.; Ying, Z.; He, L.; Haiquan, P. *Infrared Physics & Technology* **2014**.
- [36] Scribner, D.; Warren, P.; Schuler, J. In *Computer Vision Beyond the Visible Spectrum: Methods and Applications, 1999.(CVBVS'99) Proceedings. IEEE Workshop on*, pages 33–40. IEEE, 1999.

- [37] Stauffer, C.; Grimson, W. E. L. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, Vol. 2. IEEE, 1999.
- [38] KaewTraKulPong, P.; Bowden, R. In *Video-Based Surveillance Systems*; Remagnino, P., Jones, G., Paragios, N., Regazzoni, C., Eds., Vol. 1; Springer: US, 1 ed., 2002; pages 135–144.
- [39] Zivkovic, Z. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, Vol. 2, pages 28–31. IEEE, 2004.
- [40] Chang, L.; Hernández-Palancar, J.; Sucar, L. E.; Arias-Estrada, M. *Machine vision and applications* **2013**, *24*(2), 371–392.
- [41] Fassold, H.; Rosner, J. In *IS&T/SPIE Electronic Imaging*, pages 940007–940007. International Society for Optics and Photonics, 2015.
- [42] Siftgpu: A gpu implementation of scale invariant feature transform (sift). Wu, C. **2007**.
- [43] Sinha, S. N.; Frahm, J.-M.; Pollefeys, M.; Genc, Y. *Machine Vision and Applications* **2011**, *22*(1), 207–217.
- [44] Berk, A.; Bernstein, L. S.; Robertson, D. C. Modtran: A moderate resolution model for lowtran Technical report, DTIC Document, **1987**.

- [45] Johnson, K.; Curran, A.; Less, D.; Levanen, D.; Marttila, E.; Gonda, T.; Jones, J.; TACOM, U. A. In *Proceedings of Ninth Annual Ground Target Modeling and Validation Conference, Houghton, Michigan*, 1998.
- [46] Behrisch, M.; Bieker, L.; Erdmann, J.; Krajzewicz, D. In *SIMUL 2011, The Third International Conference on Advances in System Simulation*, pages 55–60, 2011.
- [47] Adams, A.; Tummala, M.; McEachen, J.; Scrofani, J. In *Signal Processing and Communication Systems (ICSPCS), 2013 7th International Conference on*, pages 1–6. IEEE, 2013.
- [48] Tan, D. K.; Sun, H.; Lu, Y.; Lesturgie, M.; Chan, H. L. *IEE Proceedings-Radar, Sonar and Navigation* **2005**, *152*(3), 116–123.
- [49] Griffiths, H.; Baker, C. *IEE Proceedings-Radar, Sonar and Navigation* **2005**, *152*(3), 153–159.
- [50] Lee, B.; Chan, Y.; Chan, F.; Du, H.-J.; Dilkes, F. A. In *MILCOM 2007-IEEE Military Communications Conference*, pages 1–6. IEEE, 2007.
- [51] Vasuhi, S.; Vaidehi, V. *Information Fusion* **2016**, *27*, 41–53.
- [52] Zhang, Y. D.; Himed, B. In *2013 IEEE Radar Conference (RadarCon13)*, pages 1–5. IEEE, 2013.
- [53] Kim, W. S. *Fordham Intell. Prop. Media & Ent. LJ* **2015**, *26*, 995.

- [54] Pell, S. K.; Soghoian, C. *Harv. JL & Tech.* **2014**, 28, 1.
- [55] Dai, C.; Zheng, Y.; Li, X. *Computer Vision and Image Understanding* **2007**, 106(2), 288–299.
- [56] Li, J.; Gong, W.; Li, W.; Liu, X. *Infrared Physics & Technology* **2010**, 53(4), 267–273.
- [57] Song, X.; Shao, X.; Zhang, Q.; Shibasaki, R.; Zhao, H.; Zha, H. *Information Fusion* **2013**, 14(3), 301–310.
- [58] Rastegar, S.; Babaeian, A.; Bandarabadi, M.; Toopchi, Y. In *2009 41st South-eastern Symposium on System Theory*, pages 64–67. IEEE, 2009.
- [59] Maini, R.; Aggarwal, H. *International journal of image processing (IJIP)* **2009**, 3(1), 1–11.
- [60] Tuytelaars, T.; Mikolajczyk, K. *Foundations and trends® in computer graphics and vision* **2008**, 3(3), 177–280.
- [61] Noulas, A.; Englebienne, G.; Krose, B. J. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **2012**, 34(1), 79–93.
- [62] Kli, V.; Zhong, X.; Barnard, M.; Wang, W.; Kittler, J. In *Information Fusion (FUSION), 2014 17th International Conference on*, pages 1–7, 2014.
- [63] D’Arca, E.; Robertson, N. M.; Hopgood, J. In *Data Fusion Target Tracking Conference (DF TT 2012): Algorithms Applications, 9th IET*, pages 1–6, 2012.

- [64] Chin, T.; Xiong, K.; Blasch, E. In *Information Fusion (Fusion), 2015 18th International Conference on*, pages 1946–1953, 2015.
- [65] Weinstein, E. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **1982**, *30*(1), 112–117.
- [66] Otnes, R. K. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **1989**, *37*(2), 306–308.
- [67] Tenenbaum, S.; Stouch, D.; McGraw, K.; Fichtl, T. In *Unmanned Systems Technology X*, Vol. 6962, page 696215. International Society for Optics and Photonics, 2008.
- [68] Danoy, G.; Brust, M. R.; Bouvry, P. In *Proceedings of the 5th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications*, pages 1–8. ACM, 2015.
- [69] Stouch, D. W.; Zeidman, E.; Callahan, W.; McGraw, K. In *Unmanned Systems Technology XIII*, Vol. 8045, page 80450F. International Society for Optics and Photonics, 2011.
- [70] Demars, C. D.; Roggemann, M. C.; Havens, T. C. *Optical Engineering* **2015**, *54*(12), 123106.
- [71] Tse, D.; Viswanath, P. *Fundamentals of wireless communication*; Cambridge university press, 2005.

- [72] Kay, S. *IEEE Signal Processing Letters* **2013**, 7(20), 721–724.
- [73] VLFeat: An open and portable library of computer vision algorithms. Vedaldi, A.; Fulkerson, B. <http://www.vlfeat.org/>, **2008**.
- [74] Yang, J.; Wright, J.; Huang, T. S.; Ma, Y. *IEEE transactions on image processing* **2010**, 19(11), 2861–2873.
- [75] Masci, F. J.; Makovoz, D.; Moshir, M. *Publications of the Astronomical Society of the Pacific* **2004**, 116(823), 842.
- [76] Hill, D. L.; Batchelor, P. G.; Holden, M.; Hawkes, D. J. *Physics in medicine and biology* **2001**, 46(3), R1.
- [77] Harris, C.; Stephens, M. In *In Proc. of Fourth Alvey Vision Conference*, pages 147–151, 1988.
- [78] Hisham, M.; Yaakob, S. N.; Raof, R. A.; Nazren, A. A.; Embedded, N. W. In *Research and Development (SCORED), 2015 IEEE Student Conference on*, pages 100–104. IEEE, 2015.
- [79] Bunting, P.; Labrosse, F.; Lucas, R. *Image and Vision Computing* **2010**, 28(8), 1203–1219.
- [80] Kern, J. P.; Pattichis, M. S. *IEEE Transactions on Geoscience and Remote Sensing* **2007**, 45(5), 1494–1505.

- [81] de Vos, B. D.; Berendsen, F. F.; Viergever, M. A.; Staring, M.; Išgum, I. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer, 2017; pages 204–212.
- [82] Yang, X.; Kwitt, R.; Styner, M.; Niethammer, M. *NeuroImage* **2017**, *158*, 378–396.
- [83] Blum, R. S. In *Information Fusion, 2005 8th International Conference on*, Vol. 1, pages 8–pp. IEEE, 2005.
- [84] Jin, X.; Jiang, Q.; Yao, S.; Zhou, D.; Nie, R.; Hai, J.; He, K. *Infrared Physics & Technology* **2017**, *85*, 478–501.
- [85] Ma, J.; Ma, Y.; Li, C. *Information Fusion* **2019**, *45*, 153–178.
- [86] Schnelle, S. R.; Chan, A. L. Fusing infrared and visible imageries for improved tracking of moving targets Technical report, ARMY RESEARCH LAB ADELPHI MD SENSORS AND ELECTRON DEVICES DIRECTORATE, **2011**.
- [87] Aguilera, C.; Barrera, F.; Lumbreras, F.; Sappa, A. D.; Toledo, R. *Sensors* **2012**, *12*(9), 12661–12672.
- [88] Cronje, J.; De Villiers, J. **2012**.
- [89] Mouats, T.; Aouf, N.; Sappa, A. D.; Aguilera, C.; Toledo, R. *IEEE Transactions on Intelligent Transportation Systems* **2015**, *16*(3), 1210–1224.

- [90] Aguilera, C. A.; Sappa, A. D.; Toledo, R. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 178–181. IEEE, 2015.
- [91] Kay, S. M. *Fundamentals of statistical signal processing*; Prentice Hall PTR, 1993.
- [92] Zitova, B.; Flusser, J. *Image and vision computing* **2003**, 21(11), 977–1000.
- [93] Crum, W. R.; Hartkens, T.; Hill, D. *The British journal of radiology* **2004**, 77(suppl_2), S140–S153.
- [94] Robinson, D.; Milanfar, P. *IEEE Transactions on Image Processing* **2004**, 13(9), 1185–1199.
- [95] Tyler, D. W.; Dank, J. A. *JOSA A* **2015**, 32(8), 1425–1436.
- [96] Tyler, D. W. May **2018**, 43(10), 2292–2295.
- [97] Evangelidis, G. D.; Psarakis, E. Z. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2008**, 30(10), 1858–1865.
- [98] Maes, F.; Collignon, A.; Vandermeulen, D.; Marchal, G.; Suetens, P. *IEEE transactions on Medical Imaging* **1997**, 16(2), 187–198.
- [99] Mattes, D.; Haynor, D. R.; Vesselle, H.; Lewellyn, T. K.; Eubank, W. In *Medical imaging 2001: image processing*, Vol. 4322, pages 1609–1621. International Society for Optics and Photonics, 2001.

- [100] Pluim, J. P.; Maintz, J. A.; Viergever, M. A. *IEEE transactions on medical imaging* **2003**, *22*(8), 986–1004.
- [101] Jiang, Y.-h. In *Radar Conference 2013, IET International*, pages 1–4. IET, 2013.
- [102] Ma, J.; Zhao, J.; Ma, Y.; Tian, J. *Pattern Recognition* **2015**, *48*(3), 772–784.
- [103] Ma, J.; Jiang, J.; Liu, C.; Li, Y. *Information Sciences* **2017**, *417*, 128–142.
- [104] Toews, M.; Zöllei, L.; Wells, W. M. In *International Conference on Information Processing in Medical Imaging*, pages 25–36. Springer, 2013.
- [105] Chen, M.; Habib, A.; He, H.; Zhu, Q.; Zhang, W. *Remote Sensing* **2017**, *9*(9), 882.
- [106] Fessler, J. A.; Hero, A. O. In *Circuits and Systems, 1993., Proceedings of the 36th Midwest Symposium on*, pages 253–256. IEEE, 1993.
- [107] Modis/aster database. https://lpdaac.usgs.gov/dataset_discovery.
- [108] Nonconventional exploitation factors data system (nefds). :<http://math.nist.gov/~FHunt/appearance/nefds.html>.
- [109] Gerace, A. D.; Goodenough, A. A.; Montanaro, M.; Yang, J.; McCorkel, J. T. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXI*, Vol. 9472, page 947214. International Society for Optics and Photonics, 2015.

- [110] Ambeau, B. L.; Gerace, A. D.; Montanaro, M.; McCorkel, J. In *Earth Observing Systems XXI*, Vol. 9972, page 99720M. International Society for Optics and Photonics, 2016.
- [111] Kolb, K. E.; Hee-sue, S. C.; Kaur, B.; Olson, J. T.; Hill, C. F.; Hutchinson, J. A. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXVII*, Vol. 9820, page 982019. International Society for Optics and Photonics, 2016.
- [112] May, C. M.; Maurer, T. O.; Sanders, J. S. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXVIII*, Vol. 10178, page 101780W. International Society for Optics and Photonics, 2017.
- [113] Howell, C. L.; Manser, K.; Olson, J. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIX*, Vol. 10625, page 106250Q. International Society for Optics and Photonics, 2018.
- [114] Packard, C. D.; Curran, A. R.; Saur, N. E.; Rynes, P. L. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXVI*, Vol. 9452, page 94520F. International Society for Optics and Photonics, 2015.
- [115] Band model method for modeling atmospheric propagation at arbitrarily fine spectral resolution. October 7 **2008**.
- [116] Robinson, D.; Milanfar, P. *Signal Processing: Image Communication* **2005**, 20(6), 554–568.

- [117] Pinar, A. J.; Havens, T. C.; Webb, A. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII*, Vol. 10182, page 101821A. International Society for Optics and Photonics, 2017.
- [118] Krotosky, S. J.; Trivedi, M. M. *Circuits and Systems for Video Technology, IEEE Transactions on* **2008**, 18(8), 1096–1105.

Appendix A

Letters of Permission

A.1 SPIE letter of permission

The content in Chap. 2 have been published by SPIE Optical Engineering. Optical Engineering has the following policy regarding republication of this material: As stated in the SPIE Transfer of Copyright agreement, authors, or their employers in the case of works made for hire, retain the following rights: All proprietary rights other than copyright, including patent rights.

† The right to make and distribute copies of the Paper for internal purposes.

† The right to use the material for lecture or classroom purposes.

† The right to prepare derivative publications based on the Paper, including books or book chapters, journal papers, and magazine articles, provided that publication of a derivative work occurs subsequent to the official date of publication by SPIE. Thus, authors may reproduce figures and text in new publications. The SPIE source publication should be referenced.

A.2 OSA letter of permission

The content in Chap. 4 has been published in the OSA - Applied Optics. OSA has the following policy regarding republication of this material: As stated in the OSA Copyright Transfer Agreement, OSA grants to the Author(s) (or their employers, in the case of works made for hire) the following rights:

† The right, after publication by OSA, to use all or part of the work without revision or modification, including the OSA formatted version, in personal compilations or other publications consisting solely of the Author(s') own works, including the Author(s') personal web home page, and to make copies of all or part of the Work for the Author(s') use for lecture or classroom purposes.

† All proprietary rights other than copyright, such as patent rights.

† If the Work has been prepared by an employee within the scope of his or her

employment or as a work made for hire, the right to make copies of the Work for the employer's internal use. Thus, authors may reproduce figures and text in new publications. The OSA source publication should be referenced.