# The Development of a Performance Assessment Methodology for Activity Based Intelligence: A Study of Spatial, Temporal, and Multimodal Considerations

by

Christian M. Lewis

B.S. Embry-Riddle Aeronautical University, 2009

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology

15 August 2014

Signature of the Author _____

Accepted by _____
Dr. John Kerekes, M.S. Degree Coordinator Date





ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

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

CERTIFICATE OF APPROVAL

---

M.S. DEGREE THESIS

---

The M.S. Degree Thesis of Christian M. Lewis
has been examined and approved by the
thesis committee as satisfactory for the
thesis required for the
M.S. degree in Imaging Science

Dr. David Messinger, Thesis Advisor

Dr. Carl Salvaggio

Dr. Derek Walvoord

Guest Member

Date

# Declaration of Authorship

I, Christian M. Lewis, declare that this thesis titled, 'The Development of a Performance Assessment Methodology for Activity Based Intelligence: A Study of Spatial, Temporal, and Multimodal Considerations' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

*"The supreme art of war is to subdue the enemy without fighting."*

Sun Tzu

**Test of a man**

*"The test of a man is the fight that he makes, The grit that he daily shows, The way he stands upon his feet, And takes life's numerous bumps and blows. A coward can smile when there's naught to fear. And noting his progress bars, But it takes a man to stand and cheer, while the other fellow stars. It isn't the victory after all. But the fight that a Brother makes. A man when driven against the wall, still stands erect, and takes the blows of fate with his head held high, bleeding, bruised, and pale, Is the man who will win and fate defied, For he isn't afraid to fail."*

An Unknown Author

*"We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness."*

Declaration of Independnce

**Our deepest fear**

*"Our deepest fear is not that we are inadequate. Our deepest fear is that we are powerful beyond measure. It is our light, not our darkness that most frightens us. We ask ourselves, Who am I to be brilliant, gorgeous, talented, fabulous? Actually, who are you not to be? You are a child of God. Your playing small does not serve the world. There is nothing enlightened about shrinking so that other people won't feel insecure around you. We are all meant to shine, as children do. We were born to make manifest the glory of God that is within us. It's not just in some of us; it's in everyone. And as we let our own light shine, we unconsciously give other people permission to do the same. As we are liberated from our own fear, our presence automatically liberates others."*

Marianne Williamson

# *Acknowledgements*

I would like to thank all the professors, staff, and my fellow students at RITs Chester F. Carlson Center for Imaging Science, for the amazing and insightful experience I have had throughout this program. I am indebted to those that took the time to provide me valuable tips and guidance through this research process and the writing of this thesis. Their constant encouragement and support gave me the drive to continue exploring avenues of research throughout my experience.

I would also like to thank the members of my committee, Dave Messinger, Carl Salvaggio, and Derek Walvoord for providing me with their insight and knowledge throughout this work. An additional thanks goes to Mike Gartley and Jason Faulring for patiently enduring the multitude of questions related to my data collection and this thesis. My gratitude goes out to the faculty and staff of the Digital Imaging Remote Sensing group and those participants in data collection that made this research feasible.

Completion of this work would not have been possible without the help and support of all those who were always willing to give their time and valuable assistance towards the completion of this thesis. Finally, my sincere thanks and appreciation goes to the United States Air Force for providing me with the opportunity to earn a graduate degree while serving my country. I appreciate the emphasis that our senior leaders have placed on education and hope that this program will continue to provide future officer's with a similar opportunity.

Above all, my deepest gratitude goes to my family for helping and supporting me through school, as well as to my girlfriend, for her encouragement and patience. Without a doubt, they are the keys to my success.

# The Development of a Performance Assessment Methodology for Activity Based Intelligence: A Study of Spatial, Temporal, and Multimodal Considerations

by

Christian M. Lewis

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Master of Science Degree
at the Rochester Institute of Technology

## Abstract

Activity Based Intelligence (ABI) is the derivation of information from a series of individual actions, interactions, and transactions being recorded over a period of time. This usually occurs in Motion imagery and/or Full Motion Video. Due to the growth of unmanned aerial systems technology and the preponderance of mobile video devices, more interest has developed in analyzing people's actions and interactions in these video streams. Currently only visually subjective quality metrics exist for determining the utility of these data in detecting specific activities. One common misconception is that ABI boils down to a simple resolution problem; more pixels and higher frame rates are better. Increasing resolution simply provides more data, not necessary more information. As part of this research, an experiment was designed and performed to address this assumption. Nine sensors consisting of four modalities were place on top of the Chester F. Carlson Center for Imaging Science in order to record a group of participants executing a scripted set of activities. The multimodal characteristics include data from the visible, long-wave infrared, multispectral, and polarimetric regimes. The activities the participants were scripted to cover a wide range of spatial and temporal interactions (i.e. walking, jogging, and a group sporting event). As with any large data acquisition, only a subset of this data was analyzed for this research. Specifically, a walking object exchange scenario and simulated RPG. In order to analyze this data, several steps of preparation occurred. The data were spatially and temporally registered; the individual modalities were fused; a tracking algorithm was implemented, and an activity detection algorithm was applied. To develop a performance assessment for these activities a series of spatial and temporal degradations were performed. Upon completion of this work, the ground truth ABI dataset will be released to the community for further analysis.

*I dedicate this work to all the children who grow up dreaming beyond the constraints of their environment.*

*To the kids on the playground who consistently take the "you can'ts" and change them into "I did's".*

*To the youth on the streets whose healthy measure of self-doubt only serves to bolster their drive for success, rather than defeat it.*

*And to the young men and women who weren't discouraged by being raised within a society of two-parent values–without the accompanying two-parent household;*

*I dedicate this work to you.*

*Let this simply serve as inadequate measure of your capacity for success.*

*Yours,*
*Someone who was told he could not succeed . . .*
*but did anyway*

## DISCLAIMER

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

# Contents

# List of Figures

# List of Tables

# Abbreviations

Remote Sensing

| | |
|---|---|
| **AoI** | **A**ctivity **o**f **I**nterest |
| **DoLP** | **D**egree **o**f **L**inear **P**olarization |
| **FOV** | **F**ield **O**f **V**iew |
| **GCP** | **G**round **C**ontrol **P**oints |
| **GIQE** | **G**eneral **I**mage **Q**uality **E**quation |
| **GRD** | **G**round **R**esolved **D**istnace |
| **GSD** | **G**round **S**ample **D**istnace |
| **HSI** | **H**yper **S**pectral **I**maging |
| **IR** | **I**nfra**R**ed |
| **LiDAR** | **L**ight **D**etection **A**nd **R**anging |
| **LWIR** | **L**ong **W**ave **I**nfra**R**ed |
| **MAPPS** | **M**ultispectral **A**erial **P**assive **P**olarimeter **S**ystem |
| **MSI** | **M**ulti-**S**pectral **I**maging |
| **NIIRS** | **N**ational **I**mage **I**nterpretability **R**ating **S**cale |
| **PI** | **P**olarimetric **I**nformation |
| **SAM** | **S**pectral **A**ngle **M**apper |
| **SSE** | **S**um **S**quare **E**rror |
| **VNIIRS** | **V**ideo **N**ational **I**mage **I**nterpretability **R**ating **S**cale |
| **WASP** | **W**ildfire **A**irborne **S**ensing **P**laftorm |

Computer Vision

| | |
|---|---|
| **FMV** | **F**ull **M**otion **V**ideo |
| **MI** | **M**otion **I**magery |

| | |
|---|---|
| **OpenCV** | **O**pen source **C**omputer **V**ision |
| **RGB** | **R**ed **G**reen **B**lue |

Department of Defense

| | |
|---|---|
| **DoD** | **D**epartment **o**f **D**efense |
| **RPG** | **R**ocket **P**ropelled **G**renade |

Other

| | |
|---|---|
| **CIS** | Chester F. Carlson **C**enter for **I**maging **S**cience |
| **PVC** | **P**oly**v**inyl **C**hloride |
| **RIT** | **R**ochester **I**nstitute **T**echonology |

# Symbols

| | | |
|---|---|---|
| $E$ | entropy | $^{J}/_{K}$ |
| $fr$ | frame rate | Hz |
| $GSD$ | ground sample distance | $^{cm}/_{pix}$ |
| $P$ | probability | % |
| $t$ | time | s, frames |
| $v$ | velocity | $^{m}/_{s}$ |
| $x$ | distance | m |

# Chapter 1

# Introduction

The intent of this work is to produce a performance assessment methodology for a new research domain known as Activity Based Intelligence (ABI). This performance assessment will consider spatial, temporal, and multimodal characteristics of physical systems when detecting activities of interest.

## 1.1 Motivation

In today's intelligence environment, sophisticated sensors are collecting larger volumes of video data over ever increasing ground swaths. The purpose is to image as many objects and actions, over as much time as possible in hopes that this aggregated data can be efficiently analyzed to produce useful information. One drawback to this age of ever expanding data is the need for someone to sift through the data. The increase in both sensors and the number of unmanned aerial systems has produced an explosion of data since 2009. Estimates indicate that each year the military acquires over "24 years' worth [of video data] if watched continuously" [22–25]. Some have estimated that this information grows at an exponential rate with increases in stored data expected to exceed 1000 exabytes (1 million terabytes) biannually [26]. Military commanders have been cited as saying "We have enough sensors," but not enough people to analyze the results, "automating the process is essential to managing the data flood" [24]. In some operations, this deluge of data has already led to unfortunate consequences in theatre [27].

This "more is better" misconception is not exclusive to our nation's military. Generally speaking, in today's market it is presumed that bigger is better, regardless of where or how the technology will be used. Camera phones provide an example. The "Mega Pixel War" began with the inclusion of cameras in cell phones and has remained the predominant quantitative metric for consumers to compare cell phone cameras to one another [28]. More pixels and higher frame rates will produce crisper images and less choppy videos. The increase in pixel count has, among other things, increased the necessary storage, without a noticeable increase in quality for most consumers [29]. To their credit, some consumers have realized that simply increasing spatial and temporal resolutions within their cell phones does not necessarily provide them with more information from their cell phones. Manufacturers have begun to shift their emphasis from placing more pixels in imagery to providing more information from imagery. For example, Google is working on a smart phone capable of performing 3D mapping of its environment [30]. Like the military commanders, some in these emerging markets have begun developing tools to analyze the activities that occur within the data [31]. This is the domain of Activity Based Intelligence.

In 2012 the Director of National Intelligence, James Clapper, indicated that ABI is not something we should be striving for, it should be a way of information gathering that we already do. [32] Further stating that "in addition to predicting actions of the future, we should have the agility and ability to perform real-time tipping and cueing based to current threats. That dynamic ability to respond is what we now call Activity Based Intelligence (ABI)" [32]. In a broad sense, ABI is concerned with the actions, interactions, and transactions of people as they move through a given scene. These activities can be complex multi-actor situations where the actions of individuals and groups are tracked, segmented, characterized, and analyzed for points of interest or as simple as two people passing by one another in an area under surveillance. The premise behind this concept is the ability to automate a series of algorithms to cue analysts towards specific times in video streams where events of interest have occurred.

However, using any sensor to derive intelligence from a particular scene is highly contingent on knowing the type of activities that are of interest. The size and speed of a target produce requirements on the type of sensor that is capable of capturing the actions those targets produce. Therefore there is an inherent link between what you are

capturing and the characteristics of the sensor performing the capture. This extends to capturing activities caused by the interactions of multiple targets.

With such a large trade space, it is nearly impossible for individuals to factor in all necessary constraints in order to optimize sensor placement and tasking. As such, part of the intent of this thesis is to learn what these constraints are by developing a common dataset involving both rudimentary and complex interactions between actors and objects in a real-world scene.

A multi-spatial, multi-temporal, multimodal tradespace will be developed to attempt to parse the problem of activity analysis and yield quantifiable results. This research will also lay the mathematical foundation required to research and develop future remote sensing systems intended for ABI-type missions. Once complete, this performance assessment methodology will provide mission planners with a tool to help determine which sensor assets should be utilized when searching for a given Activity of Interest (AoI). This implies mission planners will have access to at least one algorithm to search for each AoI under a variety of sensor requirements. A notional activity lookup table is depicted in Figure 1.1.

This ABI lookup table will continue to expand as researchers developed new techniques to evaluate activities in motion imagery. Each tuned to operate under a specific set of environmental, weather, illumination, and sensor conditions. A sufficiently robust lookup table could allow users to operate in a variety of capacities. These may range from law enforcement averting gang activity in urban environments to humanitarian missions searching for survivors during natural disasters.

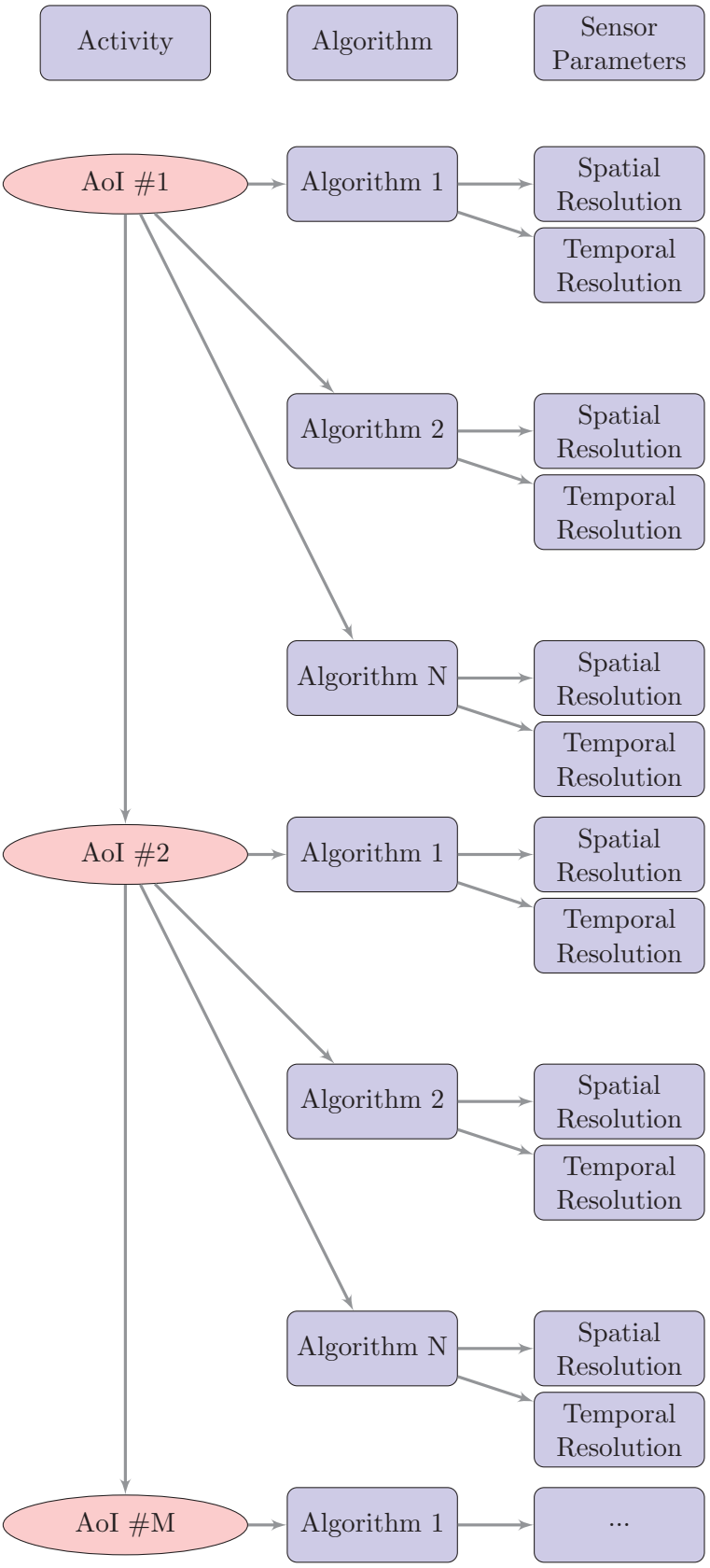



FIGURE 1.1: Notional ABI Lookup Table

## 1.2 System Acquisitions

The novelty of the Activity Based Intelligence domain means individuals attempting to solve an ABI task are faced with an unknown phenomenology, but a known physical domain. That being the case, many opt to take a route of transforming the unknown phenomenology into one more familiar. For example, if an aerial platform were searching for a car in an empty parking lot during the day, they need only make some assumptions to develop a tractable problem. The car has a predefined size, high contrast with its background, and can be seen with visible sensors. Now two metrics known as Ground Sampling Distance (GSD) and Signal-to-Noise (SNR) can be guessed and fed into an image quality equation. This will produce a requirement for the type of imaging system necessary to find said target.

However, if you were interested in finding the same car performing donuts or figure eights in the parking lot, then you would not have much to go on because the activity itself is ill-defined. Knowing that it is still a car in the same parking lot would lead you to produce the same metrics and image quality analysis. You may then be tempted to **improve** the previous results to compensate for the unknown of the situation- lower GSD and SNR. That has been the methodology going forward for technological advancements when the implementation of the advancement is not understood. Figure 1.2 graphically depicts this concept in action.

## 1.3 Trade Space

In the broadest sense, trade studies are used to access the complex interaction of varying capabilities with a predefined set of constraints. This modeling affords developers the ability to determine the ideal set of conditions under which experiments, missions, and technology should progress forward. The trade space presented here examines the optimal conditions at which activities can be characterized given a series of remote sensing modalities over a range of temporal resolutions. By focusing on a specific AoI, the performance assessment methodology can develop a notional set of spatial, temporal, and multimodal sensor parameters which would provide a high probability of detecting the activity.

FIGURE 1.2: Mapping unknown phenomenology to known phenomenology

### 1.3.1 Temporal

As technology advances, so too does the capability of capturing images at a faster rate. It is certainly possible to continue upgrading sensor platforms with the latest technology such that temporal resolution rates continue to increase without bound. That begs the question, are these platforms watching objects that move at such high speeds, that it justifies the cost of upgrading this system? It is assumed that many activities of interest will involve people and modern day vehicles. Knowing that, it stands to reason that each of these categories has a maximum speed at which it can move. Once a framing system has been developed that can match the speed of the AoI, there should be less motivation to continue increasing temporal resolution.

Furthermore, having high frame rate imaging systems has brought on the well known issue of "big data" [22–25]. Innovative solutions are currently being developed to address this issue, but if the problem that originally spawned it is not curbed, this could grow out of control. There are already more hours of data being produced than will be possible to watch in the lifetimes of our current analysts [23].

A methodical analysis of this trade space is proposed to construct the framework by which future developers can determine the necessary frame rate of new imaging systems.

### 1.3.2 Spatial

As stated above, consumers of technology may not know how to assess the utility of the technology they use. As with cell phone cameras, they may simply assume more is better [28]. Military and law enforcement are not exceptions. The recent advent of ARGUS, a 1.8 gigapixel DARPA initiative to design a sensor to provide a persistent stare capability across a roughly 40 square kilometer area, has left analysts with the same problem as the preponderance of UAV data; there is too much of it [25]. Figure 1.3 depicts a notional concept of the ARGUS imaging system.

In the author's opinion, one goal in the development of this system was to ensure that "all" data can be collected, rather than understanding what data needs collecting. While this provides a modest leap in technology, it still places the burden of turning this data into information squarely on the analysts.

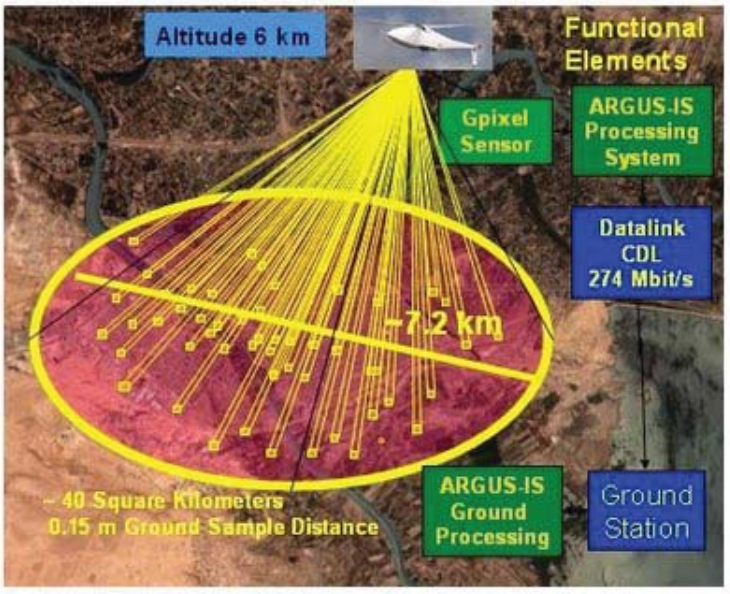



FIGURE 1.3: ARGUS concept image

This research will provide a methodology of assessing the spatial requirements of such a system that links back to the mission goals.

### 1.3.3 Multimodal

There are many different types of sensors currently in operation and under development, however there exist no requirements for what types of sensors will be necessary for future intelligence capabilities. Thus far the old adage, "bigger is better" has given the community a myopic view on how and what technologies should be developed for tomorrow [25, 28]. This has left many without a real set of future requirements stemming from the future operational purpose.

If a particular object of interest needed to be tracked utilizing a series of Motion Imagery (MI) sensor platforms, which platforms should be tasked? Along with that, what would the requirements be if one of those platforms could be incrementally upgraded to perform a specific mission? Part of the reason these questions exist is so the research and development community can have a common focus on the development of future systems.

While it is understood that innovation for innovation's sake is an admirable and requisite component in technology development, it should not be the only component. This

research will develop a framework whereby future developers and requirements managers can begin to understand the vast modality trade space. This comprehension would then allow intelligent, informed decision making in the acquisition of future sensor platforms.

# Chapter 2

# Objectives

## 2.1 Problem Statement

Two questions drove this research: Is it possible to utilize a series of multimodal sensors in a semi- or fully- automated fashion to develop intelligence based on the activities within a given scene? If so, can an objective performance assessment be developed to determine if a sensor is capable of detecting specific AoIs in motion imagery?

## 2.2 Research Objectives

The objectives of this research are twofold: To develop a semi- or fully-automated method of identifying activities within motion imagery, and to produce a performance assessment methodology whereby future researchers can understand the tradespace necessary to find specific AoIs in motion imagery.

Each activity recognition algorithm would have an associated "likelihood of detection" graph indicating how it will perform under specific spatio-temporal sensor characteristics; Figure 2.1 depicts this notional concept. For multimodal situations, Figure 2.2 depicts a similar graph that would be used to determine the optimal combination of sensors for detecting the AoI.

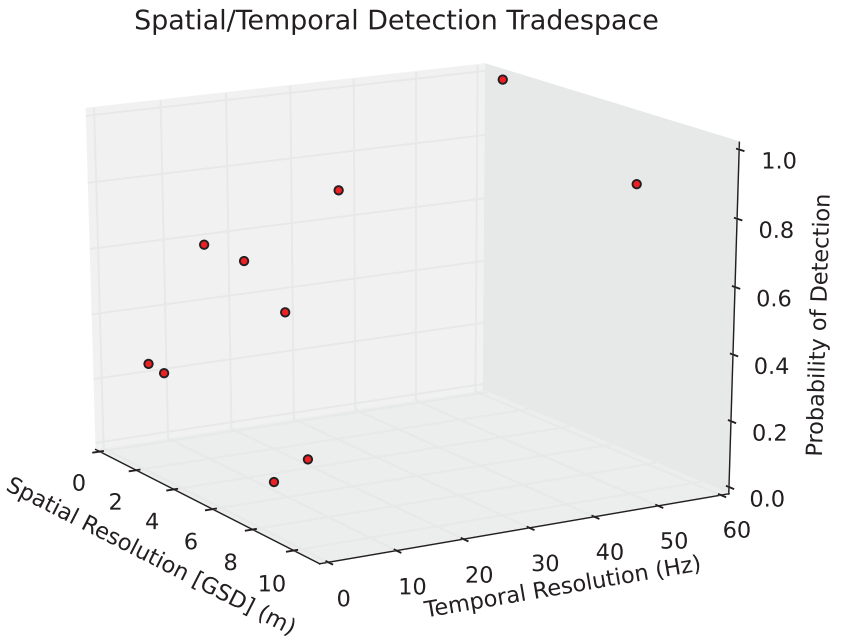



FIGURE 2.1: Spatio-Temporal Detection Trade Space

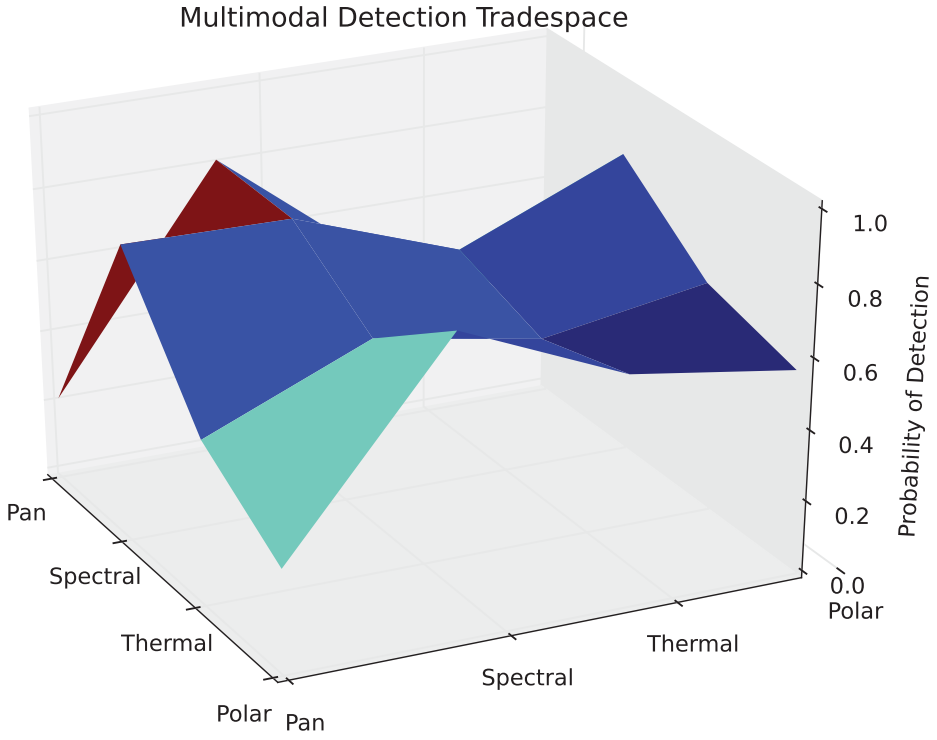



FIGURE 2.2: Multimodal Detection Trade Space

Each activity would have a list of algorithms capable of performing the recognition with varying levels of success. Sensor parameters would dictate the type of activities that could be perceived while environmental conditions would impact the likelihood of detecting the activity. Figure 2.3 expands the lookup table in Figure 1.1 by concentrating on the factors that determine the utility of each technique. By the conclusion of this research, at least one algorithm should be included for the chosen AoI.

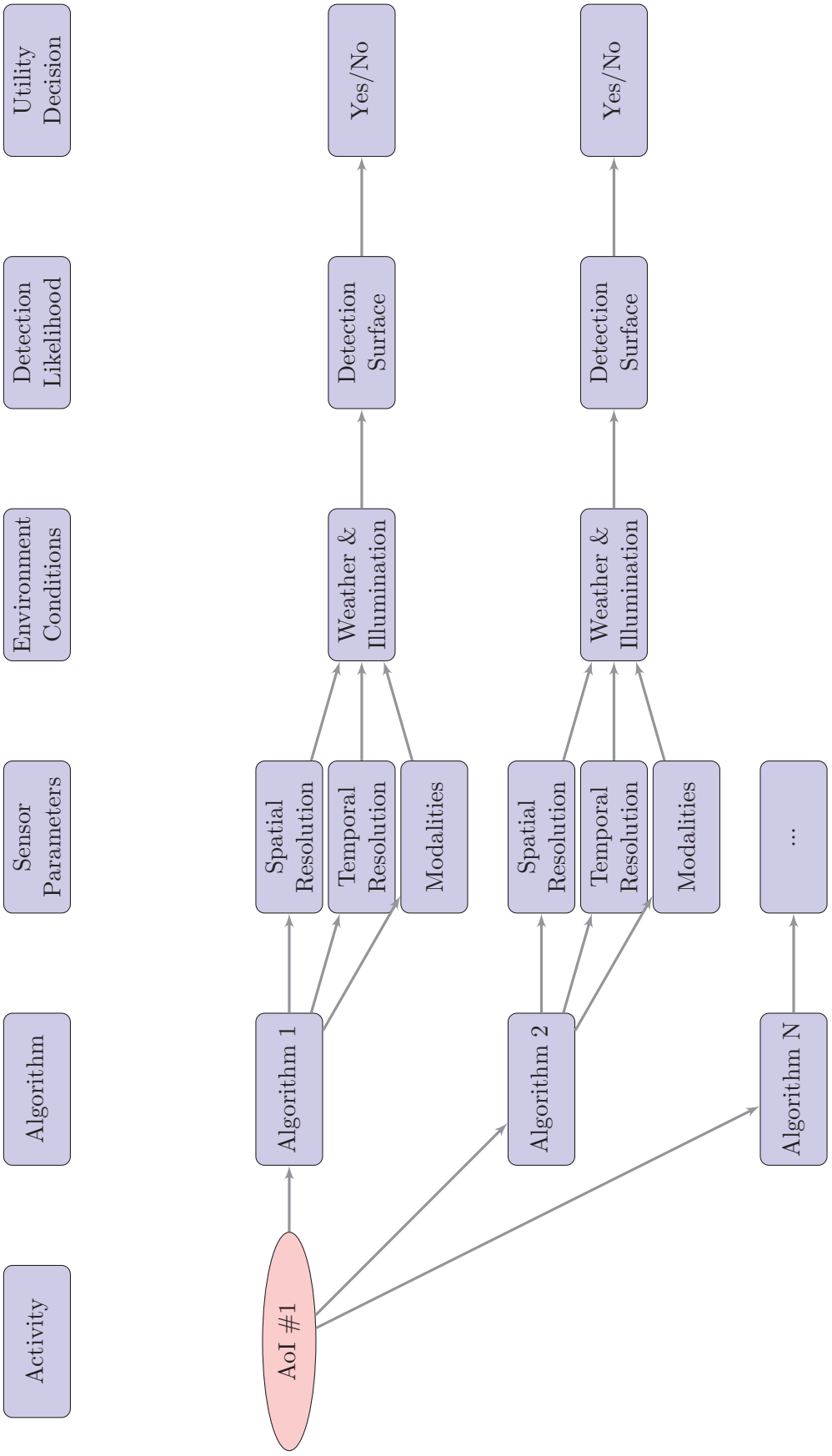



FIGURE 2.3: Notional Algorithm Lookup Table for a Given Activity

## 2.3 Tasks

Due the unique nature of this work, there exists no dataset which can be used to accomplish the research. Thus, including designing an experiment there are several steps required to complete the objectives of this research; they are:

1. Design ABI Experiment
2. Camera Calibration
3. Video Stabilization
4. Registration
5. Data Fusion
6. Tracking
7. Activity Recognition
8. Tradespace Development

## 2.4 Contributions to the Field

There currently exists no method, semi- or fully-automated, whereby activity based intelligence is developed from multi-sensor multimodal data. In addition, while there has been preliminary research into the area of activity based intelligence, there has been no consideration of the possibility of using multimodal data to augment standard visible and panchromatic sensors.

Specific contributions to the field of study will be:

- Development of a multimodal ABI dataset
- An end-to-end ABI evaluation of one activity
- Development of a limited multimodal ABI trade space
- Setting the foundation for an ABI lookup table

# Chapter 3

# Background

## 3.1 Activity Based Intelligence

Activity Based Intelligence is a developing field, notionally defined as: the inference of information from agent based interactions, occurring in a multi-temporal environment. It is primarily concerned with the actions, interactions, and exchanges of people within a scene of interest. These interactions and exchanges are then used to develop relationships between the individuals in the scene to identify actions and patterns of life.

It should be emphasized that ABI is dependent on the temporal nature of datasets. If you were to take a still photo of a crowd at the mall, it could be difficult or impossible to determine the relationships of entities within the scene. If instead if you were to capture video data, these relationships may become much more apparent. Another important aspect of temporal data is the resolution at which the data is acquired. Using the same mall example, if you took an image a day, you would perceive a very different world than if you were to take an image every hour. The same could be said decreasing from hours to minutes, and even minutes to seconds. Time lapsed photography provides an example of this concept. Figures 3.1 and 3.2 depict two forms of time lapsed photography at different rates. The first is an image of a daylily blooming over a period of 24 hours whereas the second image is that of an individual performing a stunt on a motorized bike likely lasting no longer than several seconds.

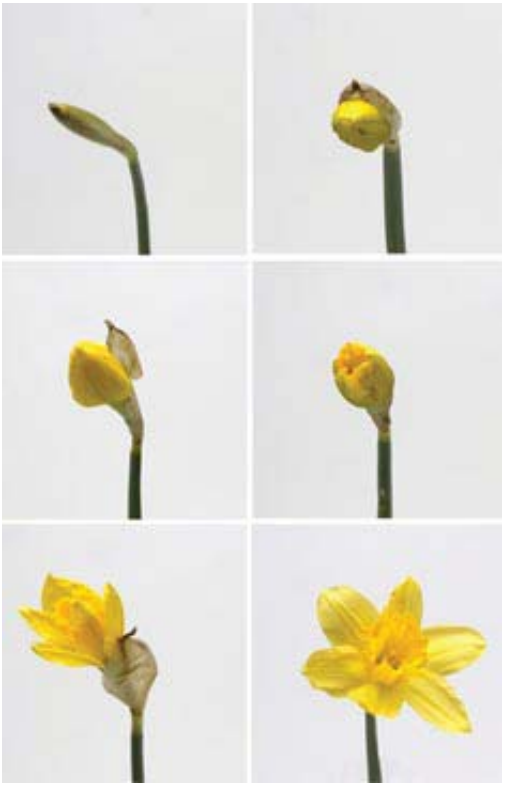


FIGURE 3.1: Kodak capture of a blooming flower [1]

FIGURE 3.2: Bike stunt [2]

The dependence on the temporal nature of the activity and the capabilities of the sensor are key to understanding what type of events can be captured with a particular imager. Section 4.4 will discuss how the actors and objects, in this dataset, were utilized and why.

### 3.1.1 State of the Field

Currently, operational ABI is a manually intensive process whereby analysts sift through large quantities of video data to develop the relationships among the individuals within the scenes. In the context of intelligence, it could be stated that this type of video analytics traces its roots to the days of photo interpretation of images from satellite imaging systems. Analysts were needed to sift through the imagery to determine the state of a nation based on its military assets, infrastructure, and even its crop production. As technology advanced, faster frame rates were possible, leading to what we now call motion imagery or video data. The proliferation of imaging equipment and video cameras has led to many forms of analysis in attempts to characterize our environment. Thermal images of blocks in New York City can be used to determine heat dissipation rates and associated electricity consumption [33]. Also, the advent of social media has led to network-based analysis that relates digital "traffic" to real world events [34]. A recent article in The Economist spoke to the ease of acquiring and launching nanosatellites carrying terrestrial (smartphone) imaging equipment [35]. This proliferation of technology has led to an explosion of analysis capabilities. The state of the field is constantly evolving.

## 3.2 Quality Metrics

Quality metrics are used as a method of evaluation to determine the utility of a particular technology to accomplish a task. Some common quality metrics of modern age computing are processing power (CPU clock speed), memory, and graphics capabilities. In cell phones, a set of quality metrics may include camera pixel size, screen resolution, or on-board storage space. In cars, quality metrics of performance may include top speed or torque.

With each technological breakthrough, people want a method of comparing similar products and ultimately knowing which product is better, or the best value. One of the recent issues with quality metrics stems from a consumerism which recognizes more as better. More processing power, higher pixel counts, and increased torque values drive our idea of performance in today's market, and yet those metrics may be irrelevant to our needs.

Since the inception of the cell phone camera in the early 2000s, mobile device manufacturers have engaged in what has been called "the megapixel war" [36]. This competition amongst manufacturers began when increasing the pixel count produced a noticeable improvement in the quality of images from cell phones. As technology improvements allowed manufacturers to place more pixels in cameras, consumers continued to assume that more pixels meant a product was better. The caveat to this trend was yes, more pixels can be better, but only if you need them. The continual improvement of imaging sensor technology and the need for its evaluation led to the development of a quality metric to compare image quality in a more objective manner. This metric was called the General Image Quality Equation (GIQE).

### 3.2.1 General Image Quality Equation (GIQE)

In order to quantify image quality, a regression-based model was developed using a collection of fundamental image and sensor attributes. This general image quality equation (GIQE) utilizes these attributes to produce a numerical rating on what is now known as the National Imagery Interpretability Rating Scale (NIIRS). These attributes are: scale, as expressed via the Ground Sample Distance of the system; sharpness, as measured by the Modulation Transfer Function (MTF) of the image; and Signal-to-Noise (SNR). Leachtenauer, et al developed the analytical form of of NIIRS as

$$NIIRS = 10.251 - a\log_{10} GSD_{GM} + b\log_{10} RER_{GM} - (0.656 \cdot H) - (0.344 \cdot {}^{G}/_{SNR}) \quad (3.1)$$

where a, and b are regressed coefficients, RER is relative edge response, H is a corrective overshoot parameter derived from the Modulation Transfer Function Correction (MTFC), and G is the noise gain of the system. This form was developed by having 10 image analysts rate 359 visible images for their quality. The regression of their results had an $R^2$ value of 0.934 and standard deviation of 0.38 which indicates the equation to be a good fit for the data.

#### 3.2.1.1 Ground Sample Distance (GSD)

Ground sampling distance is defined as the smallest distance between points on the ground that is distinguishable by a sensor. It is a geometric relationship using similar triangles that relates the GSD and the pixel pitch through the altitude (Alt) of the sensor and the focal length of the optical train. This relationship is calculated by

$$\frac{GSD}{Alt} = \frac{p}{f} \tag{3.2}$$

where $Alt$ is the altitude of the sensor, $p$ is the pixel pitch, and $f$ is the focal length. If a sensor is looking off nadir, a slant range term $R$, and corresponding angle, replaces the altitude term as show in equation (3.3)

$$R = {}^{Alt}/_{\cos\theta} \tag{3.3}$$

where $\theta$ represents the look angle of the system. Note this works even at nadir as a zero angular extent forces the cosine term to become one, thereby causing the slant range to simply become the altitude. Equation (3.2) represents the case where the sensor is nadir looking and the slant range equals the altitude. However, equation (3.4) is a more accurate representation.

$$\frac{GSD}{R} = \frac{p}{f} \tag{3.4}$$

The geometric GSD is calculated by multiplying the x and y components of the GSD and applying an angular extent $\alpha$ for non-square focal plane arrays. This is represented in its analytical form as

$$GSD_{GM} = \left[GSD_X \cdot GSD_Y \cdot \sin\alpha\right]^{1/2} \tag{3.5}$$

#### 3.2.1.2 Relative Edge Response (RER)

The relative edge response is a measure of how fast the pixel values change when going from one side of an edge to another. Figure 3.3 depicts this measure.

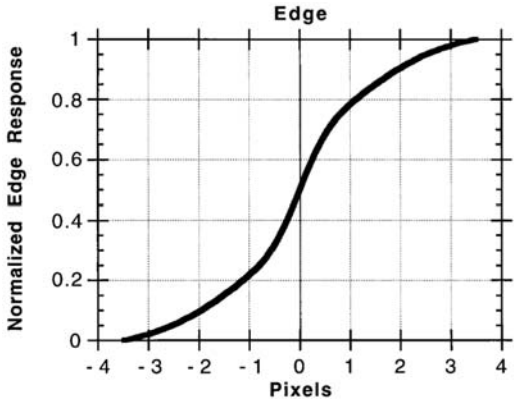



FIGURE 3.3: Relative Edge Response [3]

This value (RER) is the slope of the system's edge response.

#### 3.2.1.3 Overshoot correction (H)

The overshoot-height-based term accounts for the overshoot of the edge-response function due to the Modulation Transfer Function Correction (MTFC) factor. Take Figure 3.4 as an example. Case 1 occurs before the MTFC is applied to the dataset and case 2 after the correction has been applied. Using position 1.5 there is a 0.4 difference in the edge response of the two cases. This overshoot is captured in the overshoot correction term H. This term is measured over a range of 1.0 to 3.0 pixels from the edge in quarter pixel increments.

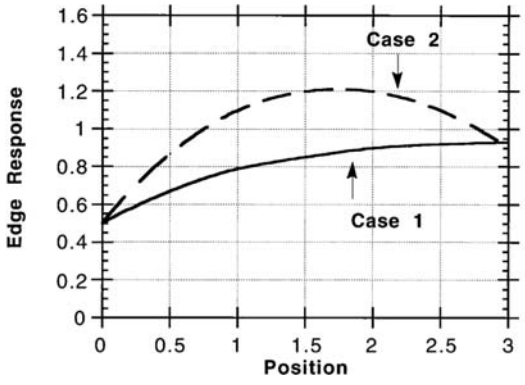



FIGURE 3.4: Overshoot [3]

#### 3.2.1.4 Noise Gain (G)

This term accounts for the noise gain induced by the MTFC and is computed by taking the Root Sum Square (RSS) of the MTFC Kernel as

$$G = \left[ \sum_{i=1}^{M} \sum_{j=1}^{N} (kernal_{ij})^2 \right]^{1/2} \tag{3.6}$$

#### 3.2.1.5 Signal-to-Noise Ratio (SNR)

The SNR is described as the "ratio of the noise of the dc differential scene radiance to the noise of the rms electrons computed before the MTFC and after calibration." [3] The analytic form was developed as

$$SNR = {}^{S}/_{N} \tag{3.7}$$

where S is the mean or peak signal of an image and N is the corresponding noise.

### 3.2.2 National Image Interpretability Rating Scale (NIIRS)

The National Image Interpretability Rating Scale (NIIRS) is the product of the GIQE equation, and is a method of mapping the results of the equation to real world items. It is a 10-level rating scale which analysts now use to quantitatively indicate their imaging needs. The full scale is presented in Figure 3.5.

**Table 1. Visible NIIRS Operations by Level—March 1994**[a]

### Rating Level 0

Interpretability of the imagery is precluded by obscuration, degradation, or very poor resolution.

### Rating Level 1

Detect a medium-sized port facility and/or distinguish between taxiways and runways at a large airfield.

### Rating Level 2

Detect large hangars at airfields.

Detect large static radars (e.g., AN/FPS-85, COBRA DANE, PECHORA, HENHOUSE).

Detect military training areas.

Identify an SA-5 site based on road pattern and overall site configuration.

Detect large buildings at a naval facility (e.g., warehouses, construction halls).

Detect large buildings (e.g., hospitals, factories).

### Rating Level 3

Identify the wing configuration (e.g., straight, swept, delta) of all large aircraft (e.g., 707, CONCORD, BEAR, BLACKJACK).

Identify radar and guidance areas at a SAM site by the configuration, mounds, and presence of concrete aprons.

Detect a helipad by the configuration and markings.

Detect the presence/absence of support vehicles at a mobile missile base.

Identify a large surface ship in port by type (e.g., cruiser, auxiliary ship, noncombatant/merchant).

Detect trains or strings of standard rolling stock on railroad tracks (not individual cars).

### Rating Level 4

Identify all large fighters by type (e.g., FENCER, FOXBAT, F-15, F-14).

Detect the presence of large individual radar antennas (e.g., TALL KING).

Identify, by general type, tracked vehicles, field artillery, large river crossing equipment, wheeled vehicles when in groups.

Detect an open missile silo door.

Determine the shape of the bow (pointed or blunt/rounded) on a medium-sized submarine (e.g., ROMEO, HAN, Type 209, CHARLIE II, ECHO II, VICTOR II/III).

Identify individual tracks, rail pairs, control towers, switching points in rail yards.

### Rating Level 5

Distinguish between a MIDAS and a CANDID by the presence of refueling equipment (e.g., pedestal and wing pod).

Identify radar as vehicle-mounted or trailer-mounted.

Identify, by type, deployed tactical SSM systems (e.g., FROG, SS-21, SCUD).

Distinguish between SS-25 mobile missile TEL and Missile Support Van (MSV) in a known support base, when not covered by camouflage.

Identify TOP STEER or TOP SAIL air surveillance radar on KIROV-, SOVREMENNY-, KIEV-, SLAVA-, MOSKVA-, KARA-, or KRESTA-II-class vessels.

Identify individual rail cars by type (e.g., gondola, flat, box) and/or locomotive by type (e.g., steam, diesel).

### Rating Level 6

Distinguish between models of small/medium helicopters (e.g., HELIX A from HELIX B from HELIX C, HIND D from HIND E, HAZE A from HAZE B from HAZE C).

Identify the shape of antennas on EW/GCI/ACQ radars as parabolic, parabolic with clipped corners or rectangular.

Identify the spare tire on a medium-sized truck.

Distinguish between SA-6, SA-11, and SA-17 missile airframes.

Identify individual launcher covers (8) of vertically launched SA-N-6 on SLAVA-class vessels.

Identify automobiles as sedans or station wagons.

### Rating Level 7

Identify fitments and fairings on a fighter-sized aircraft (e.g., FULCRUM, FOXHOUND).

Identify ports, ladders, vents on electronics vans.

Detect the mount for antitank guided missiles (e.g., SAGGER on BMP-1).

Detect details of the silo door hinging mechanism on Type III-F, III-G, and III-H launch silos and Type III-X launch control silos.

Identify the individual tubes of the RBU on KIROV-, KARA-, KRIVAK-class vessels.

Identify individual rail ties.

### Rating Level 8

Identify the rivet lines on bomber aircraft.

Detect horn-shaped and W-shapted antennas mounted atop BACKTRAP and BACKNET radars.

Identify a hand-held SAM (e.g., SA-7/14, REDEYE, STINGER).

Identify joints and welds on a TEL or TELAR.

Detect winch cables on deck-mounted cranes.

Identify windshield wipers on a vehicle.

### Rating Level 9

Differentiate cross-slot from single slot heads on aircraft skin panel fasteners.

Identify small light-toned ceramic insulators that connect wires of an antenna canopy.

Identify vehicle registration numbers (VRN) on trucks.

Identify screws and bolts on missile components.

Identify braid of ropes (1 to 3 inches in diameter).

Detect individual spikes in railroad ties.

[a]The information in this table was previously published in Ref. 3.

10 November 1997 / Vol. 36, No. 32 / APPLIED OPTICS 8323

FIGURE 3.5: National Image Interpretability Rating Scale (NIIRS) [3]

This rating scale merges the metrics used by intelligence analysts into a numerical classification in order to relate their needs to technical systems. Four categories are utilized

by analysts in this assessment:

- Detection: Identify object from its surroundings
- Classification: target vs. non-target
- Recognition: functional category (i.e. tank)
- Identification: Target is (i.e. this is a M60)

This broad-based categorization works well on traditional imaging systems operating in the visible regime. As a result of its ubiquotous use, NIIRS began to drive R&D of future systems by indicating whether a system would or would not be able to meet a specific imaging need. It also led to a few other NIIRS-esque rating scales specific to other modalities. This includes an IR-NIIRS, a Multispectral NIIRS, and a Video NIIRS. Neither the IR nor the Multispectral NIIRS will be discussed here, but their rating scales are included in appendix A.

### 3.2.3 Video NIIRS (VNIIRS)

In what appeared to be a natural extension, the still imagery quality metric was expanded for use within the multi temporal domain by Young et al [4]. However, by simply evaluating motion imagery (MI) by still imagery metrics, you lose the inherent advantage gained by having a time changing series. Young noted this, saying: "rating motion imagery using only static criteria lacks content validity ... motion imagery exploitation is concerned with timing and sequence of events" [4].

It is this concept of a "sequence of events" that lead to the development of activity based intelligence, as we are concerned with how objects act and interact with one another. In an attempt to apply a quantitative set of criteria to events of interest Young et al [4] came up with a set of VNIIRs task requirements; which can be seen in Figure 3.6. They developed this scale by having 63 motion imagery analysts judge 13 images from a set of 73 in total. The specifics of the analysis can be found in the Young et al paper entitled Video National Imagery Interpretability Rating Scale Criteria Survey Results [4]. The regression performance indicated one statistical deviation of a t-value equivalent to 0.02.

**Table 2 Selected V-NIIRS Criteria Frame Rate Requirement (10X Temporal Sampling Rule)**

| V-NIIRS | V-NIIRS Task | V-NIIRS Criteria Object | V-NIIRS Criteria Action (implied in italics) | Maneuver/ Event Duration (sec) | Minimum Sampling Rate (FPS) (10X Rule) |
|---|---|---|---|---|---|
| 3 | Visually track | convoy | *Driving in formation* | 2.7 | 4 |
| 4 | Visually track | tracked vehicles | *Driving in formation* | 2.1 | 5 |
| 5 | Visually confirm | the turret on a main battle tank | as the main gun slews during training, live fire exercise, or combat | 1.6 | 6 |
| 6 | Visually track | an identified vehicle type: car, SUV, van, pickup truck | driving independently | 1.2 | 8 |
| 7 | Visually confirm | unidentified deck-borne objects | as they are dumped over the side or stern | 0.9 | 11 |
| 8 | Visually confirm | an individual holding a shoulder fired anti-aircraft missile | as the launcher is raised to the aimed firing position | 0.7 | 14 |
| 9 | Visually confirm | the body & limbs of an individual holding a long rifle or sniper rifle | as the weapon is raised to an aimed firing position -either standing, sitting, or prone | 0.6 | 18 |
| 10 | Visually confirm | the hands and forearms of an individual holding a compact assault weapon or large frame handgun | as the weapon is raised to an aimed firing position -either standing, sitting, or prone | 0.4 | 23 |
| 11 | Visually confirm | individual's fingers and hands while aiming a shoulder fired anti tank missile | as they release safety and arm the device | 0.3 | 30 |

FIGURE 3.6: Video National Image Interpretability Rating Scale (NIIRS) [4]

Along with this rating scale, there was an attempt align the NIIRS and VNIIRS criteria. Figure 3.7 depicts this comparison of scales. The VNIIRS system was the first attempt at driving system requirements from the actions of objects and individuals within the scene.

Young also noted that utilizing time series data can lead to advances in spatial recognition: "activity discernment can lead to object recognition at spatial resolution levels less than what is required in still imagery." [4] In fact, he and his co-authors indicated an improvement of object recognition of up to 1/4 of a NIIRS rating [4]. It is currently being used to assess compression and codecs [37] and is leading to the development of a Motion Image Quality Equation (MIQE) [38, 39].

VNIIIRS defines image quality by asking two questions:

1) Can you classify the objects within the scene?
2) Can you recognize the actions occurring between the objects?

By reviewing Figure 3.6 it should become apparent that the metrics of classification and recognition are solely based on subjective visual recognition of data in the visible regime. While this concept of a video rating scale gives analysts a way to compare video streams, it still locks the analysts into the loop by requiring human recognition. The explosion of video data discussed in Section 1.1 means that this manually intensive process will only

**Table 1 Comparison of Selected NIIRS Criteria to V-NIIRS**

| NIIRS | NIIRS Criteria Task and Object | NIIRS Criteria Context | V-NIIRS | V-NIIRS Task and Object | V-NIIRS Criteria Object | V-NIIRS Criteria Action (implied in italics) | V-NIIRS Criteria Context |
|---|---|---|---|---|---|---|---|
| 3 | Identify a large surface ship by type. | In port. | 3 | Visually track the movement of | Convoy of intermediate-range ballistic missile (IRBM) transporter and support vehicles | *Making turn* | on an improved road near missile base, launch site or silo |
| 4 | Identify, by general type, tracked vehicles, field artillery, large river crossing equipment | when in groups | 4 | Visually track the movement of | individual, tracked engineering vehicles and wheeled prime mover/trailer combinations | *Making turn* | during tactical road march/deployment in the field or on an unpaved road |
| 5 | Distinguish between SS-25 mobile missile TEL and Missile Support Vans (MSVs) | in a known support base, when not covered by camouflage. | 5 | Visually confirm the rotation of | the turret on a main battle tank | as the main gun slews during training, live fire exercise, or combat | at a gunnery range, field deployment site, or battle zone |
| 6 | Identify automobiles as sedans or station wagons | - | 6 | Visually track the movement of | an identified vehicle type: car, SUV, van, pickup truck | driving independently | on roadways in medium traffic |
| 7 | Identify individual railroad ties | - | 7 | Visually confirm the movement of | unidentified deck-borne objects | as they are dumped over the side or stern | of any surface ship or fishing vessel at sea |
| 8 | Identify a hand-held SAM (e.g. SA-7/14, REDEYE, STINGER) | - | 8 | Visually confirm the movement of | an individual holding a shoulder fired anti-aircraft missile | as the launcher is raised to the aimed firing position | in the field, in a defensive position, or in the vicinity of an airfield or airport approaches |
| 9 | Identify cargo (e.g. shovels, rakes, ladders) | in a open-bed, light-duty truck. | 9 | Visually confirm the movement of | the body & limbs of an individual holding a long rifle or sniper rifle | as the weapon is raised to an aimed firing position -either standing, sitting, or prone | At a practice range, during live fire exercise, or during an engagement. |
| - | - | - | 10 | Visually confirm the movement of | the hands and forearms of an individual holding a compact assault weapon or large frame handgun | as the weapon is raised to an aimed firing position -either standing, crouched, or prone | At a practice range, during live fire exercise, or during an engagement |
| | | | 11 | Visually confirm the movement of | individual's fingers and hands while aiming a shoulder fired anti tank missile | as they release safety and arm the device | at a tactical position in a rural or urban environment |

FIGURE 3.7: VNIIRS - NIIRS Comparison [4]

become worse as time goes on. This rating scale also lacks the novelty of incorporating higher order interactions. It attempts to address the needs of the community for which it was made, by simply extending the previous NIIRS categories into the temporal domain of motion imagery.

**Action vs. Activity Recognition** Since the word "action" has come up, a digression is made to make a distinction between action recognition and activity recognition. Action recognition is generally concerned with the motions of a single individual within

a given sequence, whereas activity recognition is concerned with the interactions that individuals have in the environment and with others in the scene. An example of action recognition would be identifying someone waving their hand, whereas activity recognition would be concerned with the activity of two people saying "hello" by waving their hands.

**Motion Imagery vs. Full Motion Video** Motion imagery is a term used to describe any dataset of imagery that was captured at a rate of 1Hz or faster. Historically speaking, Full Motion Video (FMV) has been a subset of motion imagery that operates at frame rates similar to those of televisions; between 24Hz and 60Hz. [40]

#### 3.2.3.1 Spatial Degradations (GSD vs GRD)

In order to discuss the spatial degradations that occurred in this dataset, a distinction between Ground Sampling Distance (GSD) and Ground Resolved Distance (GRD) must first be made. Rearranging Equation (3.4) in terms of GSD

$$GSD = \frac{R \cdot p}{f} \tag{3.8}$$

where the slant range, pixel pitch, and focal length are represented by $R$, $p$, and $f$ respectively. By keeping the slant range constant, it is possible to change the GSD by either altering the pitch pitch, focal length, or some combination thereof. Altering the pixel pitch effectively changes the sampling rate at which the detector can physically collect data. Assuming a unity fill factor, decreasing the pixel pitch has the effect of sampling the ground at smaller distances, thereby allowing distinction between smaller objects. Increasing the pixel pitch has the opposite effect of reducing the distinction between objects. For example, with a 5cm GSD, two objects placed 6cm apart are generally distinguishable, whereas the same two objects would not distinguishable if the GSD were changed to 10cm.

Using a non-exotic lens, the focal length affects the angular extent (FOV) that can be perceived within the scene. As the focal length increases, the FOV decreases, effectively spreading the information in the smaller FOV across the focal plane array. This spread

of information stipulates that the objects within the scene are occupying more pixels, effectively being sampled more often. Figure 3.8 depicts this concept using 18, 34, and 55mm lenses [5].

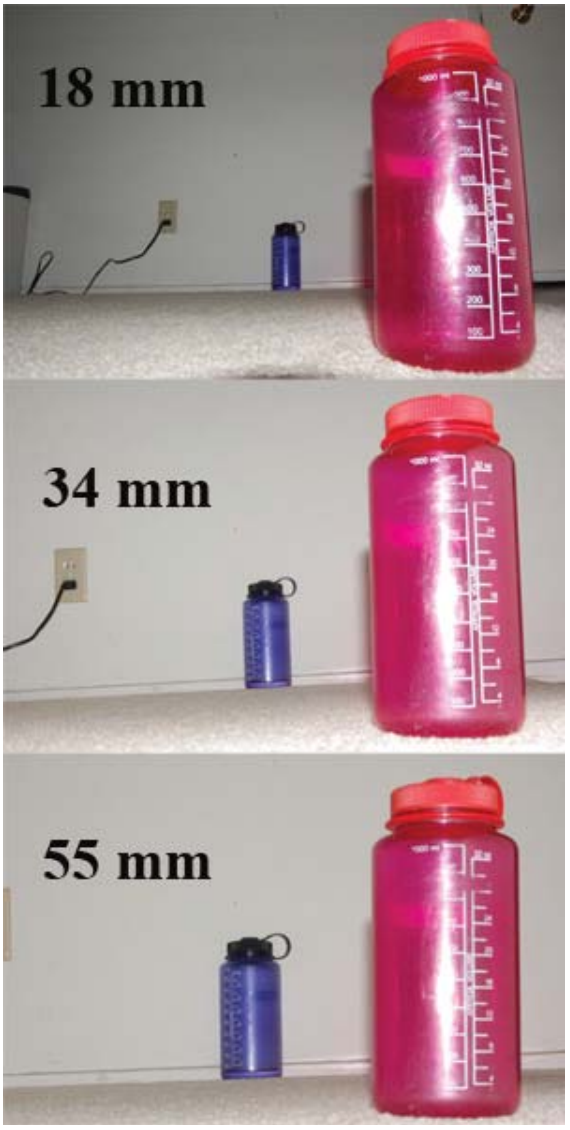



FIGURE 3.8: Focal Length and FOV [5]

In order to effectively simulate a reduction in GSD using one of the two aforementioned parameters, a few steps would need to be completed. Reducing the pixel pitch requires a general blurring of the data and downsampling to simulate the loss in sampling and mixing of the information. For example, performing a 2x reduction could be done by blurring a 2x2 square and downsampling it to a 1x1 pixel. The blur could be done by taking a mean between the four pixels or a Gaussian of the pixels within a larger extent but still nearby. Reducing the GSD by focal length requires a similar procedure, whereby

the image is blurred and down sampled, but differs in that it adds image content around the edges of the original image. This would essentially reduce the size of this image and place it within a larger image. The difference between the two techniques lies in the size of the focal plane array. Reducing the pixel pitch generally means reducing the size of the array, again assuming a unity fill factor, as larger pixels would be used to sample the image. However, reducing the GSD by increasing the focal length leaves the array unchanged by increasing the FOV of the sensor.

## 3.3 Multimodal Trade Space

There are several modalities that could be exploited to characterize AoI within a given scene. The applicable and available modalities for the problem at hand include: panchromatic imaging, multispectral imaging, hyperspectral imaging, polarimetric imaging, thermal imaging, Light Detection And Ranging (LiDAR) imaging, and Synthetic Aperture Radar (SAR). In the context of this research, each has its own strengths and weaknesses, which will be discussed below. This review is designed to provide a brief overview of each modality in order to evaluate its perceived utility in activity recognition. Once chose the modality will be incorporated into an experiment designed to develop the data for this research.

### 3.3.1 Panchromatic

Panchromatic imaging provides a good basis when working across different modalities for several reasons. Since it integrates across a broad band, the SNR of your imaging system is higher than many other sensing modalities. This increase in SNR can inversely allow for detector designers to decrease the pixel sizes within the detector, thereby increasing the GSD of the sensor. This increase provides a higher spatial resolution, which can make spatial feature detection and multimodal registration a more tractable task. It is, however, its broadband nature that reduces its usefulness in distinguishing unique characteristics of objects within the scene. As RIT currently possesses these capabilities, both in sensor and in simulation, this modality will be included in this research.

### 3.3.2 Multispectral

Multispectral imaging provides a method whereby objects within a given scene can more easily be discriminated due to the differences in their spectral signatures. This signature can be used to track objects spectrally, which is helpful when spatial segmentation may not be possible. Since RIT maintains these capabilities, both in sensor and in simulation, this modality will be included in this research. As a point of reference, there are current efforts by Bartlett et al [41] to develop motion imagery hyperspectral/polarimetric capture systems.

Ideally, the more distinction in signatures the more able the tracking algorithm will be to keep targets separate from one another. Thus hyperspectral imaging would be more desirable than multispectral imaging. However, RIT did not have a readily available hyperspectral imager at the time of this research. It has in the past utilized such technology, but the necessary time to acquire and utilize said devices was prohibitive. Therefore, multispectral imaging will suffice.

### 3.3.3 Polarimetric

Polarimetric imaging provides a method of discriminating objects whose surface and subsurface reflections cause light to change its orientation relative to its surrounds. This affords ready discernment of manmade objects from natural backgrounds [42]. Other research has shown the ability to perform object classification within a scene [43]. This modality was incorporated into this experiment for its ability to distinguish targets from natural backgrounds and due to its ability to separate objects of differing polarimetric characteristics.

Polarimetric imagery can be developed by placing a polarimetric filter in front of an imaging device. A common configuration is to have a spinning wheel with two, three, or four filters with varying angular filter orientations. Linear filters are created by placing parallel bars of conductive material at close intervals inside of a thin transmissive lens. Orienting the bars horizontally causes them to absorb horizontal electromagnetic (EM) radiation and transmit vertical EM radiation. By controlling the orientation of the filter it is possible to determine if objects in the environment favor a particular orientation. The modified Pickering method combines this orientation information to develop a polarization vector known as the Stokes vector [42]. This is written as

$$\begin{aligned} S_0 &= \frac{(E_0 + E_{45} + E_{90} + E_{135})}{2} \\ S_1 &= E_0 - E_{90} \\ S_2 &= E_{45} - E_{135} \end{aligned} \tag{3.9}$$

$$\mathbf{S} = \begin{bmatrix} S_0/S_0 \\ S_1/S_0 \\ S_2/S_0 \end{bmatrix} = \begin{bmatrix} 1 \\ \tilde{S}_1 \\ \tilde{S}_2 \end{bmatrix} \tag{3.10}$$

with $E_0$ through $E_{135}$ representing the image as seen through the four polarization filters. The numerical designation is the angle of the polarized filter. $S_0$ represents the total energy of the image, $S_1$ represents the energy difference between horizontal and vertical polarization states, and $S_2$ represents the difference between the energy in the 45 degree and 135 degree states.

A unique aspect of this modality is the ability to fuse multiple polarimetric orientations together to develop more advanced products. Two of these include the Degree of Polarization (DoP) and Degree of Linear Polarization (DoLP) as described by

$$DoP = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \tag{3.11}$$

$$DoLP = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \tag{3.12}$$

$$DoP \approx DoLP = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \tag{3.13}$$

with $E_0$ through $E_{135}$ representing the image as seen through the four polarization filters. The numerical designation is the angle of the polarized filter. $S_0$ represents the total energy of the image, $S_1$ represents the energy difference between horizontal and vertical polarization states, and $S_2$ represents the difference between the energy in the 45 degree and 135 degree states.

It is common for the four polarimetric images to be taken at different times due to the need to change filters between images. This process is called a "Division of Time" and has the benefit of using the entire focal plane array to collect data. The downside is the need to register the images to perform the DoLP and DoP evaluations. Recent research is taking advantage of the of ability to place small filters directly on the focal plane

array, alleviating the need for registration [41]. This "Division of Area" has the benefit of capturing data for all four polarization states at once. A drawback is the need to demosaic the output to reconstruct four full polarimetric images.

### 3.3.4 Thermal

Thermal imaging affords the capability of using temperature and emissivity as distinguishers between objects within a given scene. This is beneficial to this research for two reasons. First, since objects will not be placed in the scene until the time of the experiment, their innate temperatures will likely be different from those in the background. Second, specific objects can be chosen such that their emissivities afford them distinguishing characteristics from the surrounding scene. Within an ABI scenario, this data can be useful in performing multimodal registration, tracking, and activity recognition.

### 3.3.5 Light Detection and Ranging (LiDAR)

Light Detection and Ranging (LiDAR) provides a high resolution 3-dimensional model of an environment of interest. This could be useful in distinguishing specific objects within a scene, as it provides depth to the imagery. A few challenges exist with using this dataset though. First, the currently available LiDAR sensors require multiple seconds to build up a full 3D model of the scene. This prevents it from being useful in detecting activities that occur on time scales less than its ability to capture scenes. While a problem, that would not prevent it from being used. Further, regarding its capture rate, it is unclear how moving objects within the scene would affect the scene capture. The second challenge is incorporating LIDAR data with the other modalities. This would require registration and fusion of 2- and 3- dimensional data. This is possible by either creating point clouds from the 2D imagery or directly fusing the 2D imagery onto the 3D points of the LiDAR dataset [44–47]. Rather than fusing 2D imagery with 3D point clouds, it may be possible to perform tracking on the LiDAR data itself [48, 49]. While possible, recent work has indicated that neither of these techniques are mature enough for use within the temporal constraints of this thesis.

### 3.3.6 Synthetic Aperture Radar (SAR)

SAR provides an interesting capability. There has been research conducted into tracking and target recognition of manmade objects in urban and non-urban scenes which would make SAR a valuable addition to this dataset [50–53]. In order to incorporate this modality there would need to exist a multimodal dataset wherein specifically coordinated activities have been captured by SAR and other modalities concurrently. An alternate method would require a robust simulation capability that provides SAR, and other modalities the ability to image a scene characterized by predetermined activities. As neither currently exists, to the knowledge of the researcher, and a SAR system cannot be readily procured, this modality ruled out as a possibility for this work.

## 3.4 Registration

Registration is the process of transferring different datasets into a common coordinate system. In this research, the transfer (or transformation) needs to occur in both the spatial and temporal domains.

### 3.4.1 Spatial Registration

Image registration appears to be the most prominent method of transferring different datasets or images into a common coordinate system [54]. In this process we are attempting to overlay two images of the same scene that are taken at different times, from different sensors, and potentially from differing perspectives. Currently, several methods exist for accomplishing this task, including information, frequency, and feature-based approaches. Information based methods attempt to align the information content of two separate images by taking a rolling product of the images until the maximum entropy is reached. Frequency methods take the spatial content of an image to the frequency domain and use the shift theorem to align the frequencies of the two images, thereby producing the misalignment translations. Feature based methods use specific features within the images to indicate those points are in fact the same point in space. This research elected to use a feature-based method for registering the data.

#### 3.4.1.1 Speeded Up Robust Features (SURF)

SURF features, are unique scale and rotation-invariant descriptors that identify specific points in imagery which are useful to registration. This is accomplished in three steps. First, interesting points are selected within an image. These points may be corners or abnormal objects within an image. Next, the neighborhood of these points is represented by a feature vector. This distinct descriptor is robust to noise, geometric transformations, and photometric transformations. Finally, point correspondences are formed by matching these interesting points and vectors across multiple images. This match is generally determined by some distance between the vectors, such as include Euclidean or Mahalanobis distance. [55]

As the specifics can be read in Bay (2008), only a top-level review of this algorithm will be provided. Interest points are derived from a Hessian based matrix

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \tag{3.14}$$

where $L_{xx}(x,\sigma)$ is the convolution of the image with the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ .

The scale space is used to find scale invariant features. This is accomplished by upscaling the filter size rather than iteratively reducing the image size. These filters scale images by a factor of two in a parallel fashion since each works on the original image rather than a successive scale space image. This space is further divided into octaves to represent a series of filter response maps. [55]

The descriptor of the interesting points also describes the distribution of intensity content in the neighborhood of the point. A reproducible orientation is identified for each interest point by calculating the Haar wavelet response in the x and y directions within a circular neighborhood defined by a radius of six times the sampling step. This circular region is set to encompass a 16x16 orientation specific feature vector. This method has proved to be robust, reliable, and repeatable in its uses among a series of images. [55]

#### 3.4.1.2 Mutual Information Theory

Mutual information is a method of relating the entropy structures of the underlying images. Fan, Rhody, and Saber explain this technique in their paper on Airborne Image Registration [56], but it can be quickly explained as:

$$\begin{aligned} E[A] &= -\sum_{i=1}^{m} P_A(a_i) \log_2 P_A(a_i) \\ I(A,B) &= E[A] + E[B] - E[A,B] \\ R(A,B) &= \frac{I(A,B)}{E[A] + E[B]} \end{aligned} \tag{3.15}$$

where $P_A$ is the probability of a pixel value occurring within an image. A and B are images, E[A] and E[B] are the entropy associated with each image, I(A,B) is the mutual information, and R(A,B) is the scaled version of the mutual information image. The maximization of mutual information can also be attained by finding the spatial shift which maximizes the image intensities. This is done by applying a Fourier transform to the entropy images to the frequency domain and taking the difference of their phases. This is shown by

$$\begin{aligned} \mathscr{F}\{E[A]\} &= (...)e^{-2\pi i(x_A),(y_A)} \\ \mathscr{F}\{E[B]\} &= (...)e^{-2\pi i(x_B),(y_B)} \\ \frac{\mathscr{F}\{E[A]\}}{\mathscr{F}\{E[B]\}} &= (...)e^{-2\pi i(x_A,y_A)} \cdot (...)e^{-2\pi i(-x_B,-y_B)} \qquad (3.16) \\ &= (...)e^{-2\pi i(x_A - x_B, y_A - y_B)} \qquad (3.17) \end{aligned}$$

Taking an inverse Fourier transform returns the data to the spatial domain and presents the 2D x and y positional shifts. This is shown as

$$\begin{aligned}\mathscr{F}^{-1}\left\{\frac{\mathscr{F}\{E[A]\}}{\mathscr{F}\{E[B]\}}\right\} &= \mathscr{F}^{-1}\left\{(...)e^{-2\pi i(x_A-x_B,y_A-y_B)}\right\} \\ &= \delta(x_A - x_B, y_A - y_B) \end{aligned} \tag{3.18}$$

Using image A as the base image, the relative x and y translations can be attained.

$$\begin{aligned} x = x_A,\ y = y_A \\ x_0 = x_B,\ y_0 = y_B \\ \delta(x - x_0, y - y_0) \end{aligned} \tag{3.19}$$

where $x_0$ and $y_0$ are the x and y translations of the images to attain proper alignment. This correlation of information provides the maximum mutual information between the two images.

### 3.4.2 Temporal Registration

Similar to spatial registration, temporal registration is the transformation of different datasets into a common time-based coordinate system. This type of registration is needed when multiple video streams begin recording the same scene at different times or when they capture a different number of frames per second. Your essentially trying to match frames between the separate video streams.

## 3.5 Data Fusion

Data fusion is a method of taking different types of data and merging them to form information. A common example is fusing the sound of thunder with the sight of lightening to produce the conclusion that a storm is on the way. When considering image data, fusion can be accomplished at three distinct levels: pixel, feature, and decision. Since motion imagery is simply a compilation of time varying images, the same fusion

levels can be utilized. It is also noted that as with the NIIRS to VNIIRS extension, this simple extension of a single image technique to a multi-image sequence may not fully utilize the temporal characteristics of the data.

**Pixel Level** At the pixel level, data is correlated across multiple images by stacking the corresponding pixels behind one another. This common method is used when building multi- or hyperspectral data cubes. It is also the simplest to comprehend as it is directly correlating the lowest value of information from one image to another.

**Feature Level** At the feature level, specific points of interest across images are correlated as being the same or similar. This could occur if you were to take edge maps of two adjacent images and attempt to align the images by aligning the edges from one image to another. Features exists in a wide variety of descriptors and are generally just unique characteristics of a particular object in a scene. Facial detection algorithms can take advantage of the prominent features we call eyes, nose, and mouth to identify the approximate location of a face within an image.

**Decision Level** At the decision level, a specific technique has classified the information in both images and now looks to find some consensus amongst the classification. An example would involved merging the results of a clustering algorithm that was applied to two separate images. Another example could involve multispectral and polarimetric imagery, where a spectral anomaly detection algorithm and degree of linear polarization algorithm are separately used to identify points of interest within a scene. Once identified, a pixel-by-pixel weighting could be placed across the decision maps to determine which combination of pixels is both spectrally anomalous and polarimetric within the scene.

## 3.6 Tracking

Tracking occurs in two phases, first a target is detected and identified with a set of imagery. Next a maintenance step is used to correlate that target from one frame to the next.

### 3.6.1 Target Detection

Target detection is the isolation of pixels of interest from the remainder of the image at large. This basically turns into a computer vision problem, whereby the noise, in this case the background, needs to be reduced in favor of the targets of interest. Forsyth, Szelinski, and Solem [57–59] all discuss varying methods of filters, averaging, optical flows, and segmentation algorithms that could be utilized as possible solutions.

Two prior students at RIT, Zhang and Ausfeld [60, 61], utilized a difference image technique to create a foreground image for each frame of the video sequence. This foreground image essentially filters out stationary objects from moving objects.

### 3.6.2 Track Maintenance

After a set of objects are identified within each frame, an inter-frame association needs to be completed to determine how each object moved throughout the sequence. Blackman [62] explains a gating technique, whereby an object's velocity is used to predict how far it could move from one frame to another. This distance is then converted into a circular radius centered on the objects current position. Any objects in the future frame that are within this radius are considered possible updates to the current objects position. Figure 3.9 depicts this concept. Further distinctions between objects can be made by comparing the area of the detected objects in one frame to the areas of objects in future frames.

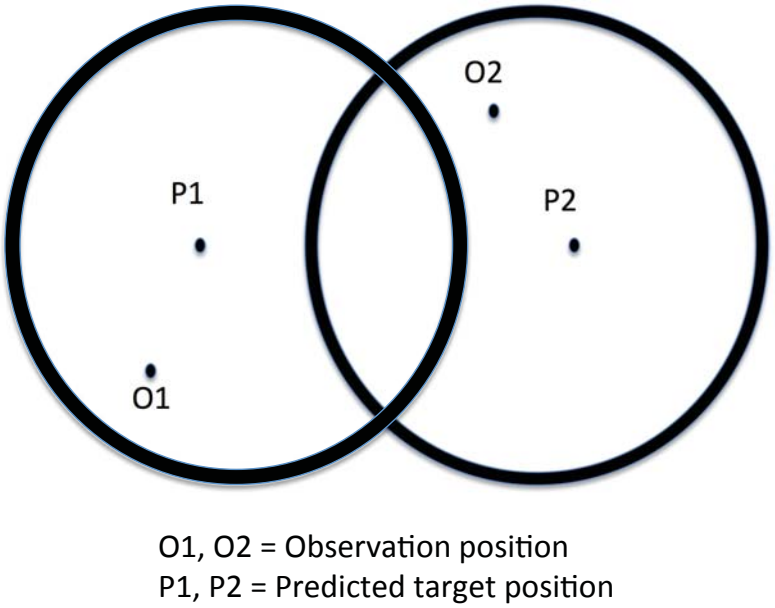



FIGURE 3.9: Gating Technique with Two Objects

## 3.7 Activity Recognition

As discussed in Section 3.2.3, VNIIRS can be considered an early activity recognition quality metric. However, being visually subjective and limited in scope of activities, it falls short of meeting the data analysis needs described in Chapter 1. However, it does take the step of deriving frame rate requirements for several of its activities. This type of analysis directly links the characteristics of the AoI to the sensor requirements necessary for capturing the AoI.

The novelty of the term Activity Based Intelligence means the most of the work done under this domain has been done in a series of well known names: Wide Area Motion Imagery (WAMI) [63], Patterns of Life [64], Social Network Analysis [65, 66], and Content-based Video Retrieval [67] to name a few. Other less prominent names include multi-target tracking, irregular warfare, and normalcy modeling. However, under each name the same basic activity recognition research has been performed.

Two authors have begun specifically addressing the use of activity recognition techniques in MI and Full Motion Video (FMV) [67, 68]. Particularly, Lash includes a high level discussion of the principles of MI and its applicability toward ABI. He continues with

some MI techniques for compressing and encoding data. Finally, he finishes with what he deems to be key technology enablers for ABI [68].

Several others have talked about using a technique called Space-Time Interest Points (STIPs) as a method for identifying specific segments within video sequences [69–72]. A STIP is a point within an MI sequence where objects are said to display unique characteristics in both space and time. A similar method of developing SIFT and SURF descriptors was applied here both spatially and temporally to determine STIP "corners" in the imagery. An example of one such corner would be a soccer ball hitting a goal post and rapidly changing direction.

Still others have discussed using spatial extents of people within an environment to develop actions and intents of actors [73]. This particular research placed a group of law enforcement officers in a prison setting and had them act out a series of high-threat inmate scenarios. The purpose of this research was to preemptively determine the imminent activity in hopes that a notification system could be set up to prevent it from occurring. Such activities include: multiple actors rapidly approaching one actor and large groups loitering in an aggressive fashion on the prison yard.

Others are using spatiotemporal data to detect patterns of life within imagery [64, 74]. These patters of life are used to develop normalcy models of a particular scene at some given time of day. Developing these models allows investigators to then extract abnormal patterns in the activity and identify behavior that needs further evaluation.

Additional ABI work includes recognizing human activity [75] within motion imagery and using graph theory approaches to detect activities within data [76]. A recent doctoral candidate reviewed event-based analytic techniques in the context of a computer science problem [77]. These are only a few examples of the several disparate domains working to develop the field of activity based intelligence that may not even know each other exist [31, 78–81].

## 3.8 Programming Languages

The work for this project was performed in several programing languages and software suites. Almost all of the work was done using the Python programming language and

Open source Computer Vision (OpenCV) library. This section simply serves to provide a basic reference for what tools were used.

**Python** The Python programming language is an object oriented language similar to that of C with an emphasis on readability. This high level opens source programming language focuses on software quality, coherence, developer productivity and a myriad of other qualities designed to making coding a relatively easy task. Its support library are maintained by the open source community and frequently updated.

**Open source Computer Vision (OpenCV)** One of the most useful tools developed for the Python coding language (among others) was the Open source Computer Vision library. This library was mainly developed to provide execution of real-time computer vision algorithms on a variety of platforms. The functionality of the library includes basic filtering operations, common tracking algorithms, and various other forms of image manipulation. The work performed in this research took advantage of several basic and few high levels tools for manipulating the motion imagery data.

# Chapter 4

# Experiment

## 4.1 Goals and Requirements

The purpose of this experiment was to develop a multimodal motion imagery dataset consisting of several AoIs. To accomplish this, several multimodal sensors were placed on the roof of a building overlooking a common walkway on a college campus. Several participants were then asked to act out a choreographed script of independent and group activities. The dataset was intentionally made large for distribution to the community for further evaluation.

The thermal, multispectral, and polarimetric modalities placed a set of requirements on the experiment that are discussed in the following sections. These requirements ranged from a spectral analysis of the contents within the scene to the inclusion of specific equipment for post-processing purposes. The unique nature of the sensors required an in-depth evaluation of the their independent and composite capabilities. Specific considerations included FOV constraints, GSD requirements, and physical proximity within the scene. The activities that were chosen also placed a set of constraints on the experiment as a whole. However, these constraints are mostly on the processing side. Those calculations will be discussed in Section 5.4.1 and logic behind the included activities will be discussed in Section 4.4.1.

In order to make a dataset available to the community, this experiment gathered more data than the author had time to evaluate. Thus, the experiment in its entirety will be

explained in this chapter and a section at the end will clearly state which portion of the problem is addressed in this research. Since the experiment occurred over a four day period, the specific time and conditions of the particular data used in this research will be included in the final section.

## 4.2 Equipment

Nine imagers, packaged into three sensor suites were used to capture the data for this experiment. Two of the three sensor suites was developed by the Digital Imaging and Remote Sensing (DIRS) group in the Chester F. Carlson Center for Imaging Science at the Rochester Institute of Technology, Rochester NY. The third was a commercial product purchased for its wide range of capabilities.

### 4.2.1 WASP-Lite

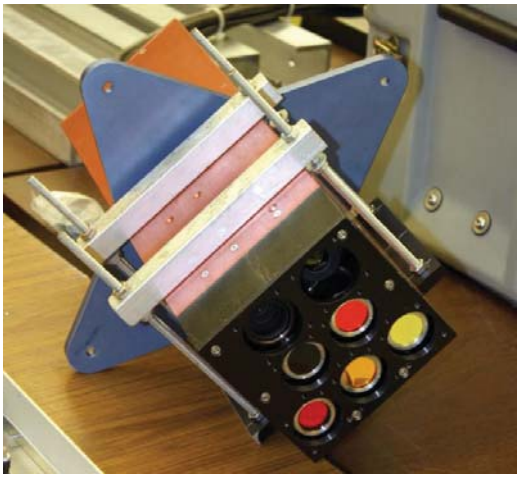


FIGURE 4.1: Wildfire Airborne Sensor Platform (WASP) [6]

The Wildfire Airborne Sensor Platform (WASP)-Lite consists of seven sensors encased in a single platform controlled together using an in-house software suite. Figure 4.1 depicts a full color view of the system, while Figure 4.2 shows a numbering of each sensor for further discussion. Each will be briefly introduced, along with the specifications relevant to the experiment being performed. It was designed to operate in a Cessna 172 flying at 3000ft with an airspeed of 90knots. Thus, many of these specifications are irrelevant to this discussion.

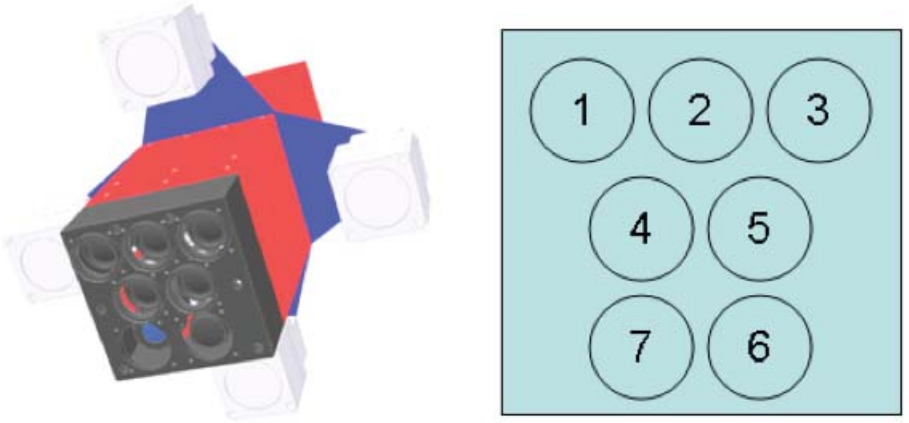



FIGURE 4.2: WASP Camera Identification [7]

Table 4.1 identifies each camera in figure 4.1 and indicates the spectral bandpass of each filtered sensor. The filters are 10$\mu$m wide, centered at the indicated filter bandpass.

TABLE 4.1: Experiment Equipment Specs

| WASP Label | Imaging System | Imager Bandpass | Filter Bandpass |
|---|---|---|---|
| Camera 1 | Spectral Imager 1 | 0.4-1.0$\mu$m | 630$\mu$m |
| Camera 2 | Spectral Imager 2 | 0.4-1.0$\mu$m | 550$\mu$m |
| Camera 3 | Spectral Imager 3 | 0.4-1.0$\mu$m | 436$\mu$m |
| Camera 4 | Spectral Imager 4 | 0.4-1.0$\mu$m | 650$\mu$m |
| Camera 5 | Spectral Imager 5 | 0.4-1.0$\mu$m | 670$\mu$m |
| Camera 6 | Hi-Res Panchromatic | 0.4-1.0$\mu$m | N/A |
| Camera 7 | LWIR | 8.0-12.0 $\mu$m | N/A |

The specific filter bandpasses were chosen based on the research of a pedestrian tracking effort completed by Herweg [8–10]. Figures 4.3 and 4.4 depict two sets of spectral reflectance values for pedestrians and common background materials in an outdoor scene. The filters were chosen to maximize contrast between the outdoor materials and our pedestrian participants by focusing on the distinctive and highly reflective nature of the pedestrians relative to other outdoor objects.

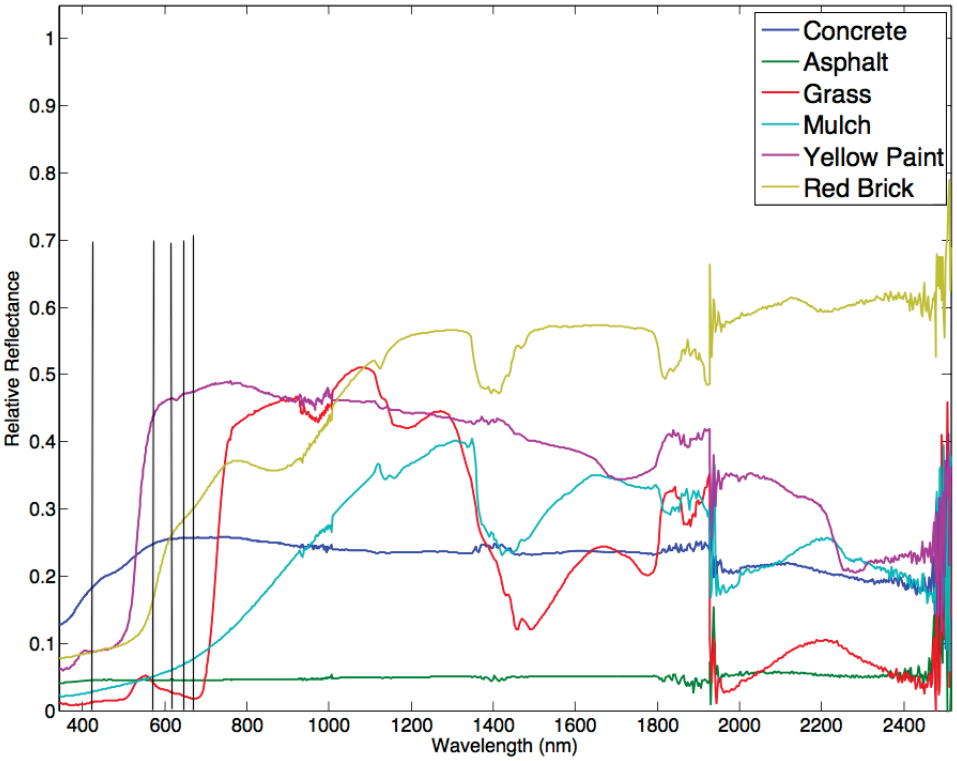



FIGURE 4.3: Reflectance Spectra of Background with Filter Centers Indicated by Vertical Lines [8–10]

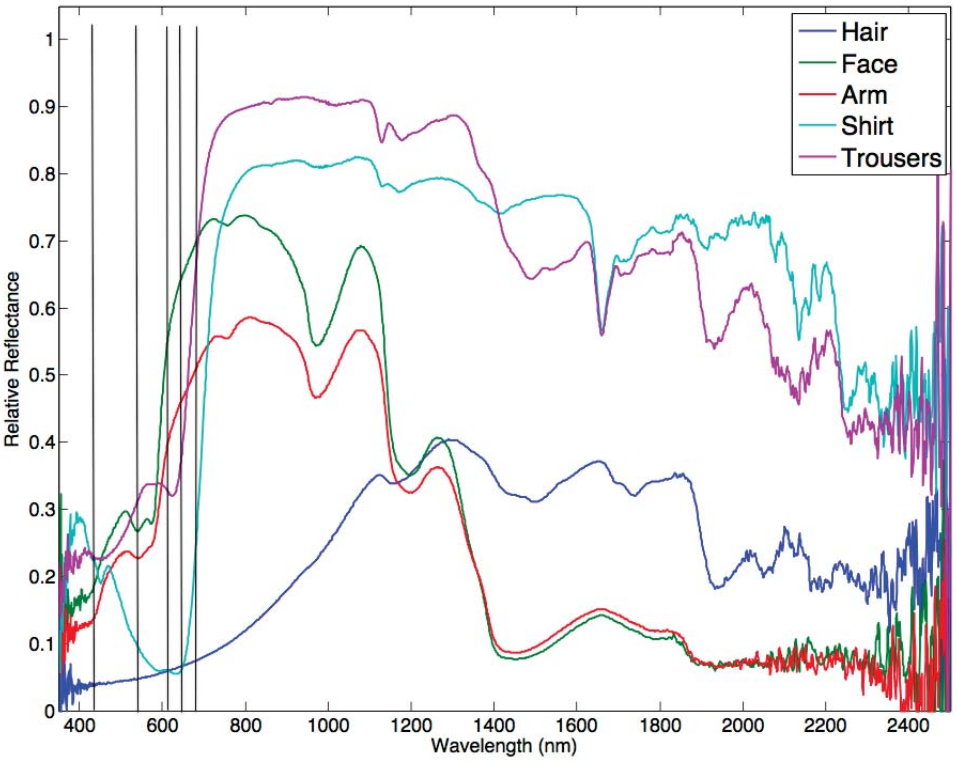



FIGURE 4.4: Reflectance Spectra of Pedestrians with Filter Centers Indicated by Vertical Lines [8–10]

### Panchromatic

Camera seven in figure 4.2 is the panchromatic sensor, designed for pan sharpening of the multispectral data. It is a Sony XCL-U1000 progressive line scanner with a 41823 Cinegon optical attachment made by Schneider [7]. It is assumed that the pixel pitch is equivalent to the pixel size unless explicitly stated.

Table 4.2: Panchromatic Camera Specifications [7, 17]

| Camera Attribute | Characteristic |
|---|---|
| Pixel Size | 4.4 x 4.4 $\mu$m |
| Array Size | 1628x1236 |
| Dynamic Range | 10 bits |

| Optics Attribute | Characteristic |
|---|---|
| Focal Length | 12mm |
| Focal Ratio (f/N) | 1.4-22 |
| Spectral Bandpass | 0.4-1.0$\mu$m |

### LWIR

Camera six in Figure 4.2 is the Long Wave Infrared (LWIR) sensor. It is a DRS E3500 uncooled Microbolometer Array with a proprietary optical interface [7]. Table 4.3 indicates the specifications of this imaging system.

Table 4.3: LWIR Camera Specifications [7, 17]

| Camera Attribute | Characteristic |
|---|---|
| Pixel Size | 25.4 x 25.4 $\mu$m |
| Array Size | 320 x 240 |
| Dynamic Range | 12 bits |

| Optics Attribute | Characteristic |
|---|---|
| Focal Length | 11mm |
| Focal Ratio (f/N) | 1.0 |
| Spectral Bandpass | 8.0-12.0 $\mu$m |

### Multispectral

Cameras 1-5, as indicated in Figure 4.2, are the multispectral sensors of the WASP-Lite imaging system. Table 4.4 indicates the characteristics of this system.

Table 4.4: Multispectral Camera Specifications [7, 17]

| Camera Attribute | Characteristic |
|---|---|
| Pixel Size | 7.4 x 7.4 $\mu$m |
| Array Size | 648 x 494 |
| Dynamic Range | 10 bits |

| Optics Attribute | Characteristic |
|---|---|
| Focal Length | 8mm |
| Focal Ratio (f/N) | 1.4-22 |
| Spectral Bandpass | 0.4-1.0$\mu$m |

### 4.2.2 MAPPS

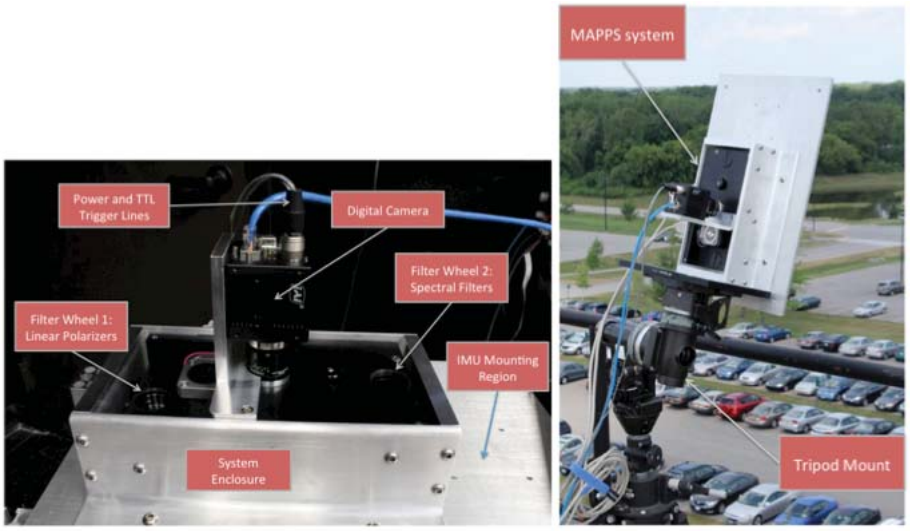



FIGURE 4.5: Multispectral Aerial Passive Polarimeter System (MAPPS) [11]

This Multispectral Aerial Passive Polarimeter System (MAPPS) is designed to produce high resolution spectral-polarimetric imagery by using a two spinning wheel design and a series of spectral bandpass and polarimetric filters. The use of spinning filter wheels makes this a Division of Time imager. The filters sit in a Sutter Lambda 10-3 dual filter wheel, capable of accommodating 10 filters per wheel. Before reaching the JAI BM-500GE CCD camera, light passes through the Schneider Optics lens. The camera specifications are listed in table 4.5. [11] The polarimetric spinning wheel is configured to cycle through the four polarimetric filters, then reverse direction and continue the sequence. Thus a full sequence collects 0, 45, 90, 135, 135, 90, 45, and 0 degree images in that order.

TABLE 4.5: MAPPS Camera Specifications [11, 18]

| Camera Attribute | Characteristic |
|---|---|
| Pixel Size | 3.45 x 3.45 $\mu$m |
| Array Size | 2456 x 2058 |
| Dynamic Range | 12 bits |

| Optics Attribute | Characteristic |
|---|---|
| Focal Length | 35mm |
| Focal Ratio (f/N) | xxxx |
| Spectral Bandpass | 0.4 - 1.0$\mu$m |


FIGURE 4.6: GoPro Hero 3: Black Edition [12]

### 4.2.3 GoPro

The GoPro imager is a commercial device developed and produced for use in sporting and other mobile events. It is housed in a small water resistant case and sold with an associated wireless transmitter. This RGB imager can operate at frames rates as fast as 120Hz and as slow as 24Hz. Its spatial resolution capabilities range from 240x240 to full 4K imagery. The specifications used in this dataset are listed in Table 4.6.

TABLE 4.6: GoPro 3 Hero Camera Specifications [19–21]

| Camera Attribute | Characteristic |
|---|---|
| Pixel Size | 1.55 x 1.55 $\mu$m |
| Array Size | 4000 x 3000 |
| Dynamic Range | 10 bits |

| Optics Attribute | Characteristic |
|---|---|
| Focal Length | 14mm |
| Focal Ratio (f/N) | f/2.8 |
| Spectral Bandpass | Visible |

Table 4.7 presents a side-by-side comparison of the specifications of all the equipment used in this experiment.

Table 4.7: Experiment Equipment Specifications

| Specifications | MAPPS [11, 18] | GoPro 3 [19] | WASP-Lite [7, 17] | | |
|---|---|---|---|---|---|
| | | | Panchromatic | LWIR | Multispectral |
| Pixel Pitch | 3.45$\mu$m | 1.55$\mu$m | 4.4$\mu$m | 25.4$\mu$m | 7.4$\mu$m |
| Frame Rate | 6Hz | 60Hz | 8Hz | | |
| FOV (deg) | 10 | 135 | 34.5 x 26.2 | 42.4 x 31.8 | 34.5 x 26.3 |
| Dynamic Range | 12 bits | 10 bits | 10 bits | 12 bits | 10 bits |

## 4.3 Experimental Setup

This section is designed to walk the reader through the steps necessary to set up the scene for the experimental collection. The following sections will go through the physical location, the scenario and actors in the experiment, as well as in-scene fiducials and meteorological conditions of the collection. All data collection was done at the Rochester Institute of Technology (RIT) in Rochester, NY between the hours of 10:00am and 4:30pm EST.

### 4.3.1 The Scene


FIGURE 4.7: Top view of experiment scene [13]

Figure 4.7 depicts an overhead view of the scene that was used in this experiment. The focus of the collection was a walkway in front of the Chester F. Carlson Center for Imaging Science (CIS) on the RIT campus. The previously described sensors were placed on the roof of the building looking down on the scene below. The participants were asked to accomplish a series of tasks on the walkways in front of the building. Figures 4.8 and 4.9 depict the locations of the sensors and participants respectively.


FIGURE 4.8: Sensor placement within scene

FIGURE 4.9: Participant routes within scene

The slant range was determined by using a Nikon N16184 Forestry Pro laser range finder. The distance from the equipment to the center of the walkway is 54m ±0.5m. Given that the sensors are not nadir-looking, the pixel sizes on the ground will change as a function of distance from the building. Using Equation (3.4) from Section 3.2.1.1, the GSD of each sensor can be calculated. Rearranging the terms of that equation, we

obtain one with GSD as a function of pixel pitch, focal length, and slant range written as

$$GSD = \frac{R \cdot p}{f} \tag{4.1}$$

By entering the values of each imager, located in Table 4.7, into Equation (4.1), the GSD can be calculated. For example, for the MAPPS sensor, the GSD is calculated as

$$\begin{aligned} GSD_{Mapps} &= \frac{R \cdot p_{Mapps}}{f_{Mapps}} \\ &= \frac{54.0m \cdot (3.45\text{E}{-6}m)}{35\text{E}{-3}m} \\ &= 0.00532m \\ &= 0.532cm \end{aligned}$$

Table 4.8 includes the GSDs of each of the sensors as set up.

TABLE 4.8: Equipment GSDs

| Attributes | MAPPS [11, 18] | GoPro 3 [19] | WASP-Lite [7, 17] | | |
|---|---|---|---|---|---|
| | | | Panchromatic | LWIR | Spectral |
| GSD(m) | 0.00532 | 0.00598 | 0.0198 | 0.125 | 0.0500 |
| GSD(cm) | 0.532 | 0.598 | 1.98 | 12.5 | 5.00 |

While this suggests that the GoPro has a better GSD than the WASP-Lite panchromatic imager, it does not take into account the fish eye lens attached to the former. Figures 4.10 and 4.11 depict the imagery side-by-side for comparison purposes; note the GoPro imagery has already been registered in this image. Figure 4.12 depicts a close-up of the white van vehicle. Notice how blurry GoPro image appears when compared to the panchromatic image.

FIGURE 4.10: Panchromatic image of scene

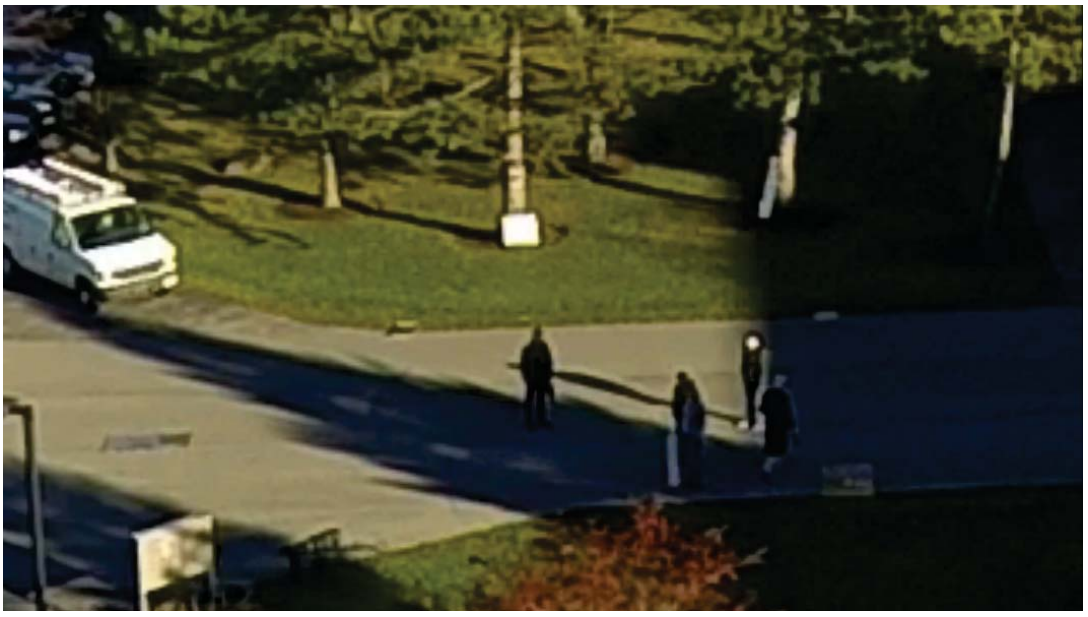


FIGURE 4.11: GoPro image of scene

(A) Panchromatic closeup

(B) GoPro closeup

FIGURE 4.12: Closeup comparison of truck in scene

### 4.3.2 Equipment Within the Scene

Within the scene, the equipment was set up on top of the CIS overlooking the walkway below. The height of the building is 14.5 meters. Figures 4.13 through 4.17 depict the setup of the equipment for the experiment. Only five of the ten images are shown in this section. The remaining images are left for view in Appendix C.

To reduce the amount of parallax in the imagery, the imagers were setup in close proximity to one another. Table 4.9 depicts the height, distance to the buildings edge, and rotations for each of the imagers. All dimensions was measured as close to the center point of the device as possible.

TABLE 4.9: Objects in Experiment

| Imaging System | Height | Distance to edge | Angles (Degrees) | | |
|---|---|---|---|---|---|
| | | | Roll | Pitch | Yaw |
| WASP Lite | 49” $\pm$ 1” | 81” $\pm$ 0.1” | $0.1 \pm 0.1$ | $17.0 \pm 0.1$ | $0.0 \pm 0.1$ |
| MAPPS | 49” $\pm$ 1” | 83” $\pm$ 0.1” | $2.9 \pm 0.1$ | $16.9 \pm 0.1$ | $0.0 \pm 0.1$ |
| GoPro | 55” $\pm$ 1” | 81” $\pm$ 0.1” | $0.1 \pm 0.1$ | $17.0 \pm 0.1$ | $0.0 \pm 0.1$ |

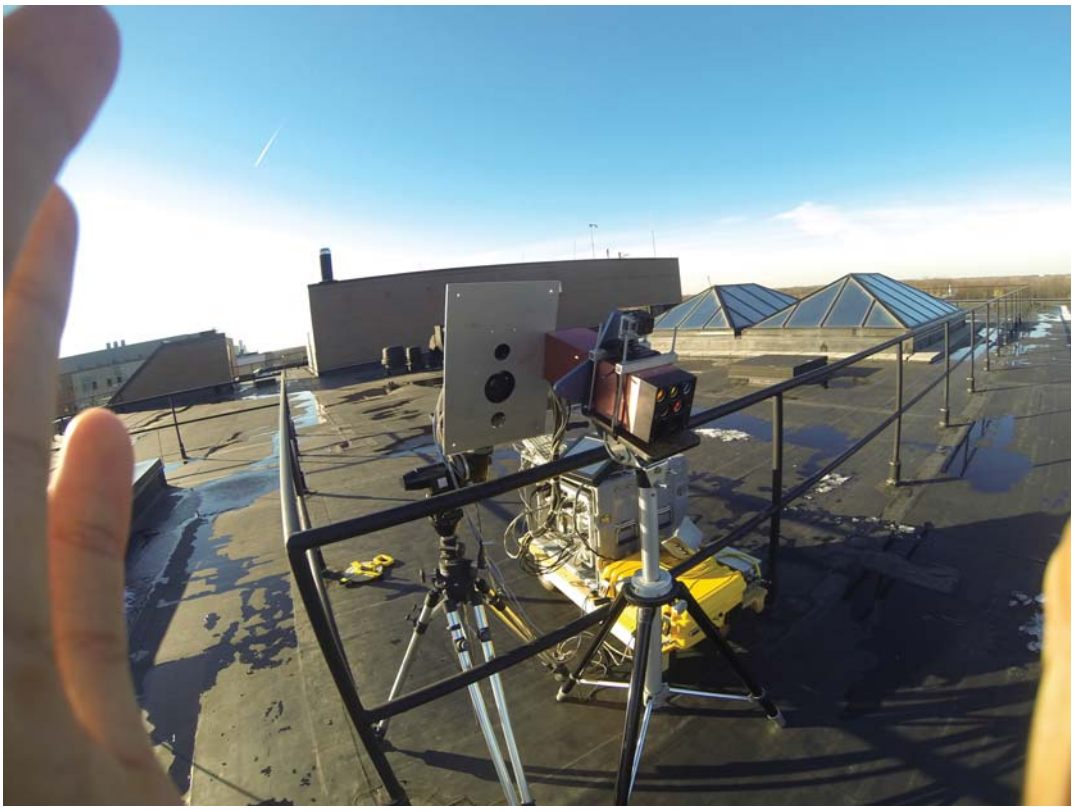


FIGURE 4.13: Experimental setup image 1

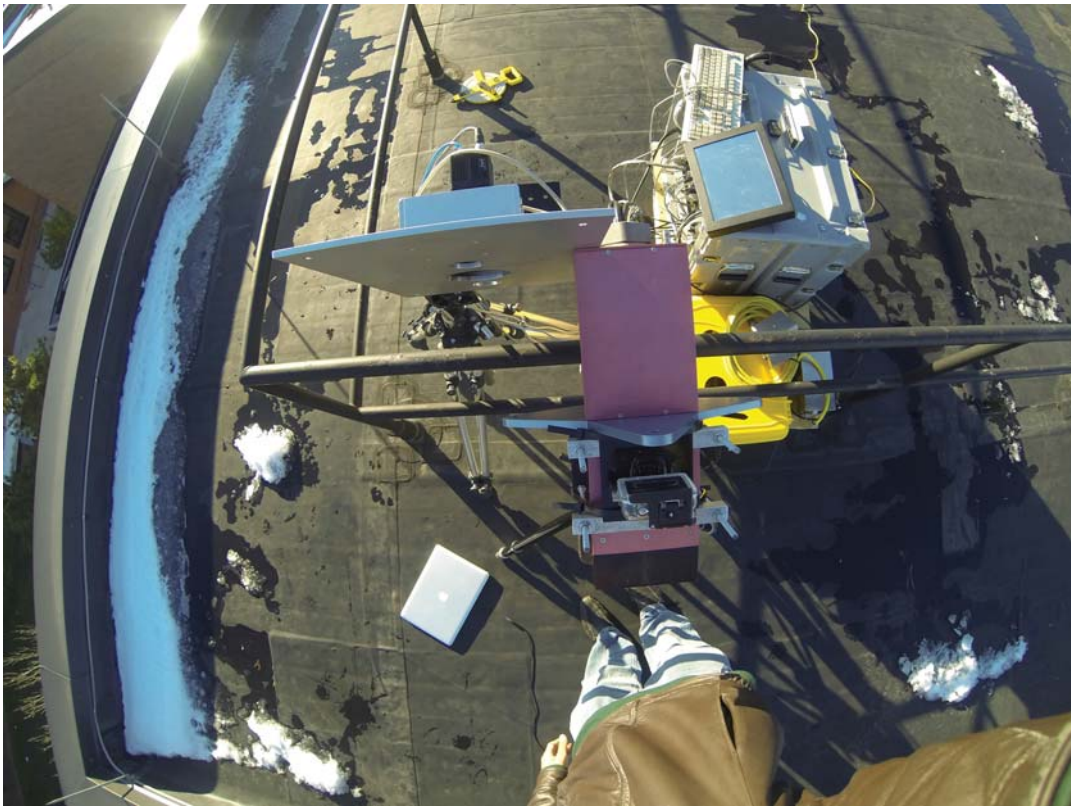


FIGURE 4.14: Experimental setup image 6

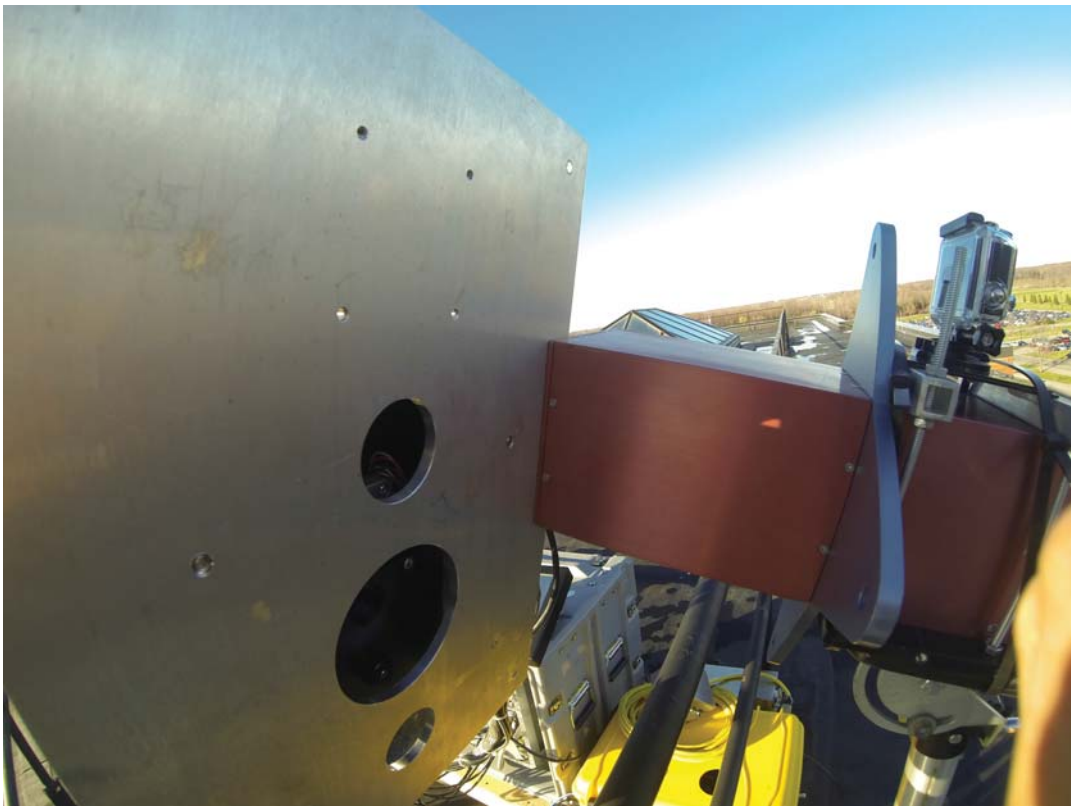


FIGURE 4.15: Experimental setup image 7

FIGURE 4.16: Experimental setup image 9

FIGURE 4.17: Experimental setup image 10

### 4.3.3 Fiducials

Fiducials are in-scene objects used to create known tie points, or Ground Control Points (GCPs), within a scene for use in registration. At the outset of this experiment, it was unknown if software-based techniques would be able to perform a proper registration on the data due to the oblique views and possible perspective differences of the imagers. As such, a series of fiducials and natural GCPs were selected ahead of time to ensure there existed adequate means to register the data.

Before taking any measurements, a series of calibration tests were performed on the imaging systems as placed within the scene. One of the purposes was to ensure the FOVs overlap and determine locations for the in-scene fiducials. The following figures depict how the sensors with smaller FOVs would fit in the scene of the sensors with larger FOVs. Since the panchromatic and spectral sensor FOVs were essentially the same, the panchromatic was used to represent those six imagers. Figure 4.18 depicts how the MAPPS FOV would look within the panchromatic sensor. Figure 4.19 depicts the panchromatic FOV within the LWIR imager. Figure 4.20, depicts the LWIR FOV

within the GoPro sensor. Finally, for registration purposes, a series of fiducials were concentrated within the overlapping FOVs depicted in Figure 4.21. This was very limiting due to the tight FOV of MAPPS. Thus, additional fiducials were placed throughout the central portion of the walkway which can be seen by the other sensors.

FIGURE 4.18: MAPPS FOV as seen through panchromatic imager

From the common overlap image, a series of locations were identified to be used as spatial registration points. As can be seen in the scene there exist few natural registration points. The corners of the walkways, the fire hydrant, the light poll, and sign are all circled in green indicating such points. The yellow circles indicate positions identified as needed additional fiduciary points for registration. In order to reduce the tripping hazard to participants but maintain the necessary number of GCPs, some of the points were created by placing boards over walkway edges. Figure 4.22 depicts all the GCPs used in this experiment. Due to the multimodal nature of the imaging equipment, a more stringent examination of the GCPs was done to ensure it can be seen from each of the sensors. The next two sections describe the visible and LWIR fiducials used within the scene.

FIGURE 4.19: Panchromatic FOV as seen through LWIR imager

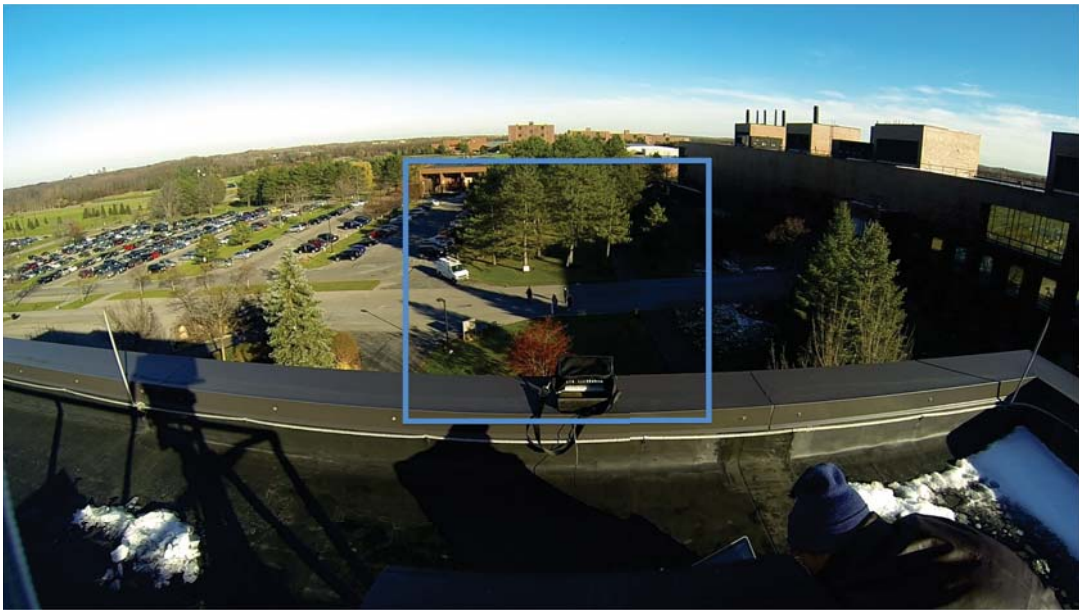


FIGURE 4.20: LWIR FOV as seen through GoPro

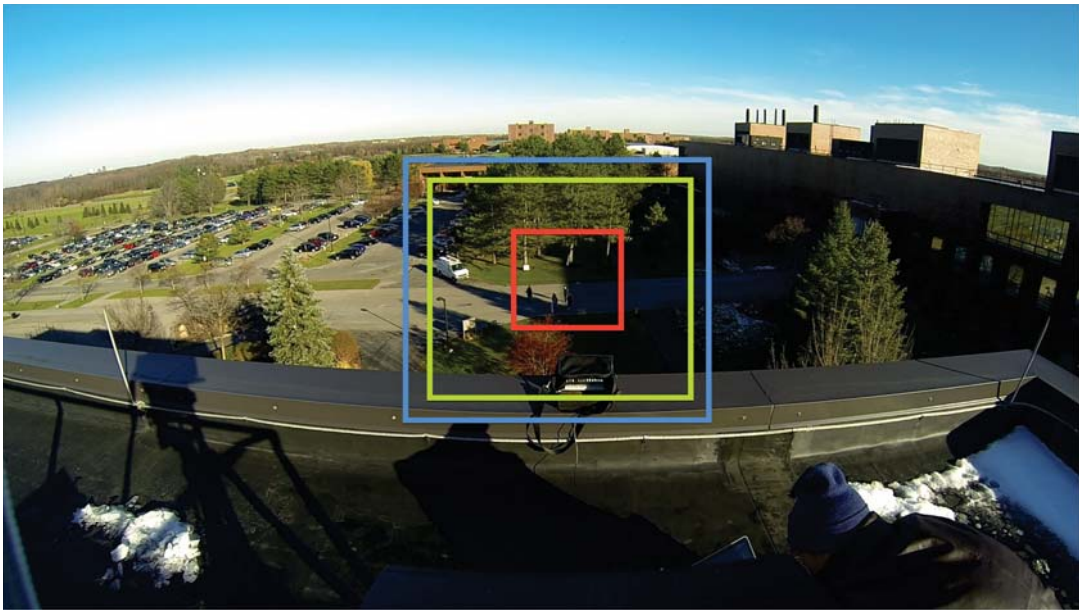


FIGURE 4.21: Platform FOV Overlap.
Blue=LWIR FOV; Green=Panchromatic FOV; and Red=MAPPS FOV

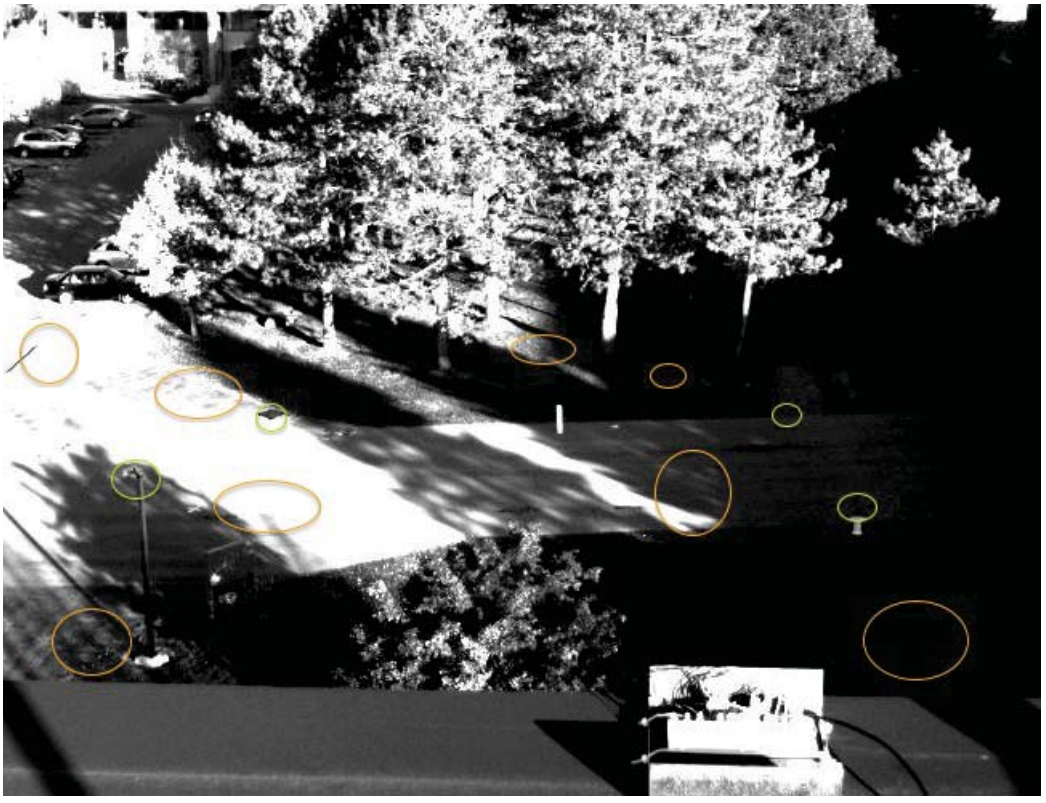


FIGURE 4.22: Ground Control Points

**Visible Spectrum Fiducials** Nearly anything that can be perceived by the Human Visual System (HVS) is useful as a visible spectrum GCP. Figure 4.23 depicts one of the fiducials within the scene. The remaining fiducials can be seen in Appendix D.

FIGURE 4.23: Fiducial E

**LWIR Fiducials** To ensure the LWIR camera can perceive the same fiducials as the visible imagers, each object needed to have distinct emissive and reflective properties when compared to the surrounding area. This was accomplished by wrapping select in-scene objects in aluminum foil and aluminum foil tape. Figure 4.23 and Appendix D depict these specific objects. The aluminum foil was selected due to its emissive properties. According to the ASHRAE handbook [82], the shiny side of aluminum foil has an emissivity of 0.05, which drastically differs with the emissivity of green grass at 0.975, water at 0.95 [83], and asphalt at 0.93 [84]. This difference in thermal emissivity provides a distinct contrast which can be used to create in scene fiducials for the LWIR imager. As a note, the Handbook of Package Engineering [85], stated that aluminum foil has a reflectivity of 95%. Thus it will still be seen in the visible regime.

**Fiducials Specifications** Table 4.10 depicts the dimensions of the fiducials and their equivalent pixel count as seen by the panchromatic and LWIR imagers. Since LWIR has the highest GSD, if a particular fiducial can be seen by this imager than it can be seen by all the imagers. The panchromatic pixel equivalents are included in Table 4.10 for comparison against the LWIR pixel equivalents.

TABLE 4.10: Dimensions of In-Scene Fiducials

| Fiducial Letter | Dimensions (cm±0.1$cm$) | | Panchromatic (Pix) | | LWIR (Pix) | |
|---|---|---|---|---|---|---|
| | Length | Height | Length | Height | Length | Height |
| A | 93.5 | 15.6 | 47.2 | 7.88 | 7.48 | 1.23 |
| B | 91.3 | 40.6 | 46.1 | 20.5 | 7.30 | 3.25 |
| C1 | 243.5 | 25.2 | 123 | 12.7 | 19.5 | 2.01 |
| C2 | 243.5 | 25.2 | 123 | 12.7 | 19.5 | 2.01 |
| C | 172.2 | 172.2 | 87.0 | 87.0 | 13.8 | 13.8 |
| D | 92.7 | 28.5 | 48.8 | 14.4 | 7.42 | 2.28 |
| E | 142.2 | 25.5 | 71.8 | 12.9 | 11.4 | 2.04 |
| F | 243.5 | 13.1 | 123 | 6.61 | 19.5 | 1.05 |
| G | 69 | 62.5 | 34.8 | 31.6 | 5.52 | 5.00 |
| H | diameter = 48.3 | | diameter = 24.4 | | diameter = 3.64 | |
| I | 121.8 | 61 | 61.5 | 30.8 | 9.74 | 4.88 |

### 4.3.4 Synchronizing Equipment Timing

Considering that each sensor suite had its own internal timing sequence, an external source was used to ensure proper syncing across the imagers. A series of LEDs actuating in a sequence matched to the fastest frame rate system was chosen to accomplish this task. Section 5.4.2 will discuss the specifics behind the timing of the LEDs and each of the sensors.

### 4.3.5 Meteorological Conditions

This portion of the experiment was accomplished on November $4^{th}$, 2013 at 10:30am EST. The conditions were clear while measurements were taken; depicted in Figures 4.24 and 4.25. During this time of day, at this time of the year, the sun's nadir is approximately 17 degrees south of the equator and 75 degrees west of the Prime Meridian. This places it low in the Rochester sky with its orientation behind the sensor. The temperature was 40 degrees Fahrenheit at the time of the collection.

FIGURE 4.24: Horizon Experiment Sky

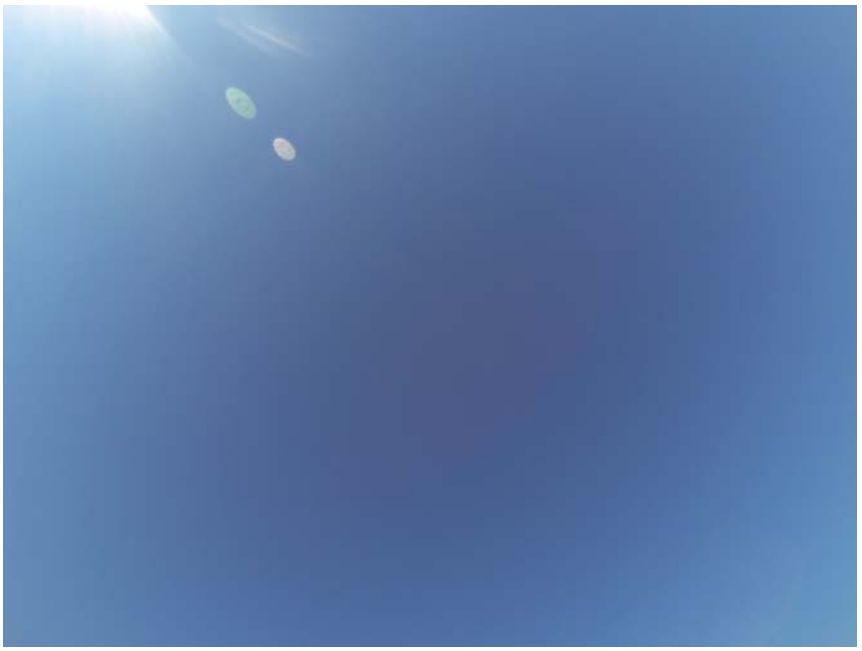


FIGURE 4.25: Overhead Experiment Sky

## 4.4 Scenario and Participants

Participants were asked to complete a series of tasks representative of activities of interest. Figure 4.26 depicts one set of instructions given to participants in the experiment. The non-explicit instructions were that individuals were to act as normal as possible

when conducting their tasks. That means walking, biking, driving, and interacting in a method that was consistent with the execution of these tasks in their everyday lives. The idea was to develop a scene that would be as realistic as possible while maintaining some measure of control of their actions. Figure 4.26 depicts the directions given to one of the participants. The remaining instructions are located in Appendix E. This section will discuss the actual events that occurred while Section 4.4.1 will describe the reason for including specific activities in this research.

A total of nine data sets were collected over a four day period. This included 278 moving people and cars with only 20 being given explicit instructions. The total execution time of the datasets was eighteen minutes and seventeen seconds; this translated into 2hr, 44min, and 55s worth of motion imagery across the nine imagers. The first data collection was the largest and included all the activities described in this research. Subsequent collections were used to collect addition data on specific AoIs. This section will discuss the conditions for the first collection and Section 4.5 will present the limited scope addressed within this research. There were a total of 15 participants with 13 being given explicit instructions and two being asked to walk around as they saw fit.

### 4.4.1 Activities

As mentioned earlier, the specific activities within a scene place minimum requirements on the imager capturing the data. At the onset of this experiment it was decided that only those activities capable of being perceived by a human reviewing the GoPro imagery would be included. Sample video data was taken of several activities and the imagery reviewed. Those activities that were recognized by a human eye were included in this experiment.

Activities were chosen to cover a wide range of spatial and temporal extents normally seen within an urban environment. Some of these activities required the use of objects with unique spectral and polarimetric properties. Table 4.11 lists the activities and characteristics that make each unique. A characteristic is defined as some unique quality that can be used to determine if the activity has occurred. For example, the mount and dismount activity is a relatively quick event and thus is said to have no appreciable temporal characteristic. The activity is comprised of a vehicle and person in close spatial

8. Begin in middle of large walkway by parking lot. Walk down the path with subject next to you. A little after crossing the gravel pathway, turn right and walk onto the bottom of the field in front of Carlson to meet up with three other subjects. Once larger group has begun game, move together to join them.

   **Begins with subject 9

FIGURE 4.26: Tasking Directions

proximity; this unique set of conditions gives it a spatial characteristic. Spectrally, each will have a different signature, but in a polarimetric context only the vehicle will have a signature. Therefore, there is a spectral characteristic to the activity, but no polarimetric characteristic. Lastly, both will have a thermal signature which can be used to determined if the activity occurs.

This list of variations was developed with respect to the contexts under which these activities were being performed. Changing the temporal scale in which these activities were executed would alter the nature of the expected variations. This research is intuitively operating under human time-scales. Thus, hours, days, and weeks are long periods of time, whereas seconds and minutes are short periods of time.

Temporal variation occurs in any type of ongoing activity; this would include groups loitering for appreciable amounts of time and sporting events. In this context this does not include the quick nature of the object exchange activity. Spatial variations tend to occur in an activity that covers a large spatial extent; these included large area sporting events, and object exchanges where the object travels across a large portion of the scene. Spectral variations are those that would provide unique changes in spectral signatures; this includes object exchanges.

For spatial variations "people", "bicycles", and "cars" were executing specific activities throughout the collection. The mount and dismount activity represented a spatially large vehicle interacting with a spatially small individual. This range of spatial extents can be used to develop a notional spatial tradespace for capturing activities occurring within an urban environment. Furthermore, by having people mount and dismount vehicles, additional research can be done on identifying activities where varying spatial extents interact with one another.

For temporal variations, people were asked to walk, run, and bicycle throughout the scene; representing an increase in speed with each successive activity. A sporting event was also included to capture short duration, fast pace actions that are indicative of larger activities. This range of temporal extents can be used to develop a notional temporal tradespace for capturing activities within an urban environment.

People were also asked to interact with one another in a specific fashion to demonstrate specific AoIs. For instance, several participants were asked to stand together in a group and chat amongst themselves. Some of the participants were asked to leave and execute another portion of this scenario. Other participants external to the group were asked to join the group at some predefined point within the experiment. These activities are indicative of people loitering with members of the group coming and going. This loitering activity can be used to build relationships amongst the group members and further analyzed to define their interactions within the larger context of the scene [73].

Objects were included to represent several activities. The simulated briefcase and duffel bag were utilized in exchange situations where one person began the scenario with the object and another ended the scenario with the object. The difference between the two is in how they were exchanged. The simulated briefcase was directly passed from one

individual to another, while the duffel bag was dropped at one point in the sequence and picked up at a later time.

The PVC pipe was included for use in a simulated RPG scenario. RPGs are round objects which are known for having a strong polarimetric signature [86]. While the PVC pipe does not share the exact dimensions of polarimetric characteristics as an actual RPG, it was deemed comparable enough for use in this research. The actual situation was to occur in a vegetated part of the scene devoid of highly polarized objects. Also, although it is well known that vehicles produce strong polarimetric signatures, the narrow FOV of MAPPS will prohibit their being captured by a polarimetric sensor.

### 4.4.2 Participant Objects

Some of the participants were asked to utilize specific objects while moving throughout the scene. Table 4.12 provides a brief description of the object and its purpose in this research. Each of these items was chosen to maximize its ability to be detected throughout the scenario. Colors such aks bright orange, red, and white were used to contrast the typical colors appearing throughout these collections: brown, green, blue, etc.

#### 4.4.2.1 Simulated Briefcase

The simulated briefcase was used in an object exchange scenario in this experiment. To execute this, one of the participants carried the item in their hand facing the imagers and began walking in the scene. The item was placed in the hand facing the imaging equipment to prevent an occluded sequence. That participant then passed this object off to another participant and continued walking in the scene. The second participant placed the object into the hand facing the imagers and continue walking throughout the scene. Figure 4.27 depicts the front of the simulated object; note that the back is identical.

TABLE 4.11: Activities in the Experiment

| Activity | Purpose | Activity Chacteristics | | | | |
|---|---|---|---|---|---|---|
| | | Temporal | Spatial | Spectral | Polarimetric | Thermal |
| Mount/Dismount Vehicle | Vehicle-person interaction | No | Yes | Yes | No | Yes |
| Merge/Separate | Person-person interaction | Yes | Yes | Yes | No | Yes |
| Group Loitering | Group-based interactions | Yes | Yes | Yes | No | Yes |
| Object Exchange | Small object transition | No | Yes | Yes | No | No |
| Bag Drop | Large object transition | No | Yes | Yes | No | No |
| Simulated RPG | VNIIR object simulation | No | No | No | Yes | No |
| Group Sport | High spatial/temporal changes | Yes | Yes | No | No | Yes |


FIGURE 4.27: Simulated briefcase

#### 4.4.2.2 PVC Pipe

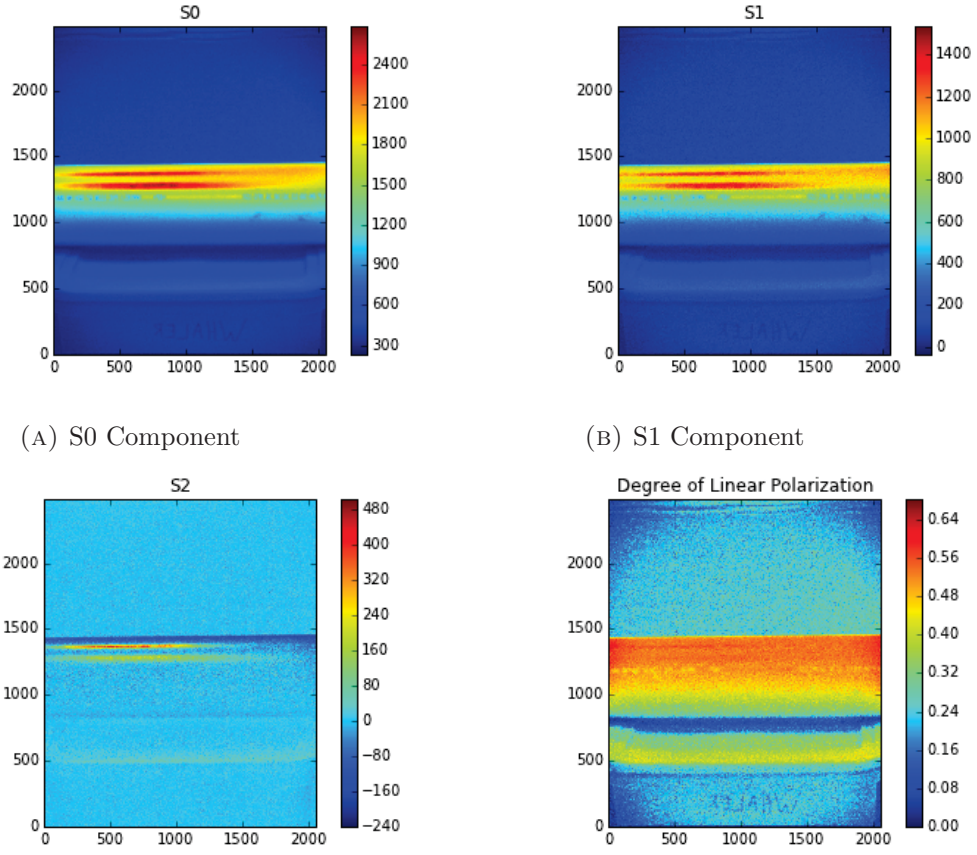
A Polyvinyl Chloride (PVC) pipe was included for use in a simulated Rocket Propelled Grenade (RPG) launch activity. This activity was chosen for inclusion in this experiment because the NIIRS and VNIIRS quality metrics include a metric for identifying an RPG launch. This common activity will allow for future comparison between the aforementioned metrics and the performance assessment methodology described in this research. The person holding the PVC pipe was instructed to stop at a central portion in the scene and lift the pipe onto their shoulder, thus simulating launch preparations. They kept the pipe on their shoulder in a skyward direction while slightly moving the object around as if to aim at a target. A short time later, the participant removed the pipe from their shoulder and resumed walking across the scene. Figure 4.28 depicts the side and front views of the PVC pipe.

**Laboratory Measurements** In order to determine if the object of interest contained the necessary polarimetric signature, an in-lab analysis was performed. Since the exact sun-target-sensor geometry could not be determined beforehand, the PVC pipe was placed in the center of a laboratory setting where illumination emanated from a series of extended sources on the ceiling. Due to the small nature of the room, it is expected that this angle of illumination was less than 45 degrees off nadir, in a 360 azimuth. Figures 4.29a, 4.29b, 4.29c, and 4.29d depict the S0, S1, S2, and DoLP results respectively.

(A) PVC pipe lengthwise (B) PVC pipe top view

FIGURE 4.28: PVC pipe imagery

Note, due to the object's stationary nature in a controlled environment, there is no need to register the data. In a scene with the object moving, the four frames would need to be registered before creating the DoLP.




(A) S0 Component (B) S1 Component

(C) S2 Component (D) Degree of Linear Polarization

FIGURE 4.29: Polarimetric Lab Results of Object

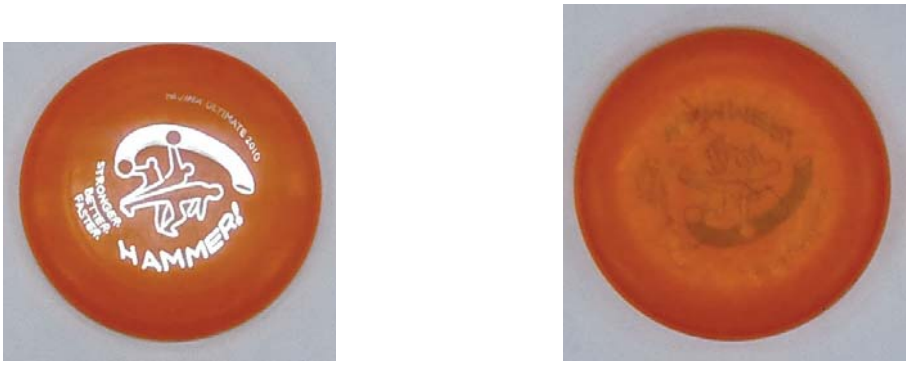
#### 4.4.2.3 Duffel Bag

The duffel bag was used in a bag drop scenario in a similar manner to that of the simulated briefcase. A participant held this item in their hand and walked toward the middle of the scene. At some point the participant set the bag down and continued walking. Later, another participant walked up to the bag and picked it up. They then continued walking through the scene. Figure 4.30 depicts the front of the duffel bag; note that the back is identical.

FIGURE 4.30: Duffel Bag

#### 4.4.2.4 Frisbee

The Frisbee item was used to include a fast paced group sporting event in the dataset. Participants were asked to congregate in a grassy area and throw the Frisbee to one another as they saw fit. Figure 4.31 depicts the front and back of the Frisbee.




(A) Frisbee front (B) Frisbee back

FIGURE 4.31: Frisbee imagery

TABLE 4.12: Objects in Experiment

| Object | Purpose | Dimensions (cm$\pm$0.1$cm$) | | |
|---|---|---|---|---|
| | | Height | Length | Width |
| Simulated Briefcase | Object Handoff | 26.0 | 38.5 | 7.7 |
| PVC Pipe | Simulated RPG | 16.8 | 85 | 16.8 |
| Duffel Bag | Bag Drop | 26.7 | 53.3 | 26.7 |
| Frisbee | Group Sport | 3.04 | diameter = 27.3 | |

## 4.5 Research Scope

This subset of data was collected on November $14^{th}$, 2013 at 4:00pm EST. The sun's nadir was approximately 17 degrees south of the equator and 135 degrees west of Prime Meridian. Figures 4.32, 4.33, and 4.34, depict the oblique, top, and side views of the scene respectively.

Figures 4.35, 4.36, and 4.37 depict the setup of the equipment in the experiment. Take note of the sun at an angle directly behind and to the left of the sensor suite. The temperature was 50 degrees Fahrenheit at the time of the collection with clear skies above, as seen in Figure 4.35. Table 4.13 lists the activities in this portion of the data collection. Of those included in the larger experiment, only the object exchange and simulated RPG were included.

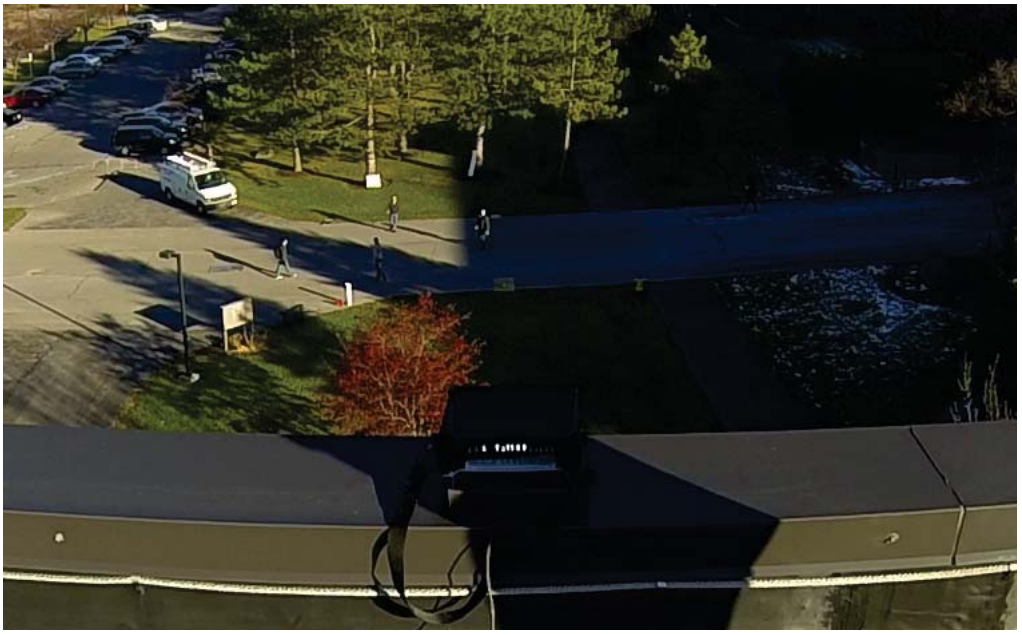


FIGURE 4.32: Oblique view of scene

FIGURE 4.33: Top view of scene from Google Maps [13]

TABLE 4.13: Activities Specific to the Scope of this Research

| Activity | Purpose | Variations | | | |
|---|---|---|---|---|---|
| | | Temporal | Spatial | Spectral | Polar |
| Object Handoff | Small object transition | No | Yes | Yes | No |
| Simluate RPG | VNIIRS object simulation | No | No | No | Yes |

FIGURE 4.34: Side view of scene

FIGURE 4.35: Back view of sensor setup

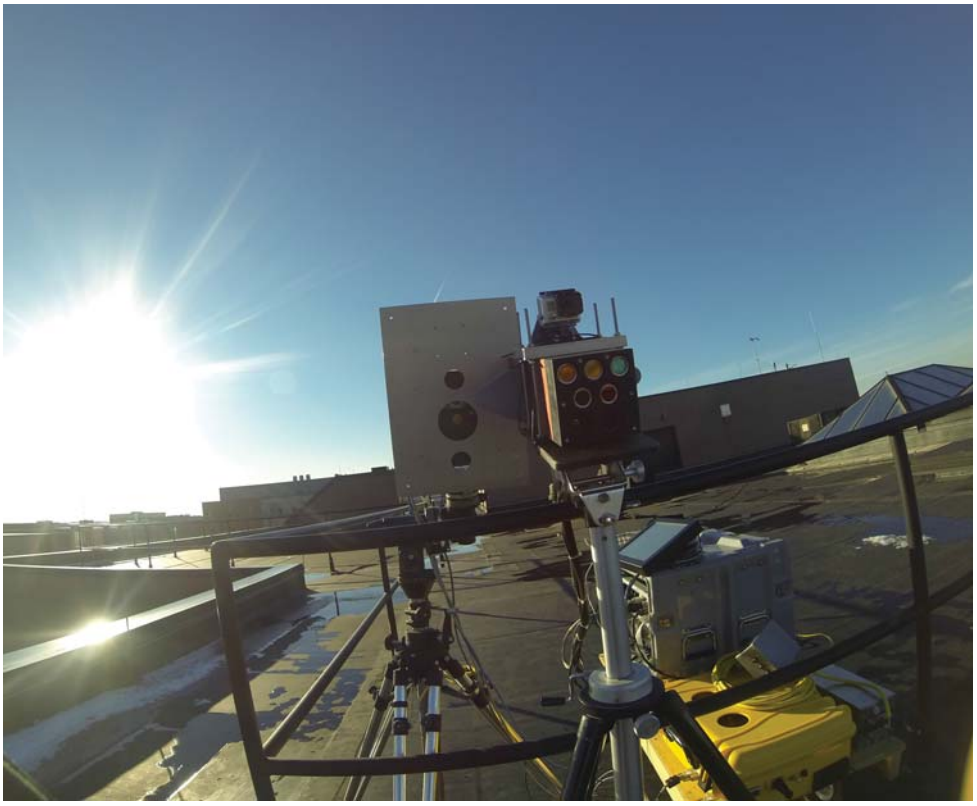


FIGURE 4.36: Front view of sensor setup

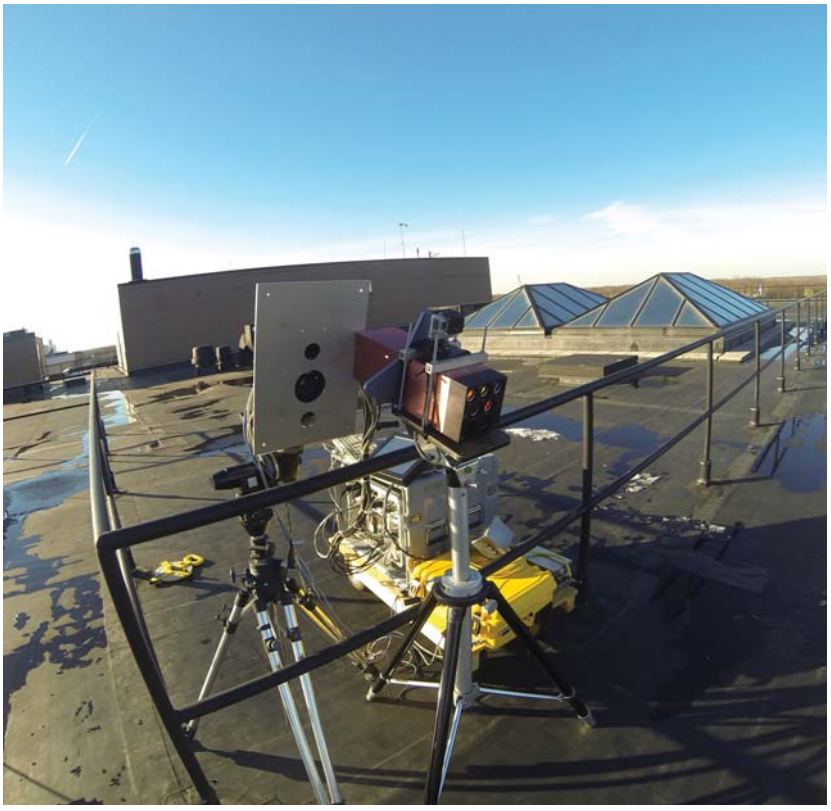


FIGURE 4.37: Diagonal view of sensor setup

# Chapter 5

# Methodologies

## 5.1 Flow of Data Processing

Figure 5.1 depicts the top level flow methodology of this activity recognition research. This process begins with the raw data collected from the imaging equipment. Once obtained, the cameras need to be properly calibrated to remove distortions and aberrations within the imagery due to lens effects. Following calibration, the video sequence needs to be stabilized if there were environmental factors (i.e. wind, building vibration, etc) that induced motion in the imaging data. After stabilization, the images must be registered and the data fused for exploitation. In order to limit exploitation to moving people and objects within the scene, a tracking algorithm is implemented. Having these positions, it is then possible to perform activity recognition.

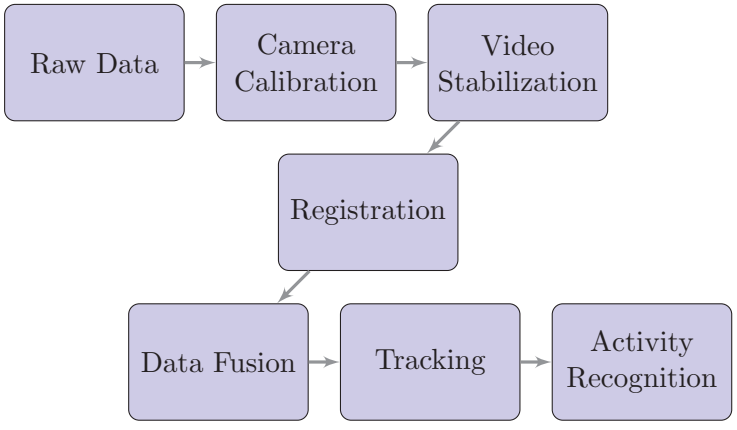



FIGURE 5.1: Processing Flow Diagram

Figure 5.2 depicts the full flow diagram of the processing specifically involved in this research. This diagram includes intermediary steps necessary to achieve the results in chapter 6.

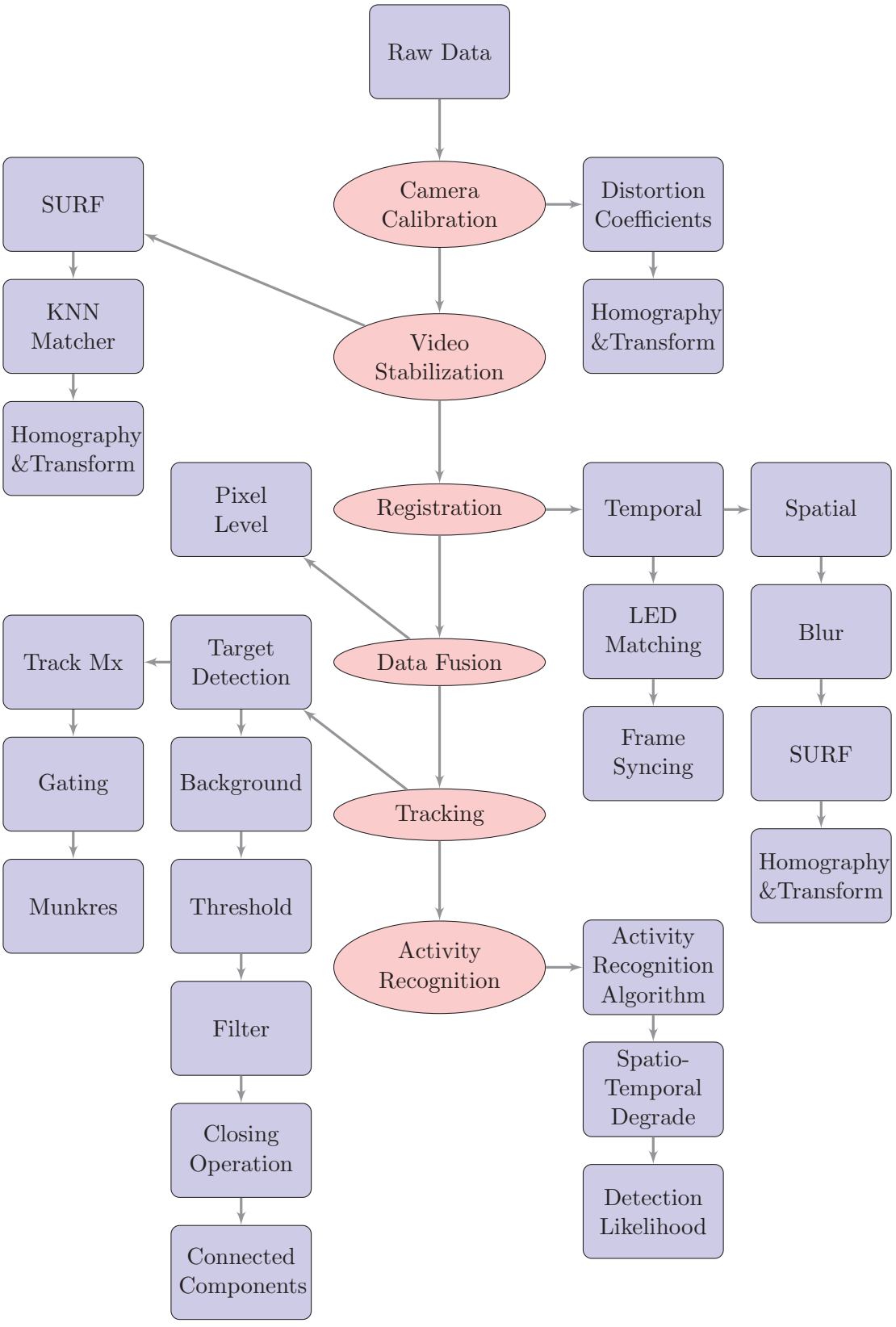



FIGURE 5.2: Processing Flow Diagram with Intermediary Steps

## 5.2 Camera Calibration

The purpose of camera calibration is to remove the lens distortion included in the imagery from the attached optical train. Upon visual inspection of the WASP-Lite and MAPPS imagers, the data were found to have a negligible amount of lens distortion relative to their intended uses. The GoPro imagery displayed large amounts of barrel distortions due to the fisheye lens. Thus, this section describes the process necessary to remove those distortions.

To do so, the GoPro was taken to the RIT Calibration Cage and a series of images were taken at various locations and orientations. Then, the Australis software was utilized to develop the calibration coefficients necessary to remove the lens distortions. Finally, the distortions were removed and a notionally calibrated image sequence was produced.

**RIT Calibration Cage** The RIT calibration cage is a three-dimensional calibration structure consisting of a series of visible and infrared LEDs. Figure 5.3 shows an image of the calibration cage as taken by the GoPro imager.

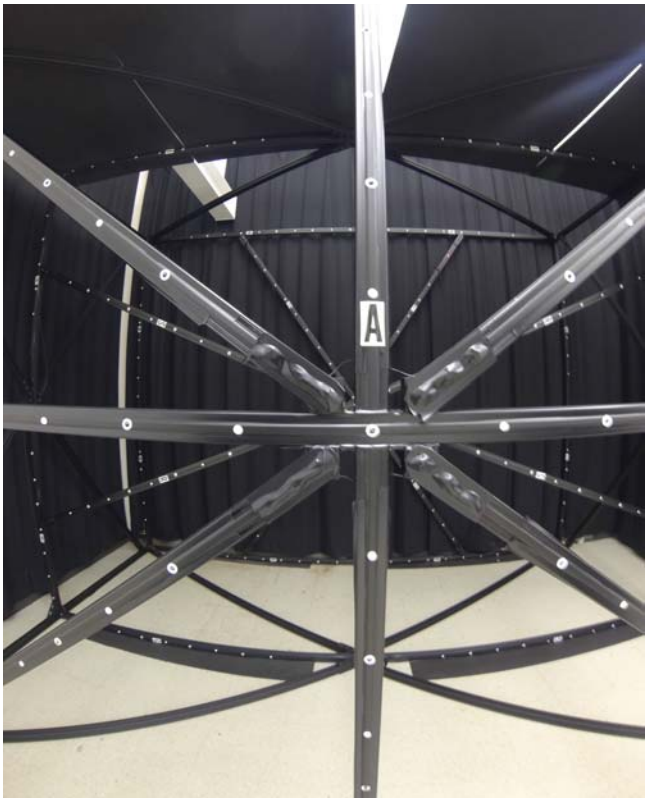



FIGURE 5.3: RIT Calibration Cage

**Australis** Australis is "A software system for automated off-line digital close range photogrammetric image measurement, orientation/triangulation and sensor calibration cage [87]." This software was used to take a series of images of the RIT calibration from several perspectives and orientations in order to determine the calibration coefficients of a system. These coefficients remove distortions in the radial and tangential directions of the imagery taken from the imager in question. Figures 5.4 and 5.5 depict the three dimensional digital version of the RIT calibration cage in straight on and diagonal views.

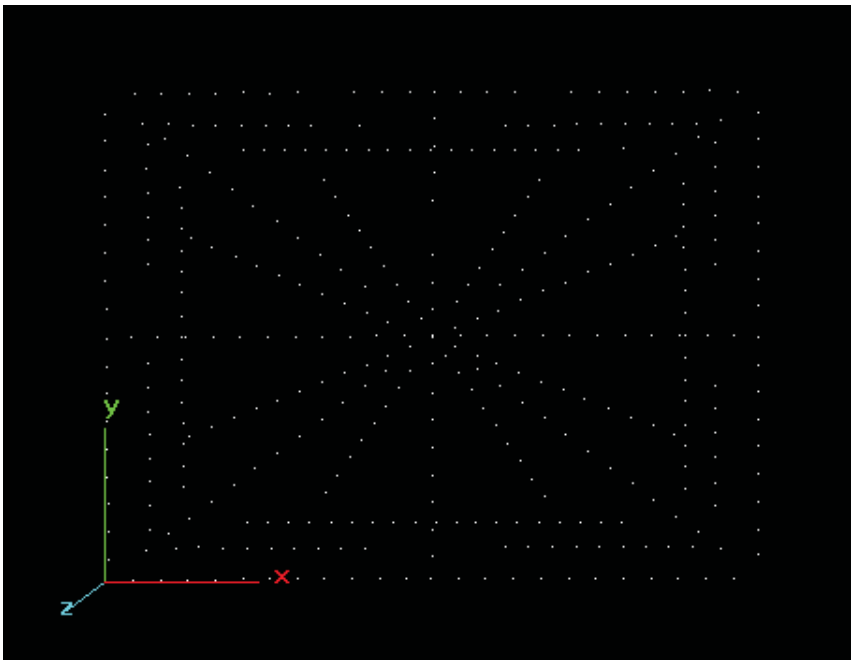



FIGURE 5.4: Digital Version of RIT Calibration Cage

Figure 5.6 depicts an output view of some of the camera position and orientations as calculated by the bundle adjustment software within the Australis framework. The output of the bundle adjustment software is a list of distortion coefficients; depicted in Figure 5.7. Table 5.1 depicts the distortion coefficients for the GoPro imager. The coefficients K1, K2, and K3 adjust for radial distortions in the image and the tangental coefficients, P1 and P2, adjust for the decentering of the alignment of the array.

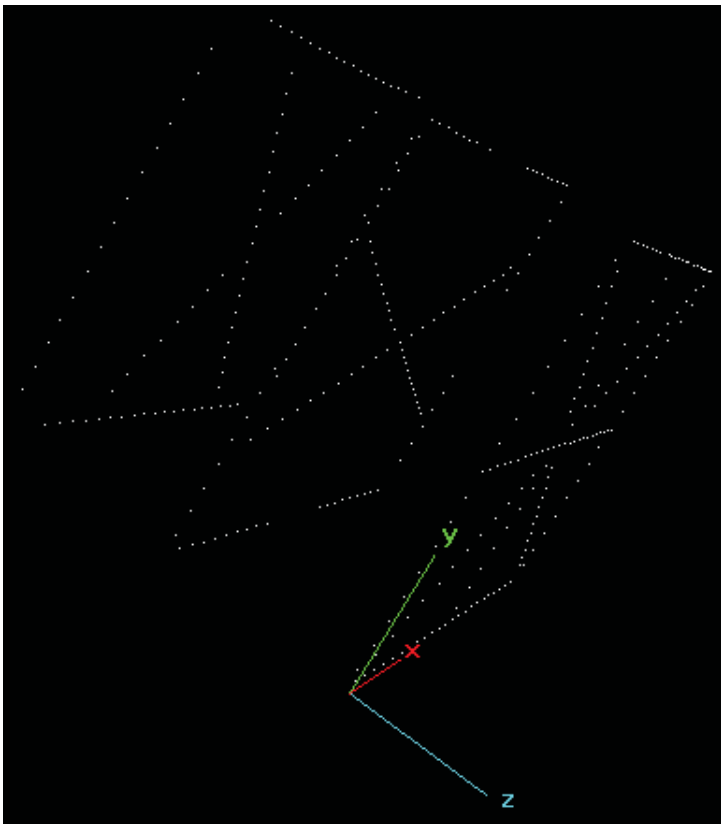



FIGURE 5.5: Rotated Digital Version RIT Calibration Cage

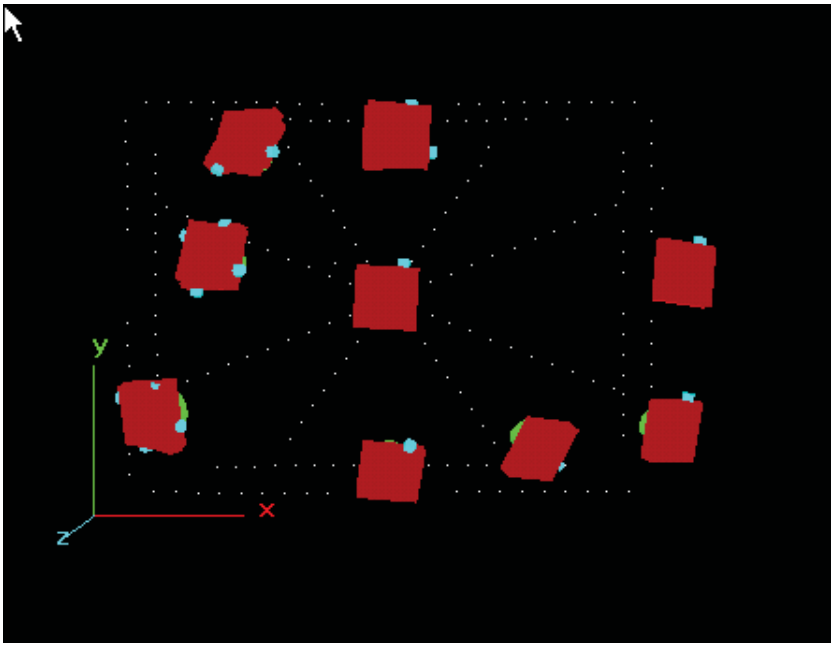



FIGURE 5.6: Camera Locations using Australis Camera Calibration

```
                        Australis Bundle Adjustment Results: Camera Parameters

                                    12 December, 2013   07:56:59

Project:  C:\Users\Public\Documents\Project 1.aus

Adjustment: Free-Network
Number of Points: 210
Number of Images: 18
RMS of Image coords:    0.89 (um)

Results for Camera 1    Go-Pro      Lens

Sensor Size        Pixel Size (mm)
  H    4000           0.002
  V    3000           0.002

  Camera    Initial      Total          Final        Initial         Final
 Variable    Value     Adjustment       Value       Std. Error     Std. Error
    C        2.7660      0.00377        2.7698       1.0e+003     4.705e-004 (mm)
   XP        0.0271     -0.00015        0.0270       1.0e+003     2.265e-004 (mm)
   YP        0.1335      0.00027        0.1338       1.0e+003     2.559e-004 (mm)
   K1  5.07200e-002   3.428e-004  5.10628e-002       1.0e+003     9.419e-005
   K2  1.08470e-004  -1.355e-004 -2.70112e-005       1.0e+003     2.070e-005
   K3  1.52087e-004   1.040e-005  1.62488e-004       1.0e+003     1.481e-006
   P1  1.07631e-004   8.072e-006  1.15703e-004       1.0e+003     1.841e-005
   P2 -1.15467e-004   6.887e-006 -1.08579e-004       1.0e+003     1.916e-005
   B1 -2.37071e-004   8.997e-005 -1.47096e-004       1.0e+003     4.204e-005
   B2 -2.18761e-004   5.486e-005 -1.63903e-004       1.0e+003     4.071e-005
```

FIGURE 5.7: Output of Australis Bundle Adjustment

TABLE 5.1: Distortion Coefficients

| Camera Variable | Initial Value | Final Value |
|---|---|---|
| Focal Length | 2.7660 | 2.7698 |
| K1 | 0.05072 | 0.05106 |
| K2 | 1.0847E−4 | −2.7011E−5 |
| K3 | 1.5209E−4 | 1.6249E−4 |
| P1 | 1.1076E−4 | 1.1570E−4 |
| P2 | 1.1547E−4 | −1.0858E−4 |

**Sensor Calibration** In the sensor calibration step, the above coefficients are applied to the distorted imagery originally produced by the imager. Figure 5.8 depicts a example "before and after" image of the correction applied to a fisheye lens using RITs calibration cage; this image was created by Brent Bartlett [14]. In an effort to achieve the same goal, we took the first frame in the image sequence and applied the technique above. Figure 5.9 depicts the before calibration image of the data. Notice the high degree of bow in the building edge.

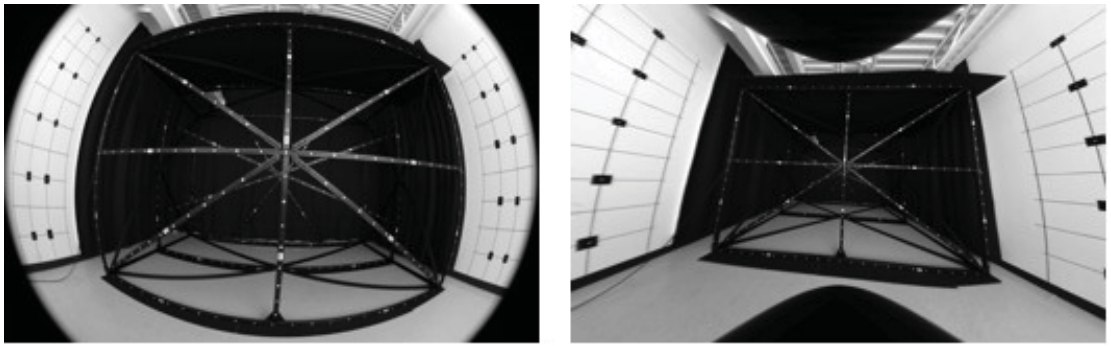


FIGURE 5.8: Fisheye lens calibration before and after [14]

FIGURE 5.9: Before GoPro Camera Calibration

After attaining the calibration coefficients and applying them to the imagery, it was noticed that it did not have the intended effect. Figure 5.10 shows this result. The black around the edge is indicative of an image with some level of distortion removed.

Further investigation revealed that the GoPro video streams and still image data collection modalities use different pixel bin sizes to capture the scene. This binning of pixels

FIGURE 5.10: Original Distortion Correction

causes irregularities in the application of the derived coefficients. Through empirical trials, a visually appeasing adjustment was achieved. Figures 5.9 and 5.11 depict the before and after of the same scene.

FIGURE 5.11: After GoPro Camera Calibration

As can be seen, while it does straighten the curved edge, there are oddities about the edges of the image, and we cannot be completely sure that similar oddities are not present in the central portion of the image. As such, the original uncorrected imagery was selected and only the central portion of the images were used. A closer look at the

center of the image reveals what appear to be minor radial distortions due to the fisheye nature of the lens. Figure 5.12 shows this effect.

FIGURE 5.12: Full Scene Center Closeup

While a fully undistorted image would be ideal for use in this experiment, the individuals and objects used in this research appear to be of sufficient size to alleviate the need to have the central portion of the image completely undistorted.

## 5.3 Video Stabilization

Since the data were taken outside, the sensors were susceptible to the same atmospheric conditions as the objects in the scene. This included both sustained wind and short gusts. In order to correct for induced motion in the sensor, a video stabilization process was implemented. Manually reviewing the first frame provided an indication that the sensor was stable at this collection, therefore it was used as the base for future frame stabilization. Figure 5.13 depicts this processing flow. This process occurred on each frame. First a SURF feature detector was used to find common features within the sequence. Then a Nearest Neighbor algorithm was implemented to find three neighbors for each SURF feature in the base image. RANdom Sample Concensus (RANSAC) was

used to develop the homography. Finally, the homography was used to implement a perspective transform, removing atmospherically induced motion.

Although this technique was used to stabilize a sequence of images in a video stream it simply represents a method of registering one image to another. Thus, it will be referenced in Section 5.4.4 when talking about multimodal spatial registration.

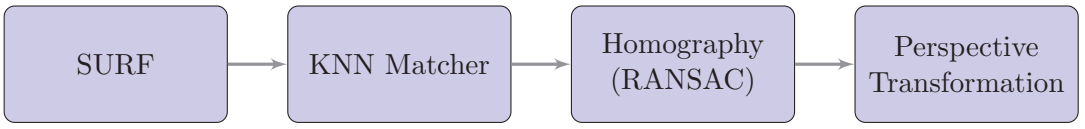



FIGURE 5.13: Image Stabilization Flow Diagram

The motion imagery collection, for this experiment, was conducted over multiple days allowing for various environmental conditions. Two of the four collections required video stabilization due to noticeable oscillations in the videos. This dataset was not one of them, so video stabilization was not applied.

## 5.4 Registration

Registration of the data within this experiment occurs in two phases. First, the data need to be temporally registered so concurrent events can be correlated across the various imagers. Second, the data need to be spatially registered so common events within the video streams occur in the same space. Take, as an example, two cameras watching a ballet. One is set on the left side of the audience and the other on the right side. One is set to begin recording at the beginning of the ballet, while the other is set to begin recording five minutes later. After the performance, if you were to play these videos on side-by-side monitors you would notice a lag between videos and a difference in perspective. The purpose of this step is to make these two videos streams appear as though they began at the same time and were placed in the same location in the audience.

### 5.4.1 Registration Accuracies

Determining the necessary registration accuracy depends on the type of activity being imaged and the capabilities of the imager. Of the two activities being analyzed the object exchange will be used as the base for the registration accuracy calculations. This is due to the spatial and spectral extent of the activity. The requirements for the simulated RPG activity will be addressed in Sections 5.5.3 and 5.7.2.

Beginning with the physical layout of the object exchange, assume a human will be walking forward with the object held lengthwise in full view of the imager as depicted in Figure 5.14. To ensure proper spatial and temporal registration, we would like each object of interest to have a minimum of 50% overlap with itself. Spatially, this means that 50% of the object needs to occupy the same pixel coordinates each image modality. Temporally, within each matched set of frames across the video sequences, at least 50% of the object needs to occupy the same pixel coordinates. An objective would be 75% of overlap with itself. Since we are evaluating spectral signatures of particular objects, this constraint will be placed on full pixels of overlap to avoid a pixel unmixing situation. The details of the effects of this full pixel requirement will be discussed later in this section; first the variables need to be defined.

As stated in Robinson [88], the average velocity of a marching soldier is said to be

$$v_{march} = 1.5^{m}/_{s}$$

Using the hand-off object dimensions described in section 4.4.2, the object's height, length, and width are known. However, since most of the action occurs in the horizontal direction, only the length of the object will be used. The object of interest here has a primary dimension of

$$l_{obj} = 38.5cm \pm 0.1cm$$

The object exchange aspect of the research includes the use of the GoPro and WASP-Lite sensors. The frame rate of the GoPro is

FIGURE 5.14: GoPro image of human holding object of interest

$$fr_{GoPro} = 60Hz$$

while that of WASP-Lite is

$$fr_{WASP-Lite} = 8Hz$$

As part of this discussion, we will assume that the object is translated in a linear fashion between adjacent frames. The translation distance for the object is calculated as

$$
\begin{aligned}
x_{obj}(t) &= v_{march} \cdot t \\
x_{obj}(1s) &= 1.5m/s \cdot 1s \\
&= 1.5m
\end{aligned}
\tag{5.1}
$$

#### 5.4.1.1 Temporal Registration

Determining how far it translates with respect to our sensor is done by taking the inverse frame rate as the time between frames. The temporal registration of this data is done with respect to the highest frame rate imaging system. Thus the GoPro is used and the object's per image translation is calculated as

$$
\begin{aligned}
x_{obj}\left(\frac{1s}{60}\right) &= 1.5m/s \cdot {}^{1s}/_{60} \\
&= 0.025m \\
&= 2.5cm
\end{aligned}
$$

Imaging through the GoPro, the object moves 2.5cm each frame. These translations can be remapped into pixel space by using the GSD of one of the sensors. The GSD of the multispectral sensor will be used. The pixel space equivalent of the object is calculated by

$$
GSD_{spectral} = 5.00\frac{cm}{pix}
$$

$$
\begin{aligned}
l_{obj}(pix) &= \frac{l_{obj}(cm)}{GSD_{spectral}} \\
&= \frac{38.5cm}{5.00\frac{cm}{pix}} \\
&= 7.7pix
\end{aligned}
$$

Using the above constraint of avoiding mixed pixels, the length of the object fills seven pixels at a time. However, we will further reduce this by saying that it must be full pixels at all times. Since we cannot say with certainty that the leading and trailing edges will not both be partial pixels, we stipulate that only six pixels are considered in the registration requirements. Moving at the above velocities, the object's pixel-based translational velocity is

$$\begin{aligned} v_{obj}\,(^{pix}/_{frame}) &= \frac{V_{obj}(^{cm}/_{frame})}{GSD_{spectral}} \\ &= \frac{2.5\frac{cm}{frame}}{5.0\frac{cm}{pix}} \\ &= 0.5^{pix}/_{frame} \end{aligned} \tag{5.2}$$

Four temporal registration tolerances can be calculated by using a combination overlap and pixel fill requirements. These tolerances are calculated in terms of frames. Beginning with the partial pixel requirements the number of frames needed to ensure half of the object overlaps itself is calculated as

$$\begin{aligned} l_{obj-partial-pix}(pix) &= 7.7pix \\ \frac{1}{2}l_{obj-partial-pix}(pix) &= 3.85pix \\ t_{half-overlap}(frames) &= \frac{\frac{1}{2}l_{obj-partial-pix}(pix)}{v_{obj}\,(^{pix}/_{frame})} \\ &= \frac{3.85pix}{0.5^{pix}/_{frame}} \\ &= 7.7frames \end{aligned} \tag{5.3}$$

Since the data is not being interpolated between frames, the actual requirement must be rounded to a discrete frame value

$$t_{half-overlap}(frames) = 8frames$$

By considering only full pixels, a similar number of frames can be calculated as follows

$$\begin{aligned} l_{obj-whole-pix}(pix) &= 6.0pix \\ \frac{1}{2}l_{obj-whole-pix}(pix) &= 3.0pix \\ t_{half-overlap}(frames) &= \frac{\frac{1}{2}l_{obj-whole-pix}(pix)}{v_{obj}\,(^{pix}/_{frame})} \\ &= \frac{3.0pix}{0.5^{pix}/_{frame}} \\ &= 6frames \end{aligned} \tag{5.4}$$

Performing the same evaluation, an objective number of frames can be determined. By requiring a $^{3}/_{4}s$ overlap of the object, the calculations indicate that only $^{1}/_{4}$ of the object is not overlapping. This $^{1}/_{4}$ is used to calculate the number of frames by

$$\begin{aligned} l_{obj-partial-pix}(pix) &= 7.7pix \\ \frac{1}{4}l_{obj-partial-pix}(pix) &= 1.925pix \\ t_{3/4-overlap}(frames) &= \frac{\frac{1}{4}l_{obj-partial-pix}(pix)}{v_{obj}\,(^{pix}/_{frame})} \\ &= \frac{1.925pix}{0.5^{pix}/_{frame}} \\ &= 3.85frames \end{aligned}$$

Since the data is not being interpolated between frames, the actual requirement must be rounded up to a discrete frame value

$$t_{3/4-overlap}(frames) = 4frames$$

By considering only full pixels, a similar number of frames can be calculated as follows

$$
\begin{aligned}
l_{obj-whole-pix}(pix) &= 6.0pix \\
\frac{1}{4}l_{obj-whole-pix}(pix) &= 1.5pix \\
t_{3/4-overlap}(frames) &= \frac{\frac{1}{4}l_{obj-whole-pix}(pix)}{v_{obj}\,(pix/frame)} \\
&= \frac{1.5pix}{0.5^{pix}/_{frame}} \\
&= 3frames
\end{aligned}
$$

Table 5.2 consolidates the partial and full pixel calculations for both the threshold and objective temporal registration requirements in terms of frames. Table 5.3 depicts the same in units of milliseconds.

TABLE 5.2: Temporal Registration Requirements (frames)

| Requirements | Partial Pixels | Full Pixels |
|---|---|---|
| 1/2 Object Overlap | 8 frames | 6 frames |
| 3/4 Object Overlap | 4 frames | 3 frames |

TABLE 5.3: Temporal Registration Requirements (ms)

| Requirements | Partial Pixels | Full Pixels |
|---|---|---|
| 1/2 Object Overlap | 133.33ms | 100ms |
| 3/4 Object Overlap | 66.67ms | 50ms |

By relating the WASP-Lite imaging suite to the GoPro imager, it is possible to determine how many GoPro frames occur between WASP-Lite images. At a frame rate of 8Hz the inter-frame capture time can be calculated by

$$
\begin{aligned}
t_{b/t-Images-WASP-Lite}(^{s}/_{frame}) &= \frac{1s}{8frames} \\
t_{b/t-Images-WASP-Lite}(^{ms}/_{frame}) &= \frac{1000ms}{8frames} \\
&= 125^{ms}/_{frame}
\end{aligned}
$$

The same inter-frame capture time can be calculated for the GoPro in the following manner

$$\begin{aligned} t_{b/t-Images-GoPro}(^{s}/_{frame}) &= \frac{1s}{60frames} \\ t_{b/t-Images-GoPro}(^{ms}/_{frame}) &= \frac{1000ms}{60frames} \\ &= 16.67^{ms}/_{frame} \end{aligned}$$

A ratio of these two values can be used to determine how many GoPro frames occur between every WASP-Lite frame.

$$\begin{aligned} \frac{t_{b/t-Images-WASP-Lite(s/frame)}}{t_{b/t-Images-GoPro}(^{s}/_{frame})} &= \frac{125^{ms}/_{frame}}{16.67^{ms}/_{frame}} \\ &= 7.5 \\ t_{b/t-Images-WASP-Lite} &= 7.5 \cdot t_{b/t-Images-GoPro} \end{aligned}$$

Since there are no intermediate frames, this occurs every $8^{th}$ frame. In the actual syncing, it is likely that the frames will align every $8^{th}$ then $7^{th}$ then $8^{th}$ again to balance out the timing.

#### 5.4.1.2 Spatial Registration

With the temporal registration understood, we need to determine how its accuracies or inaccuracies affect the spatial registration. Figure 5.15 depicts how misregistering the data will affect the movements spatially.

MAPPS is a lower frame rate sensor operating at 6Hz. Thus expanding this logic, at frame 10 rather than frame 8, it will have a spatial registration error of 25cm. At the threshold and objective values, the spatial error caused by mis-registration can be calculated as

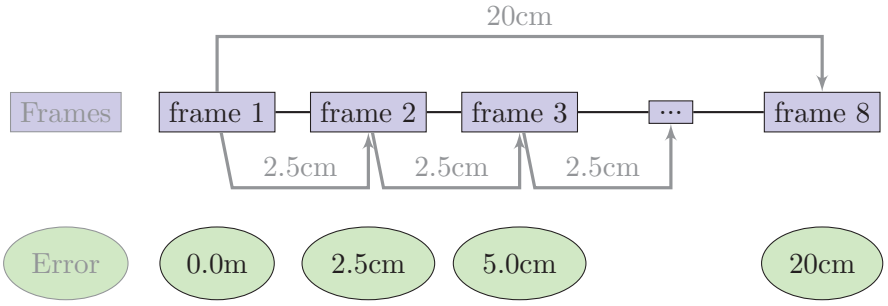



FIGURE 5.15: WASP-Lite Temporal Registration Error

$$
\begin{aligned}
x_{obj}(cm) &= v_{obj}(^{cm}/_{frame}) \cdot \#\ of\ frames \\
x_{obj-minimum}(cm) &= 2.5^{cm}/_{frame} \cdot 6 frames \\
&= 15cm
\end{aligned}
\tag{5.5}
$$

Replacing the above with an object velocity in pixels per frame, the per frame translation can be calculated as

$$
\begin{aligned}
x_{obj}(pix) &= v_{obj}(^{pix}/_{frame}) \cdot \#\ of\ frames \\
x_{obj-minimum}(pix) &= 0.5^{pix}/_{frame} \cdot 6 frames \\
&= 3pix
\end{aligned}
\tag{5.6}
$$

Since spatial registration techniques will be utilizing an interpolation method, it is more accurate to assess object locations to less than a pixel. Making that assumption, it is assumed that at least $^1/_{10}{}^{th}$ of a pixel is filled at either end of the object. This translates to a 0.5cm remainder in spatial registration.

#### 5.4.1.3 Registration Budget

The purpose of leaving this remainder can be seen in the remaining budget for the spatial registration. Essentially, even if the temporal registration can only meet the minimum requirements for alignment (6 frames), there is still a small amount of registration budget remaining for the spatial aspect to accomplish the task. Otherwise, meeting the

minimum threshold value would require a perfect spatial registration of the data. This is not likely with current techniques. Figures 5.16, 5.17, 5.18 depict the effect temporal registration has on spatial registration. The three graphs depict the same data plotted in different units.

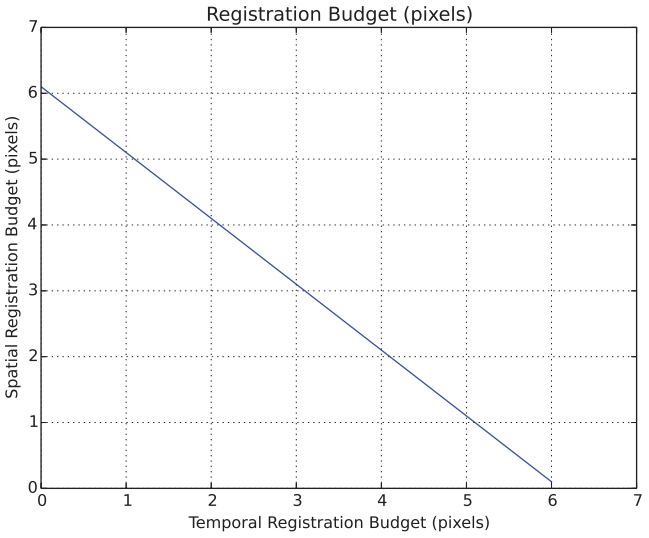



FIGURE 5.16: Registration Budget in Pixels

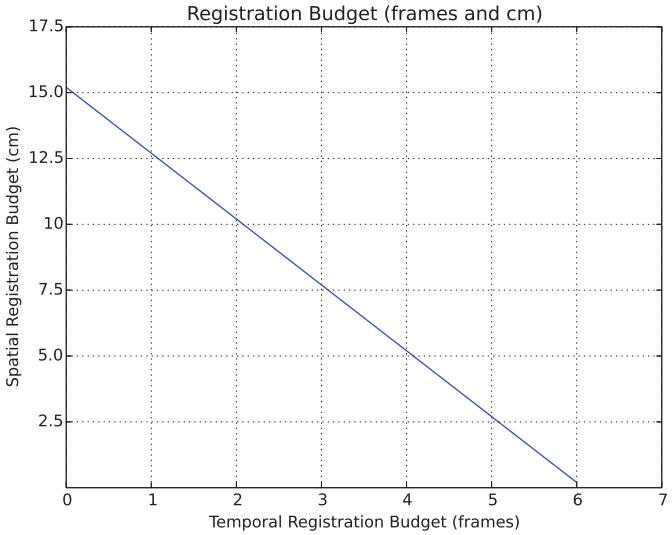



FIGURE 5.17: Registration Budget in frames and cm

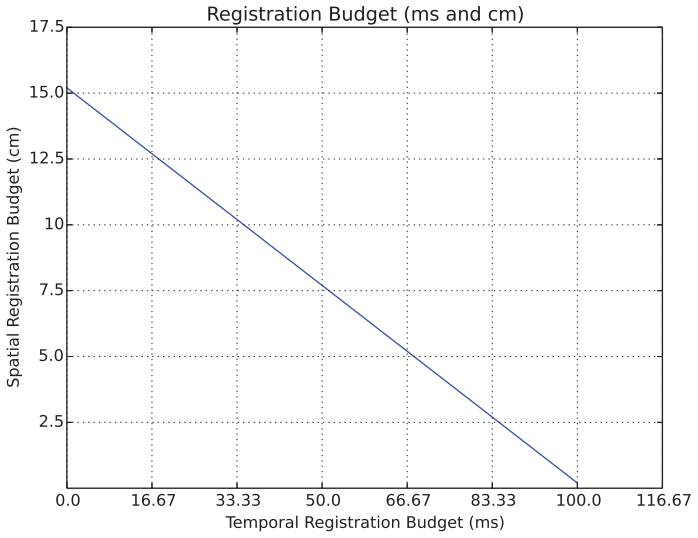



FIGURE 5.18: Registration Budget in ms and cm

### 5.4.2 Temporal Registration

Figure 5.19 depicts how this data association may look temporally. Since the frame rates are not equivalent, the data cubes will only contain data from multiple modalities when the modalities are present, i.e. every 6th frame. Due to the need to utilize multiple frames for polarization products, another layer of temporal alignment needs to occur. This will be discussed in section 5.5.3. In order to perform the temporal registration, a series of Light Emitting Diodes (LEDs) were included in the scene. Figure 5.20 depicts the setup of the LEDs.

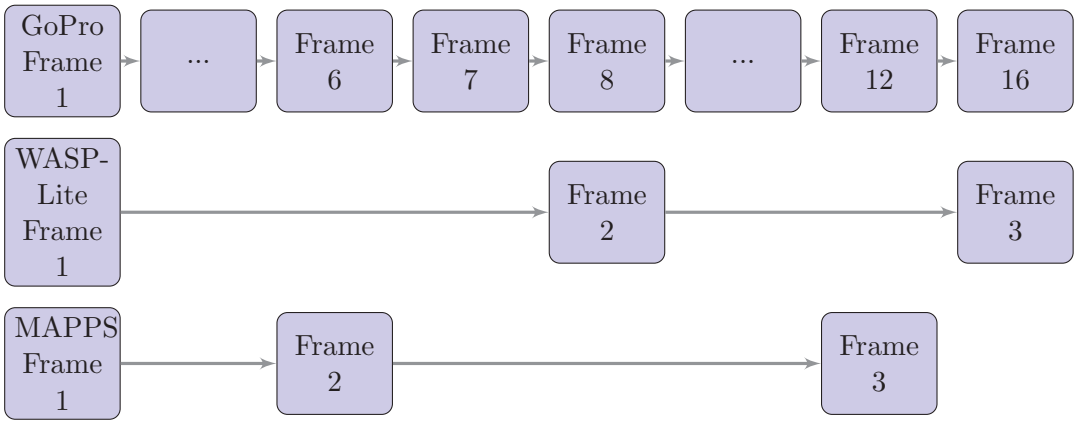



FIGURE 5.19: Temporal Data Association

#### 5.4.2.1 Light Emitting Diodes (LEDs)

FIGURE 5.20: LED Setup

The LEDs were divided into two groups; the three on the left acted as counters for the sequence performed by the eleven on the right. The group on the right begins by turning on each diode beginning from right and moving left until all diodes are engaged. Once all eleven are active, the sequence continues by turning off each diode in the same sequence they were turned on. This ensures 22 unique states. Once this sequence has been completed, the rightmost diode of the other group becomes active. The right group again goes through its sequence and upon completion, the middle diode of the left group becomes active. This continues in the same fashion, with all three turning on from right to left, then turning off from right to left. In this fashion, the three left LEDs allows for six unique combinations. Together, these groups produce 132 unique combinations.

To ensure a uniquely lit diode in each consecutive frame, the diodes must engage at a rate greater than or equal to the fastest framing system. By setting the sequencing of the diodes such that they remain on rather than rapidly turning off, the issue of Nyquist sampling is avoided. Stated differently, by purposely keeping LEDs active much longer

than the framing capability of the system, we avoid having to ensure the framing systems are sampling the diode sequence at Nyquist rates or better.

To gain the maximum temporal resolution possible, the LEDs were set to actuate (engage or disengage) at a rate equivalent to the fastest framing rate sensor in the experiment. This was set to be 1/60 of a second to match that of the GoPro. This ensures that each frame of this sensor has a unique timestamp which can be matched to another in the other modalities. With 132 unique combinations, the sequence begins anew every 2.2 seconds. The experiment occurs over a period of 45 seconds, meaning the sequence occurs roughly 20.5 times. It is presumed that there is enough distinction between actions of a scene to allow visual based temporal syncing to within a one second accuracy; LEDs will be used for finer distinctions.

### 5.4.3 Multimodal Considerations

The only multimodal concerns using the LEDs come from the thermal imagers inability to perceive the relatively low change in temperature from the rapidly actuating devices. However, since it was previously confirmed that the imagers within WASP-Lite image to within 1/60 of a second of each other, for this application, it is acceptable to use the timing of another imager to temporally register the data. This induces an acceptable error accuracy of $\pm 1/60$ of a second.

### 5.4.4 Spatial Registration

As described in Section 4.3.2, the equipment was placed in such a manner as to reduce or even eliminate the parallax issues within the region of interest. However, due to the oblique imaging angle, there is no single (x,y) coordinate shift that would align each plane of the scene. For this reason, we focused on the central portion of the scene as portrayed in Figure 5.21.

Since none of the sensors were viewing the scene through a common optic, each sensor needed to be properly registered to the base. In this instance, the WASP-Lite high resolution panchromatic imager was chosen to act as the spatial registration base. The

FIGURE 5.21: Region of Interest within FOV

reasoning was twofold: first, all the WASP-Lite sensors are co-boresighted (they are mechanically aligned to have parallel optics), meaning a simple pixel shift would be enough to properly register them; second, this broadband sensor had the highest resolution that covered the entire FOV of the region of interest.

We note that the registration occurred on the first image of each temporally registered sequence. In order to determine the correct transformation matrix to apply to each of the sensors, a few basic assumptions need to be understood. First, the hardware was set up such that each sensor was parallel to every other, thereby reducing the odds of perspective issues between adjacent views. Second, only the multispectral sensors are using the "exact" same camera and optical train, thereby guaranteeing duplicative FOVs and GSDs amongst them. That being said, the FOV and GSDs of the panchromatic imager are different from all the others, thereby forcing a reliance to match specific features amongst the imagery to properly register. The SURF detection algorithm was used to perform this task.

#### 5.4.4.1 Feature Matching

Using the assumption that each sensor was placed parallel to every other, it was expected that enough common matching features amongst the imagery would result in an

affine transformation matrix for registration. Understanding that each imaging modality had a different GSD from that of the panchromatic, it was necessary to reduce the panchromatic GSD to match that of the other imagery before applying the SURF algorithm. This reduction and SURF application was performed in a two step process for each pairing of imagers (i.e. panchromatic and multispectral camera 1).

First the panchromatic image was blurred by using a standard odd size averaging blur kernel. Then the SURF algorithm was applied to both images. Once the set of features was detected in each image, the two nearest neighbors were retained as possible point correspondences. Finally, a closeness rating of 0.7 was used to determine which pairs were close enough to be kept as good features.

This process was applied several times by changing the size of the blur kernel and counting the number of 'good' features that remained after the process was complete. Figure 5.22 depicts the results of several blur and SURF iterations between the panchromatic and GoPro imagery.

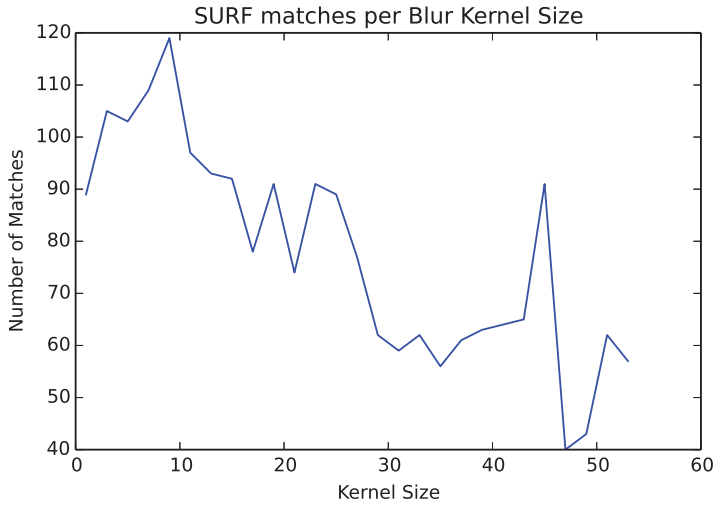



FIGURE 5.22: Blur and SURF Results

This figure shows that there are several peaks that are produced as the image becomes more and more blurred. These are comparable to adjusting a lens to focus on a particular object. Not knowing which would produce the best image, the features from the top three blur kernels were used to perform the registration on the GoPro imagery. This

registration was done by finding the homography matrix using a RANSAC to determine the best fit between the panchromatic and GoPro matched features.

To determine which blur kernel provided the best results in an automated fashion, the Sum Square Error (SSE) of the panchromatic image and the newly registered GoPro imagery was analyzed. First, the panchromatic image and a grayscale version of the GoPro imagery were peak normalized and overlaid on a common axis. Then the SSE of the image registration was evaluated by

$$SSE = \sum_{x=0,y=0}^{x=n,y=m} (I_{x,y}^{GoPro} - I_{x,y}^{Pan})^2 \tag{5.7}$$

where x and y represent spatial locations and n & m are stand in variables representing the full spatial extent of the image (i.e. at 1600x1200 pixels n=1200 and m=1600).

Figure 5.23 depicts the visual results of the various blur kernels in a three-channel (RGB) image. To simulate this, the Red and Blue channels were filled with the panchromatic image and the Green channel was filled with the greyscale registered GoPro Image. The left side depicts the registered imagery with non-common overlap included, whereas the right side masks out non-overlapping portions of the scene. The titles of each image indicate the blur kernel size and amount of Sum Square Error (SSE).

Once the appropriate transformation matrices were developed for each of the multispectral cameras and the GoPro, they were individually applied to each of the images in the image sequence. Appendix B depicts the results of each of the spatial registrations for the multispectral imagers.

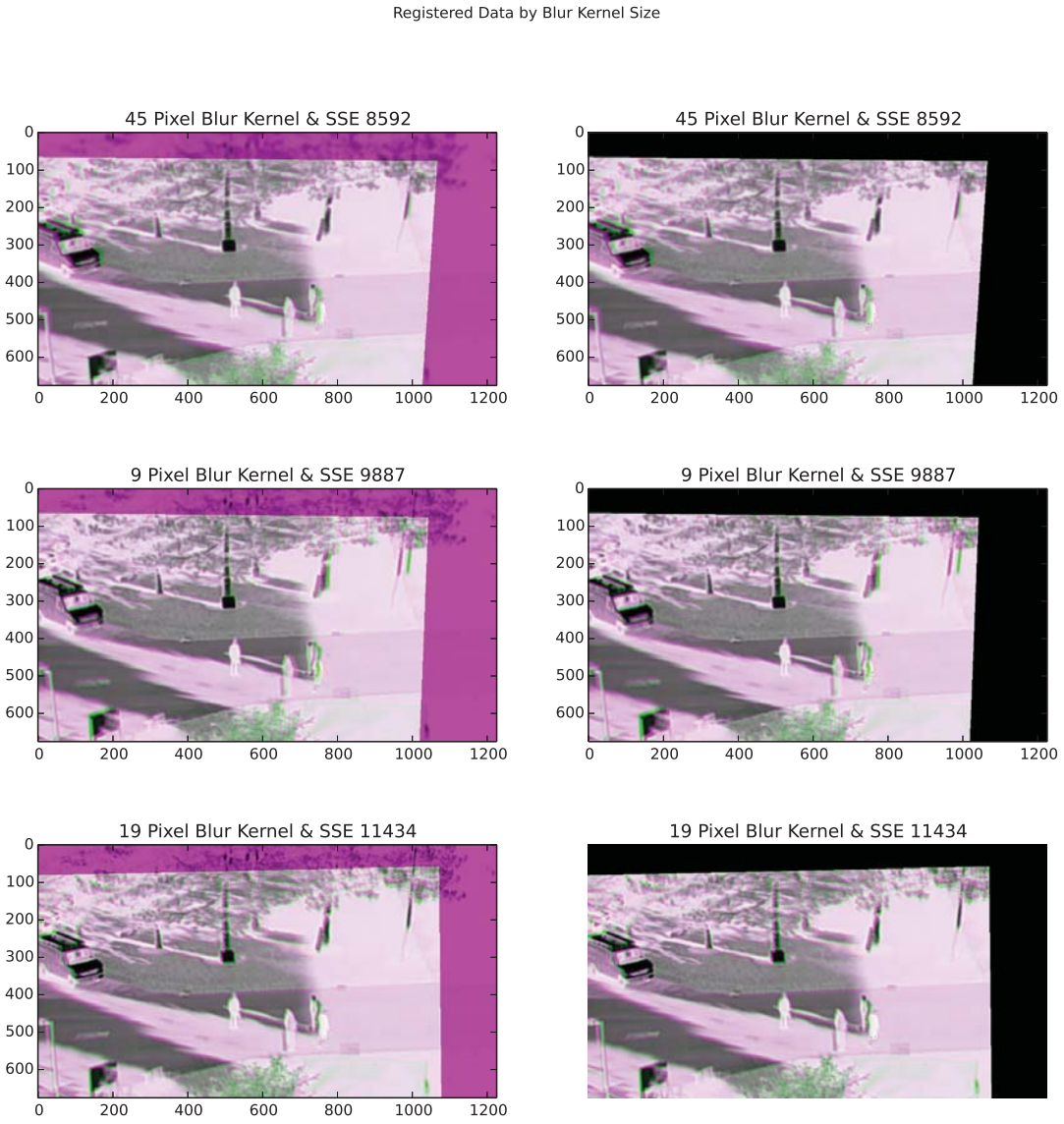



FIGURE 5.23: Registration results from varying blur kernel sizes. Note, the left contains the entire image from both imagers, whereas the right masks out non-overlapping portions of imagery. The Red and Blue channels were filled with the panchromatic image and the Green channel was filled with the greyscale registered GoPro Image. The titles of each image indicate the blur kernel size and amount of Sum Square Error (SSE).

## 5.5 Data Fusion

As stated in the background section, there are three levels of fusion that can occur: pixel, feature, and decision. This research will concentrate on the pixel level fusion wherein each modality will be placed into a multimodal data cube for further evaluation.

### 5.5.1 Pixel Level

Upon proper registration of the disparate modalities, each was stacked behind the others in a multimodal data cube representing the scene. Due to the differences in the temporal resolution, not all modalities will initially be represented in each data cube produced. The GoPro will be the basis for each cube, with empty placeholders (channels comprised of all zeros) being used to keep a consistent order among all the cubes in the temporal data set. Figure 5.24 depicts a multimodal data cube of one of the frame's in this dataset.

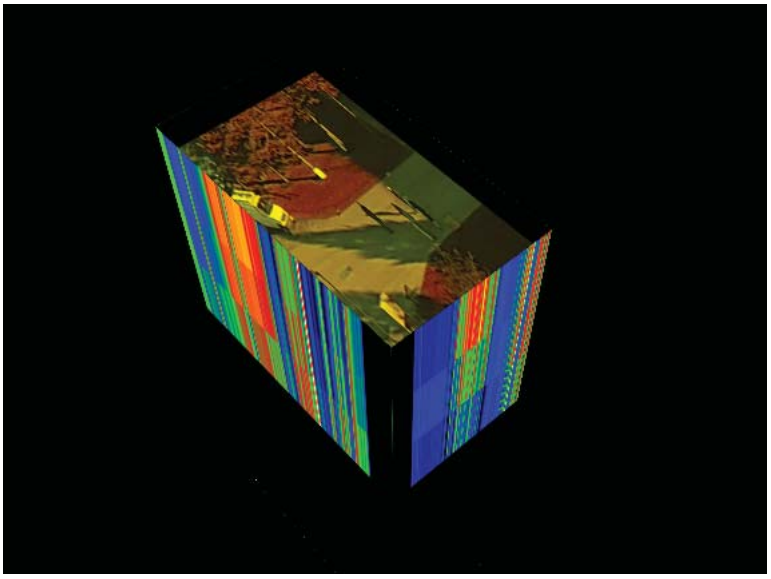


FIGURE 5.24: Multimodal Data Cube

### 5.5.2 Change Detection

Once the GoPro imagery is evaluated for tracking purposes, only those pixels indicating foreground objects will be considered for further evaluation. This binary change detection image will be placed on top of the data cube acting as a mask for the information in the adjacent modalities. Multimodal information will be subsequently tagged as belonging to the pixels indicating motion in the tracking phase and activity recognition will occur in the next phase.

### 5.5.3 Polarimetric Data Fusion

The data from the MAPPS sensor needed to be temporally aligned before polarimetric analysis could occur. The sensor takes an image for each polarimetric filter then rotates to the next filter in sequence. Figure 5.25 depicts the flow chart for collection and processing of the MAPPS data and Figure 5.26 depicts how often a polarimetric set is available to the data cube. For each Stokes vector, a degree of linear polarization is analyzed. In this case, we are making the inherent assumption that the circularly polarized component (S3) is roughly equal to zero, thereby equating the Degree of Polarization (DoP) to our Degree of Linear Polarization (DoLP).

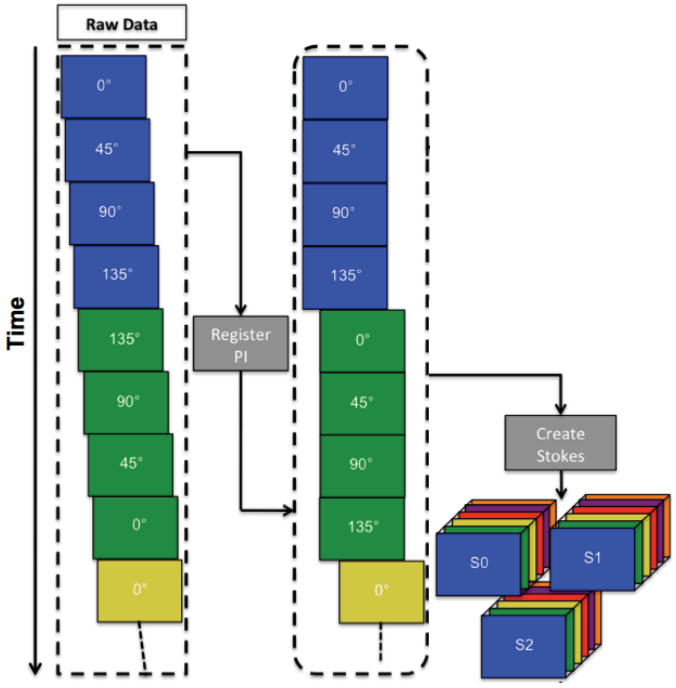



FIGURE 5.25: Multiplexed Processing Sequence [11]

With the multimodal data cube, per pixel evaluations of objects of interest can be interpreted. A object depicting a high DoLP relative to its surrounds will be the discriminator to determine what objects are interesting. Once a series of pixels has been tagged as interesting the target detection algorithm will be cued to track this grouping of pixels through the remaining sequence of data. Since this activity is only concerned with moving polarimetric pixels, the tracking algorithm will associate the polarimetric "signature" with the moving object in the scene. Therefore, even if the signature

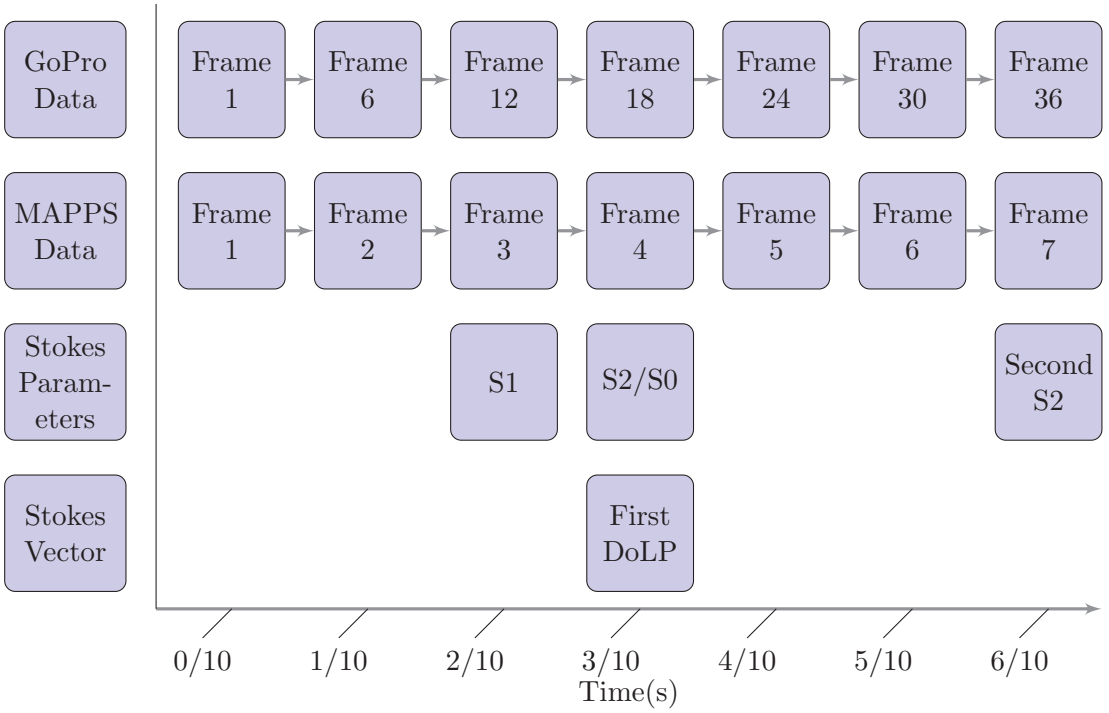



FIGURE 5.26: Temporal Data Association

fades due to a change in the sun-target-sensor geometry, the object will still be tracked throughout the remainder of the video.

We note that it is possible for a single object to exhibit a range of high DoLPs throughout a collection period. This change in DoLPs is due to the changing position of the sun over time. Due to the short temporal span of this experiment, this is unlikely to have occurred. After developing the DoLP imagery, a tracking method will be applied to the polarimetric data. These tracks will be correlated with the tracks from the GoPro imager to match moving people across the two systems. Afterwards, Those people carrying objects depicting a high DoLP will be identified in the GoPro imagery as having a polarimetric signature.

## 5.6 Tracking

Tracking, as described in this section, is broken up into two sections: target detection and track association.

### 5.6.1 Target Detection

Target detection is the act of identifying specific points of interest in a given scene. This is accomplished by walking through a series of steps in the target detection sequence depicted in Figure 5.27. This essentially becomes a computer vision problem, in which the noise (i.e. background) needs to be reduced in favor of the targets of interest. Forsyth, Szelinski, and Solem all describe varying methods of filters, averaging, optical flows, and segmentation algorithms that could be used as possible solutions[57–59]. A combination of these are included in the background suppression element depicted in Figure 5.27. A background image of the entire video was developed by averaging all the images on a pixel-by-pixel basis. Then a difference image is constructed from the current frame and background image. Both Zhang and Ausfeld used similar techniques when assessing change detection in their polarimetric and infrared research, respectively [60, 61]. Then the remaining steps in Figure 5.27 are applied using empirically derived values. While helpful, this process does not completely isolate moving objects of interest from background clutter (i.e., leaves on trees). Finally, tracks are maintained using a gating technique for further analysis.[62]

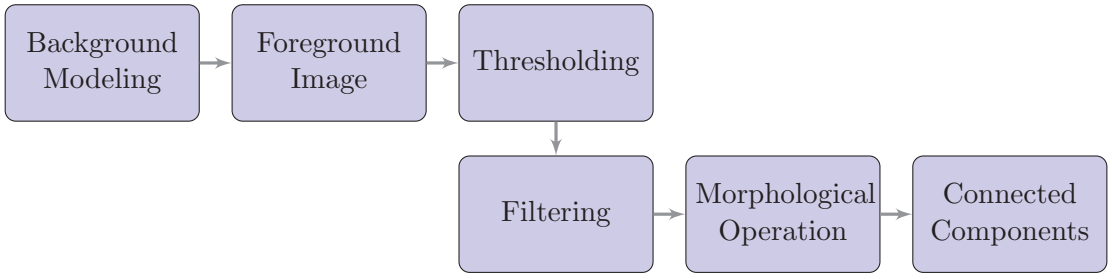



FIGURE 5.27: Target Detection Flow Diagram

#### 5.6.1.1 Background Modeling

The background of our given sequence is modeled by taking the pixel-by-pixel average of all the frames within the video sequence. Here we describe the background as the image that would occur if all the moving objects within the sequence were removed. We note that 'background' of the scene is contingent on how long and how often the imaging sensor captures an image of the scene. Over a period of months, the dominant change

in the image would likely be the trees senescence cycle. Comparatively, over hours and minutes it may be a change in the number of parked cars and people passing through. Figure 5.28 depicts that modeled background for the evaluated data.


FIGURE 5.28: Background of the video sequence

#### 5.6.1.2 Foreground Image

Once a background was developed, a foreground image was produced for each frame in the sequence. To accomplish this, each frame in the sequence was differenced with the background image. Since both images have objects moving through low and high intensity areas, there exists the possibility of obtaining both positive and negative values. An absolute value of this image was taken to ensure all values were positive. Figure 5.29 depicts one of the foreground images in the sequence.

#### 5.6.1.3 Thresholding

The foreground image primarily contains moving objects, indicative of actual targets and noise (leaves moving in trees). The noise is caused by a number of factors including: slight per pixel intensity changes, leaf movement on trees, and shifting shadows. In order to reduce the noise due to subtle shifts in intensities, shadows, and slight leaf movements,


FIGURE 5.29: Foreground of first frame in the video sequence

the foreground image was thresholded. Empirically, a threshold of 20 digital counts was determined to provided an adequate amount of noise reduction while retaining a reasonable amount of actual targets in the image. Figure 5.30 depicts the output of this step.


FIGURE 5.30: Thresholding of foreground image

#### 5.6.1.4 Filtering

Another method of reducing the previously mentioned noise is to implement a filter. A median filter was chosen to find and remove pixels under the median digital count of the remaining pixels. The thought behind the median was to indicate that legitimate targets would produce a more easily detected difference from the background in the steps above. Leaves blowing in the wind would not produce large values in the foreground image. Figure 5.31 depicts the output of this step.


FIGURE 5.31: Median Filter of threshold image

#### 5.6.1.5 Morphological Operations

Morphological operations were used to ensure confidence in detecting the humanoid targets within the scene, while removing additional noise throughout the image. To accomplish this morphological closing was performed with a large elliptical kernel, empirically determined to represent the silhouette of a person within the scene. Figure 5.32 depicts the output of this step.


FIGURE 5.32: Morphological Operation of Median Filter

#### 5.6.1.6 Connected Components

Finally a connected components analysis was used to differentiate large objects from closely spaced smaller clusters of objects. Normally, this analysis is utilized to differentiate different islands of objects within a scene, however, in this particular situation, it was used to differentiate the length of the connected segments of a particular island. It can be seen that the noise from the above step was still small relative to the actual targets of interest in the center of the image. Therefore, by counting the length of the connection in each island, and setting an empirically derived minimum connectivity, we can filter out smaller remaining noisy elements within the scene. This became a binary image, where all islands above the threshold were set to one, and everything else was set to zero. Figure 5.33 depicts the output of this step.

#### 5.6.1.7 Target Locations

Once the final binary connected component image was created, an OpenCV function called "findcontours" was used to wrap the individual islands and determine their centers. This function works by implementing a topological border following technique developed by Suzuki and Abe [89]. It was implemented to follow the borders of structures

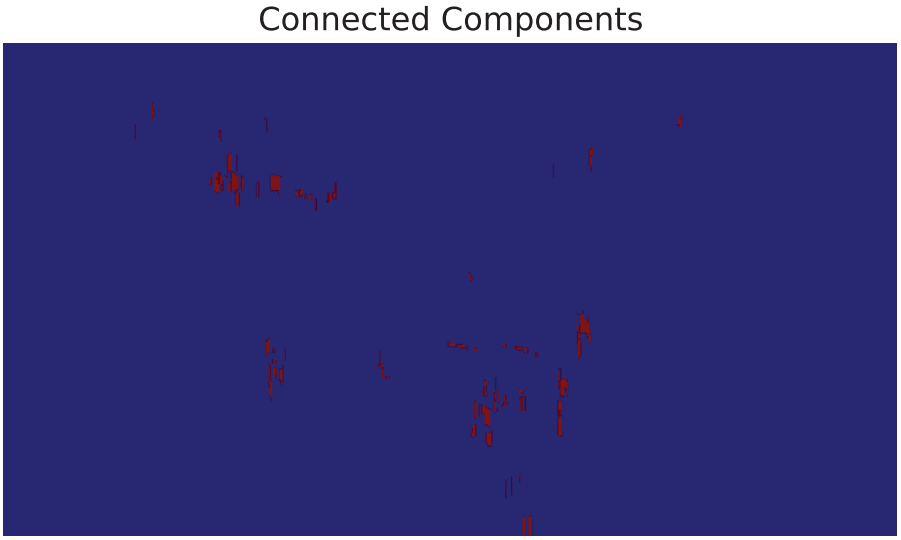



FIGURE 5.33: Connected Components of Morphological Image

in binary images in order to determine their most external outline. Once completed, the center of the object is determined and saved for further analysis. Figure 5.34 depicts the location of the centers outlined by red circles for easy identification.

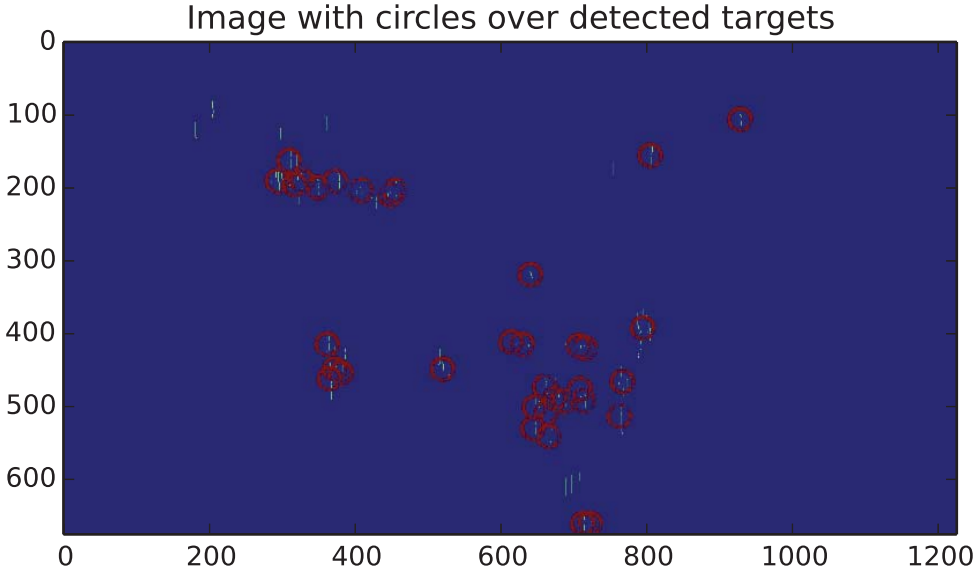



FIGURE 5.34: Centers of identified targets

#### 5.6.1.8 Consolidation

As indicated in the above Figure 5.34, there can be many targets identified within the image. In order to reduce the number reduce the number of duplicative and false, a comparison of points is performed. Even with the previous cleaning steps above, there are still many points where noise and multiple parts of the same person are identified. In order to further reduce the noise, a stipulation is placed that there must be at least one point within the area. This area was defined by creating a ellipse that represents the silhouette of a humanoid within our video. Then each point was compared against all others and a series of pairs were formed. Since a series of pairs may have common points within, an analysis was performed in which all common points were consolidated. After consolidation, the average location was found and used as the target's location. Figure 5.36 depicts the output of this step.

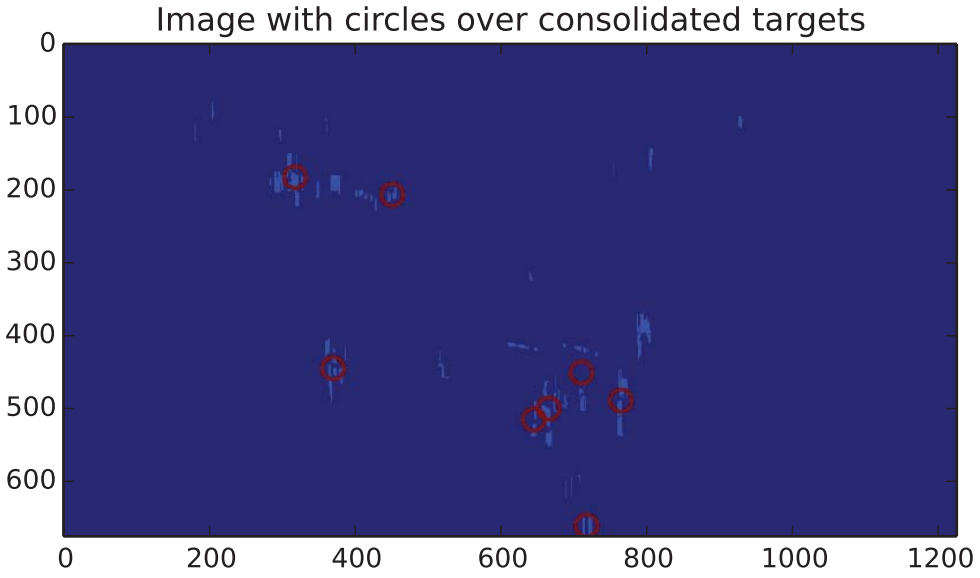



FIGURE 5.35: Consolidate centers of identified targets

### 5.6.2 Track Maintenance

Having detected targets in each frame, the challenge is to now associate the tracks from one frame to another. When doing so, we can say that a track either belongs to a previously tracked target or does not. In order to determine whether it is a previously tracked target, we compare the current location of the tracked object to the locations of all the new detections in the image. These values are placed into a Munkres assignment matrix [90, 91] with previously tracked objects.

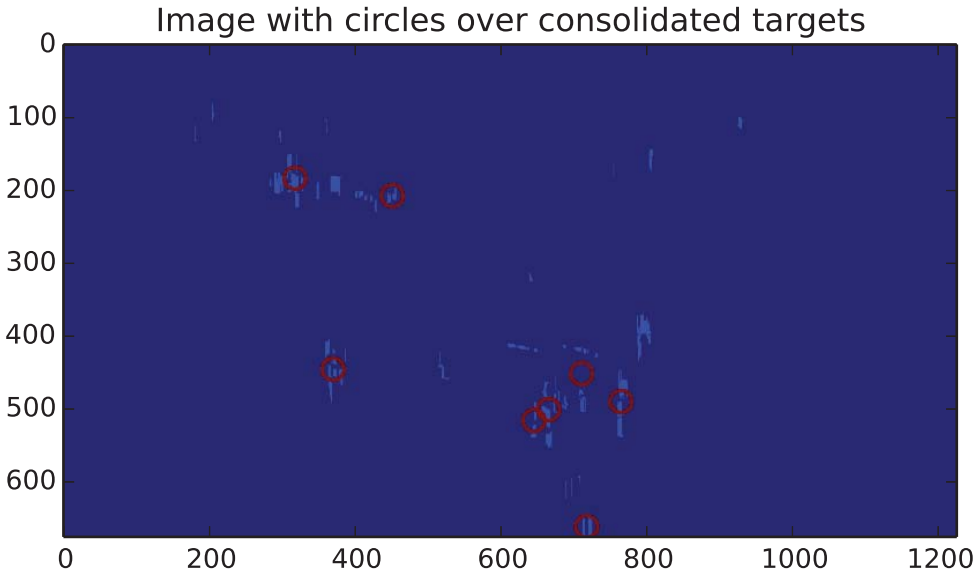



FIGURE 5.36: Consolidate centers of identified targets

#### 5.6.2.1 Munkres Assignment Algorithm

The Munkres assignment algorithm, also known as the Hungarian Method, is a combinatorial optimization algorithm that solves the assignment problem in a polynomial time rather than exponential. In this execution it attempts to minimize the cost associated with assigning a series of object locations to previously identified objects. Equation (5.8) depicts a matrix where previously found objects are given the choice of updating to one of three new positions. Each number in the matrix represents a range between the last known position of the object and a newly provided position.

$$\begin{array}{c} \\ \text{New Position}_1 \\ \text{New Position}_2 \\ \text{New Position}_3 \end{array} \begin{array}{c} \begin{array}{ccc} \text{Obj}_1 & \text{Obj}_2 & \text{Obj}_3 \end{array} \\ \begin{pmatrix} 5 & 9 & 1 \\ 10 & 3 & 2 \\ 8 & 7 & 4 \end{pmatrix} \end{array} \tag{5.8}$$

#### 5.6.2.2 Manual vs. Automatic Tracking

Once applied, the automatic tracking algorithm did not perform in an optimal fashion. The performance of the target detection algorithm can be evaluated by using signal detection theory. This required a manual target detection of each person in each frame. A true positive rate is defined as each correct detection the algorithm picked when compared to the manually detected location of the person within an image. When comparing the automatically detected position vs. the manually detected position, a target within 30 pixels was counted as a detection. A false positive is defined as any detected target that is not within that 30 pixel area of a manually detected target. The 30 pixel threshold was empirically derived.

On average, the true positive rate of the detection algorithm was only 14%, with 3.86 false positives for every target in the image. This led to about 17.3 false positives per image. Since the activity recognition algorithm needs a higher detection rate, it was decided that the manual target detection dataset would be used in lieu of the automated target detection dataset.

### 5.6.3 Tracking Results

Using the manually detected targets, the Munkres Assignment algorithm was able to achieve a 100% correct assignment of each of the people in the scene. Figures 5.37 through 5.40 depict a successful track association sequence where each of the individuals within the scene maintains a constant numerical indicator above their head. Person 1 and Person 3 are the two individuals engaging in the object exchange. In Figure 5.38 Person 3 can be seen handing off the object to Person 1. In Figures 5.39 and 5.40 a passerby is tracked through the scene and is represented by the number four over their position.

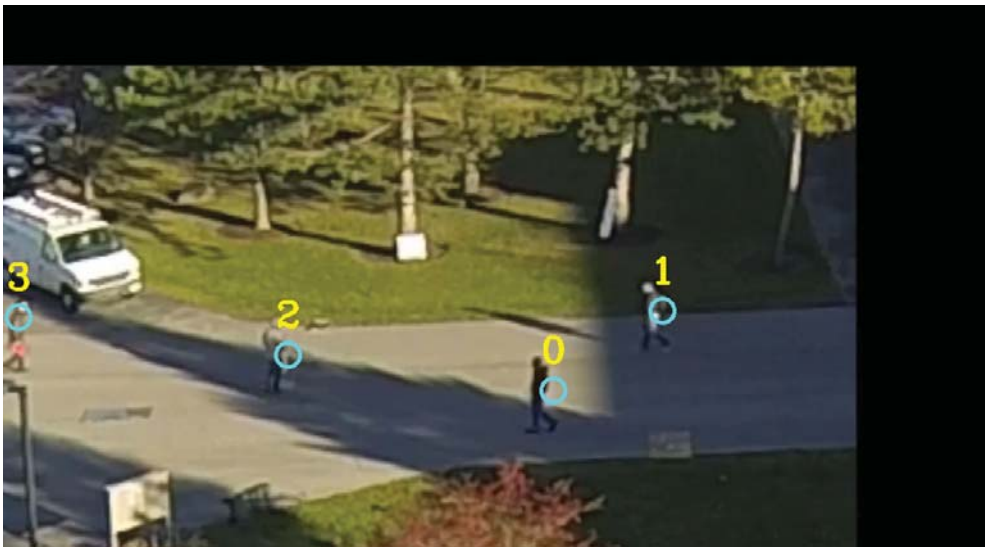



FIGURE 5.37: First Frame in Tracked Sequence

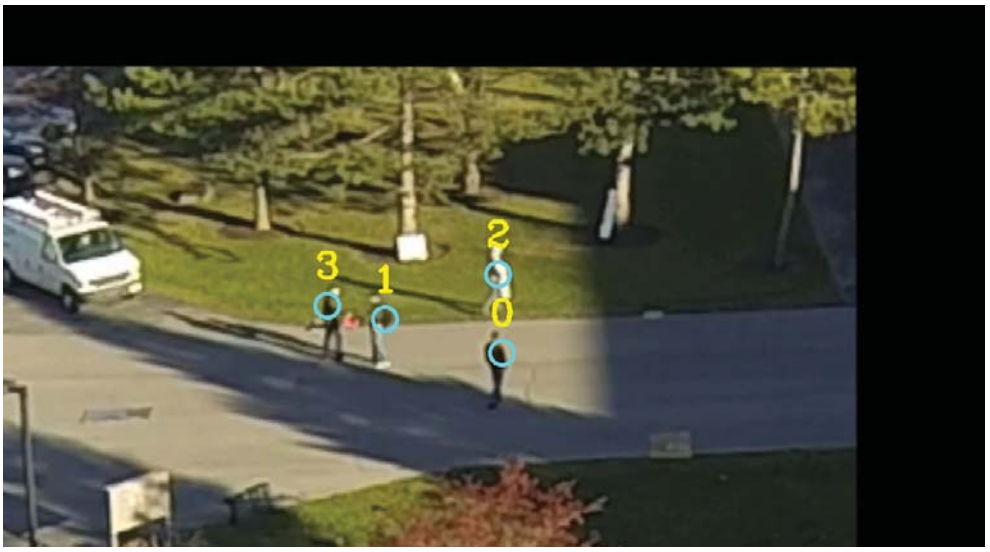



FIGURE 5.38: Object Exchange in Tracked Sequence

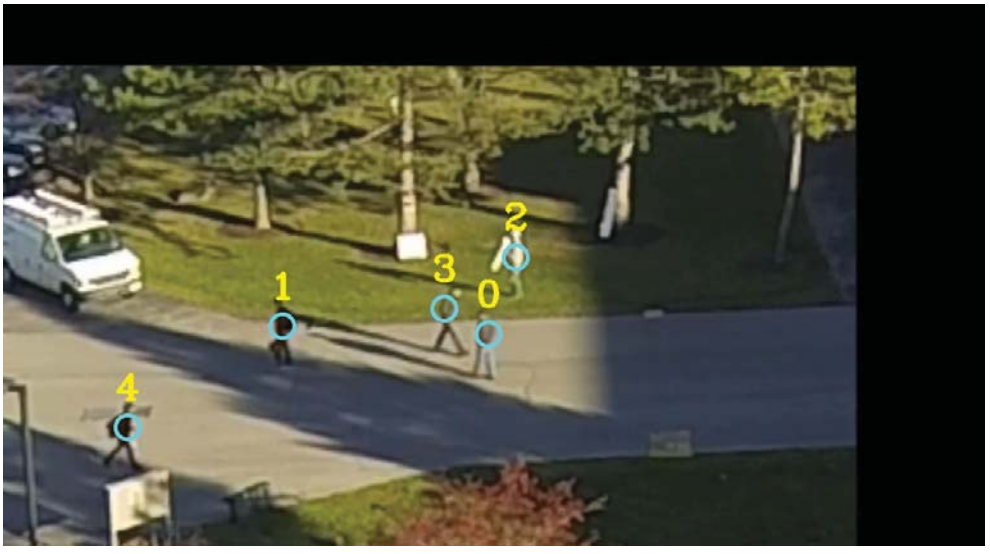



FIGURE 5.39: Post Object Exchange in Tracked Sequence

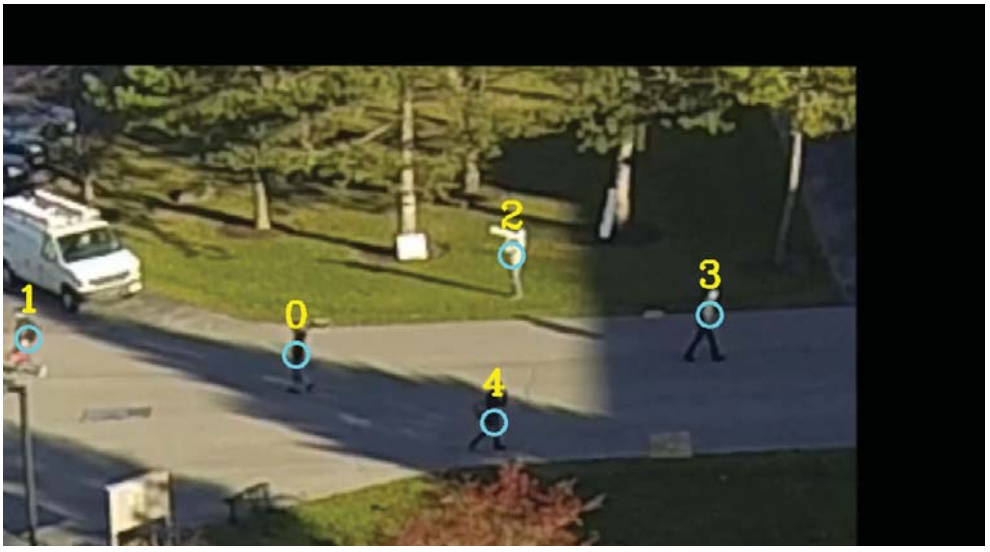



FIGURE 5.40: Additional Person in Tracked Sequence

## 5.7 Activity Recognition

In this portion of this research, there are two activities considered as interesting. The first is an object exchange between two people and the second is the detection of an object with a high DoLP. Section 5.7.1 will address the object exchange while Section 5.7.2 address the polarization activity.

### 5.7.1 Object Exchange

Once the tracking algorithm produced an adequate set of tracks for the moving objects within the scene, an activity recognition algorithm was applied. The spectral signature of each person was computed in the first few frames and compared against all future frames for signs of change. If two signatures within a close spatial proximity depicted a change at some point in the sequence, then an exchange is said to have occurred. By using Spectral Angle Mapper (SAM) between two spectral signatures, an angle can be used to determine how the spectral signature of a person changes over time. This technique's illumination invariant nature makes it possible to compare signatures of a person moving in and out of shadow. The following steps are used to determine the existence of an object exchange:

1. Develop a pixel mask indicative of foreground objects. This is done by using the pixels in the threshold image derived from the target detection workflow; Figure 5.30 depicts the threshold image.
2. Apply the mask to all bands in the data cube.
3. Apply a bounding box around the detected locations of each person within the image.
4. Take the band-by-band mean of the pixels within the bounding boxes
5. Place the means into a vector. This is considered the spectral signature associated with the detected person.
6. Perform this technique for every frame and every object in the sequence.

7. Average each object's spectral signature from the first 10% of frames in the sequence. This average will act as the reference signature for each individual.

8. Calculate the spectral angle between each object's reference signature and the signature found in each frame.

9. Determine if any person has a spectral angle above an empirically derived threshold.

10. Reduce the number of people being evaluated by using a spatial filter. If two people are not within a close spatial proximity, then it is not possible for them to exchange an object.

Figure 5.41 depicts the above steps in the workflow. Note that as part of this workflow, there are steps performed on each band, each person detected in the scene, and each frame in the sequence. The flow begins by taking the threshold image from the target detection workflow and performing a series of operations.

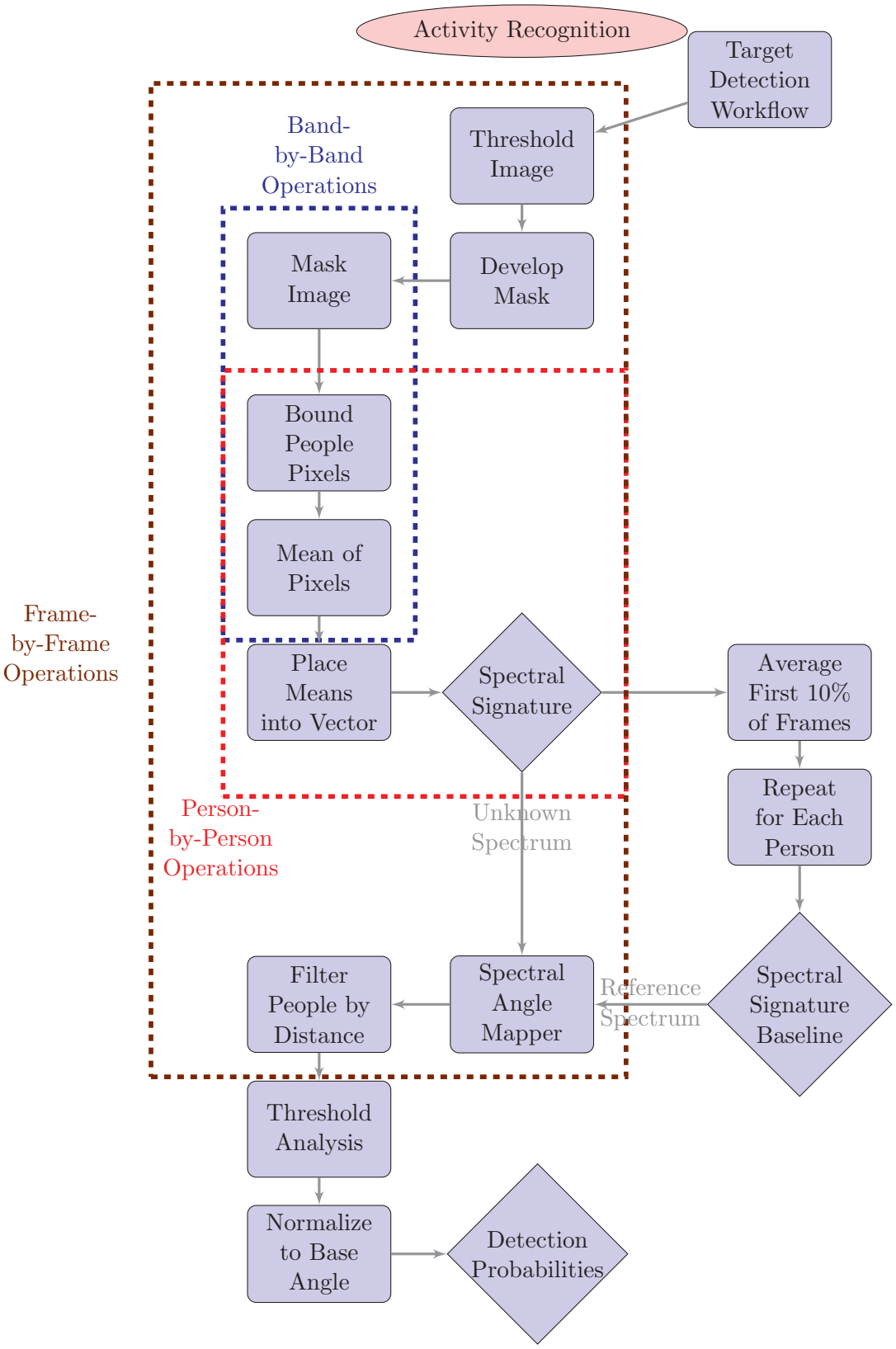



FIGURE 5.41: Object Exchange Activity Recognition Flow Diagram; The dotted boxes indicate where the type of operation is performed. The flow begins by taking the threshold image from the target detection workflow as indicated in the upper right hand corner of the figure.

#### 5.7.1.1 Band-by-Band Operations

**Mask Image** The mask is developed by changing the threshold image into a binary image. Any pixel with a value is changed to a one and any without a value is kept zero. For example, Figure 5.42 is a sample image from the video sequence. By taking the threshold image of this particular frame and making it a binary image, the image mask is created. Figure 5.43 depicts this mask.

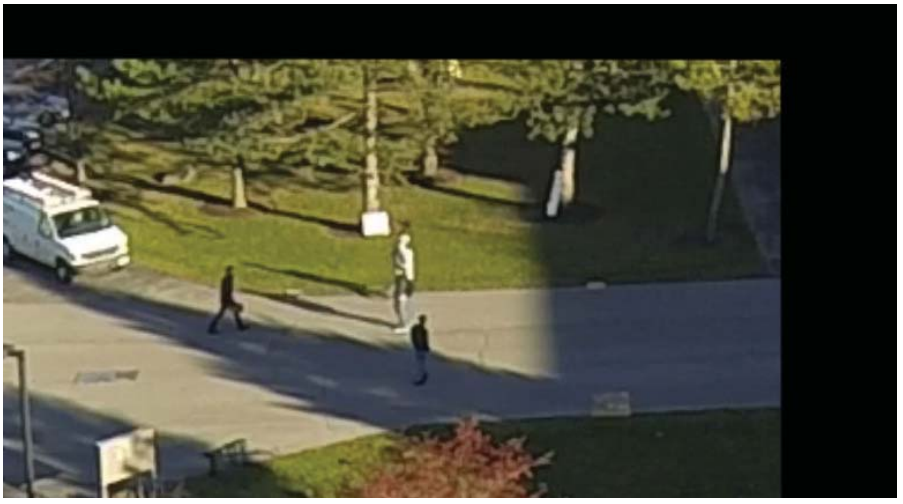


FIGURE 5.42: Image to be Masked

Once the mask is developed, it is applied to the image to remove all background data. Figure 5.44 depicts the final background image used in the masking. This is accomplished by multiplying the mask by each channel in the image. Figure 5.45 depicts the inverse mask, which is easier to interpret.

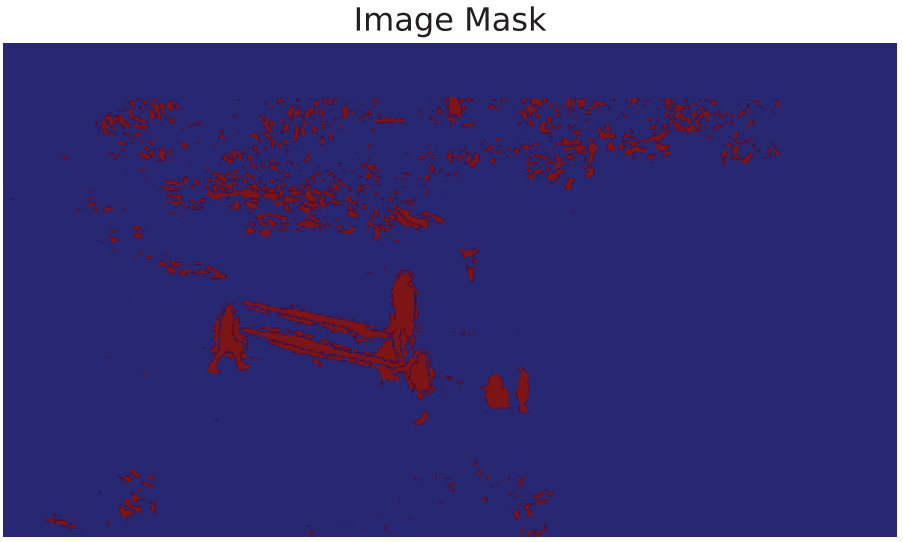



FIGURE 5.43: Image Mask


FIGURE 5.44: Masked Image


FIGURE 5.45: Inverse Masked Image

**Bound People Pixels** Once the image has been masked, a bounding box is created around a point indicative of a detected target. The box is intentionally made such that it will encompass most of an individual in each frame. Using the image above, Figure 5.46 depicts the two individuals used to describe the bounding situation. The following side-by-side figures depict the actual size of the bounding boxes used to create the object spectral vectors. Figures 5.47a and 5.47b depict a side-by-side image of Person 3 in Figure5.46.

The bounding box could have been made larger to ensure it retained all pixels related to the object, but empirical results showed that doing so includes other undesired foreground. Figures 5.48a and 5.48b depict a side-by-side image of labeled Person 1 in image 5.46. Notice the difference in the amount of information included between Figures 5.47b and 5.48b. When an individual is near others or surrounded by shadows, the previously defined algorithms include that pixel information in the spectral mean content. One method to avoid this is by using a smaller bounding box. However, doing so has presented some adverse empirical results indicative of a loss in spectral signature uniqueness. A full range evaluation on the proper size of the bounding box was not completed, but is left for the assessment of future researchers.

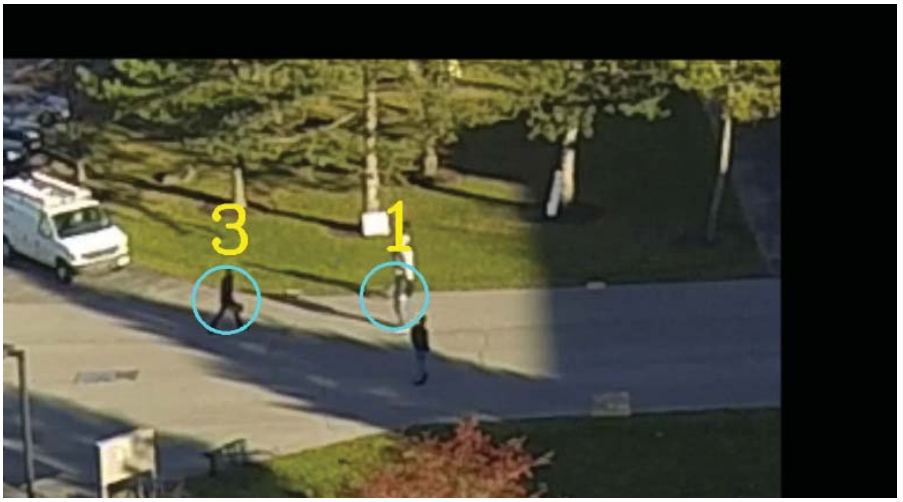



FIGURE 5.46: Inverse Masked Image with Individuals labeled

(A) Original Image (B) Masked Image

FIGURE 5.47: Bounding Box Around labeled Person 3

The size of the box is 60 pixels in the x-direction and 100 pixels in the y-direction, centered on the detected target location. The manual tracking kept the detected target location on the upper body of the people walking through the scene. There is some

(A) Original Image (B) Masked Imaged

FIGURE 5.48: Bounding Box Around labeled Person 1 with Cluttered Surroundings

variability in the tracked position, but its effects appear to be minimal as most of the head, torso, and legs can be seen in these frames.

Using a panchromatic resolution image, we calculate this window to be roughly 118.8cm across and 198cm high. The spectral data has a GSD roughly 2.5 times greater than that of the panchromatic, thus providing roughly 47.5cm across and 79cm high. These dimensions are enough to cover a significant portion of the person and the object being held.

**Mean of Pixels** Once the bounding boxes have been created around each of the people pixels, the means of each band are taken. Each mean is then placed into a vector denoting that person's spectral signature at that frame.

#### 5.7.1.2 Person-by-Person Operations

Due to the existence of multiple objects within the scene, many of the band-by-band operations that have been previously completed, have to be redone for each person in

the scene. The prior two operations of bounding the people pixels and taking the mean of each band, are also done on a person-by-person basis.

**Spectral Signature** As described in Section 5.7.1.1, the the mean of each channel is written as a vector. This vector now constitutes the spectral signature of an individual person. This process is performed for each person identified within the frame.

**Reference Spectral Signature** In order to develop a reference spectral signature, each person's spectral signature is averaged over the first 10% of the frames in the sequence. The purpose of this is to develop a robust signature unique to the individual despite extraneous foreground clutter. We also note that since the person is moving, it is unlikely that the imagers will ever see two positions of the exact same orientation or spatial extent. Thus, as the person moves through the scene their body and clothes will reflect different levels of radiance back to the sensors. This reference signature, or baseline signature, is only completed once for each person in the scene and then used in future frames.

#### 5.7.1.3 Frame-by-Frame Operations

Aside from the reference signature, each of the steps above was performed on a single frame. The next set of steps involves evaluating inter-frame data.

**Spectro-Temporal Interpolation** Due to the mismatch in temporal resolutions, the spectral data from WASP-Lite was only interspersed throughout the GoPro framing data. The GoPro equipment was operating at 60Hz while the WASP-Lite was set for 8Hz operation. While laboratory results confirmed these frame rate before the experiment was conducted, the WASP-Lite equipment was actually operating at a variable rate centered around 5.45Hz. Figure 5.49 depicts how the data originally came out of the process; note the drops where zeros were placed between spectral signatures.

There are 35 frames of spectral data over the 600 frames of GoPro imagery indicative of the object exchange. To fill in the gaps, a spectro-temporal interpolations was performed.

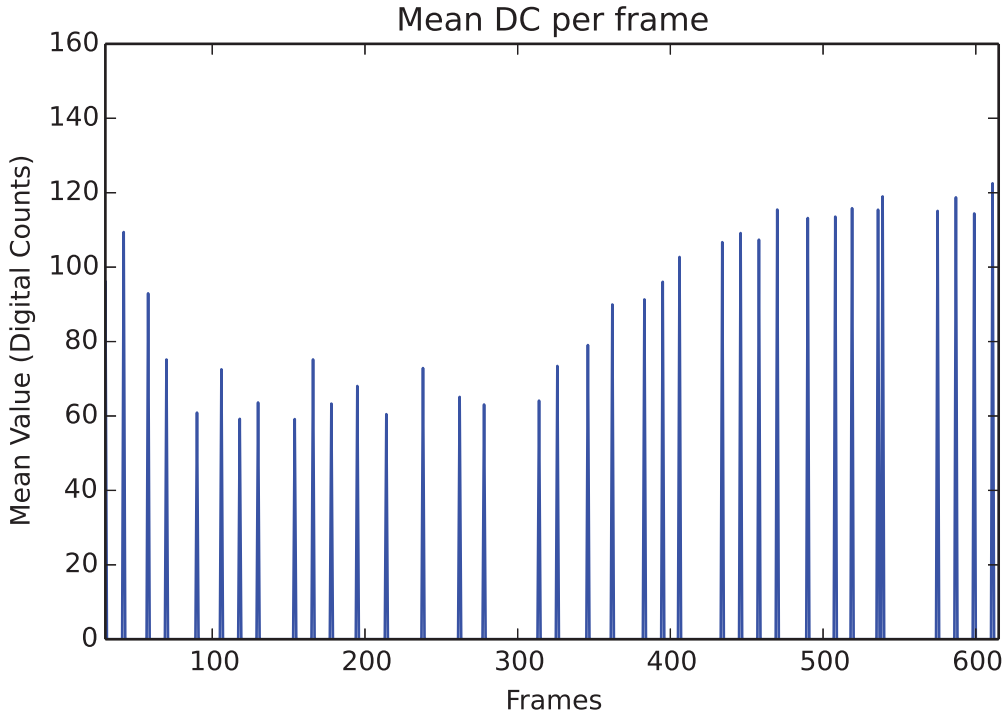



FIGURE 5.49: Original Mean Digital Counts per Frame for 630$\mu$m Imager

For brevity, the intermediary steps are left in Appendix F. Figure 5.50 depicts the results of the interpolation. 565 frames worth of spectral data were developed in this process.

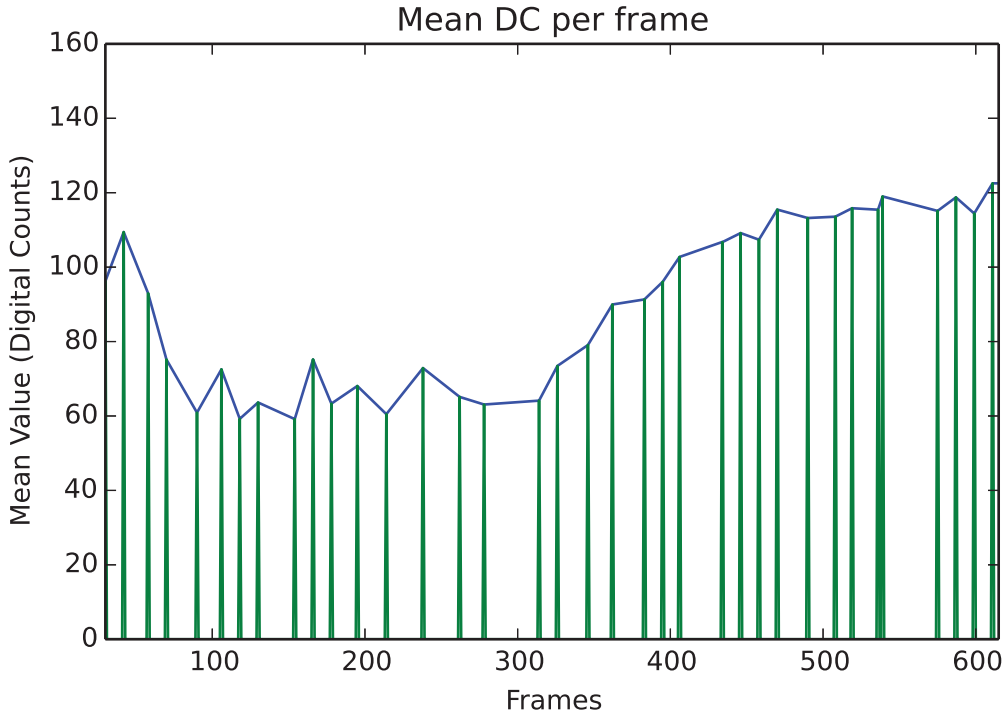



FIGURE 5.50: Interpolated Mean Digital Counts per Frame overlaid on Original Data

**Spectral Angle Mapper** The Spectral Angle Mapper (SAM) is a signature matched detection algorithm used to compare a reference spectral signature to that of an unknown signature. The spectral angle between two spectra can be computed by

$$r(\mathbf{x}) = \frac{(\mathbf{s}^T\mathbf{x})^2}{(\mathbf{s}^T\mathbf{s})(\mathbf{x}^T\mathbf{x})} \tag{5.9}$$

where s denotes the reference spectrum, and x denotes the unknown spectrum.

This represents the square of the normalized projection of the unknown spectrum onto the reference spectrum. Understanding that both the inverse cosine and square root are monotonic functions [92], this can be written as

$$r_{SAM}(\mathbf{x}) = -\cos^{-1}\left(\frac{(\mathbf{s}^T\mathbf{x})}{\sqrt{(\mathbf{s}^T\mathbf{s})(\mathbf{x}^T\mathbf{x})}}\right) \tag{5.10}$$

This suggests that smaller angles are more similar to the reference spectrum. In this research the spectral angle of a person at one point in time is being compared to the

spectral angle of the same person at a later point in time. In our case, we know that the angle between the two spectra should be small because they are theoretically coming from the same target. What is of interest is when the angle becomes large; this is indicative of a change in the spectral signature of the person.

**Filter People by Distance** Since this research is interested in finding an object exchange between two participants within the scene, the spectral angle charts can be reduced to retaining only those people that passed by each other within a preset elliptical distance.

#### 5.7.1.4 Threshold Analysis

Once the people have been filtered, it is possible to determine if an activity has occurred by evaluating the temporal change in spectral angle. At some point when the object changes hands, we expect there to be an increase in the immediate and overall angular difference of the data. In this research, it was decided that the spectral angle should increase by 10% compared to the pre-exchange mean spectral angle to be considered a change. Due to the controlled nature of this experiment, the time at which the exchanged occurred is well known. It is then possible to compare the mean spectral angles of the data before and after the exchange transpires. After confirming that in increase of the spectral angle has occurred, the post-exchange spectral angle is used for further evaluation.

#### 5.7.1.5 Spatio-Temporal Degradations

Once the post-exchange angle has been established, it is possible to perform spatial and temporal degradations on the data before reassessing the angular disparity associated with an object exchange. The spectral angle without the degradations would be considered a 100% likelihood of detection, and as degraded angles diverge from this value, the likelihood of detection would decrease.

It is important to note that the degradations were only performed on the data after the tracking occurred. To be clear, this means that all steps up to the activity recognition

portion of the methodology depicted in Figure 5.2 were done on the original 60Hz 5cm dataset. This is important to note as spatial degradations would have affected camera calibration, video stabilization, registration and tracking, likely causing systemic difficulties. The temporal degradations would likely have caused issues in the video stabilization and tracking steps. This portion of the research was interested in developing an activity recognition algorithm, it was more important to degrade the data going into the algorithm than into the process. Furthering this reasoning, it was stated earlier that each of the steps in the methodology could be done with any number of algorithms. Thus it can be assumed that the data was put through all prior steps with a similar level of success.

**Spatial Degradations** In this experiment, a change in pixel pitch was chosen to perform a reduction in the spatial resolution. However, it was noticed that downsampling the array caused issues with the location of the tracks developed from the tracking algorithm. Therefore the array size was kept the same and the blur was used to change the effective resolution of the imagery. The proper nomenclature with this change in resolution is Ground Resolved Distance (GRD), and will be used from here on. Section 3.2.3.1 discussed the difference between the two. Since the GSD of the original data is 5cm, each adjacent pixel extent in the blur kernel will change the GRD of the image data by a factor of 5cm. Thus a 2x2 blur kernel will provide a GRD of 10cm, a 3x3 will result in 15cm, and so forth.

**Temporal Degradations** The temporal degradations are developed by taking the interpolated temporal data, described in Section 5.7.1.3, and skipping select frames. At 60Hz every frame is included in the spectral angle analysis, however at 30Hz, only half of the original frames are included in the analysis. Without interpolating more frames into the dataset, it is only possible to attain temporal degradations in integer values divisible by the total number of frames. Table 5.4 depicts the frame rates included in this analysis, the associated number of frames, and the steps size between frames. The Step Size column depicts when the next frame in the sequence was included. Thus, at 1Hz the next frame used was 60 frames away in the sequence. To obtain the number of

frames skipped between each included frame, simply subtract one from the inter-frame step size.

TABLE 5.4: Frame Rates, Frame Count, Step Size, and Skipped Frames

| Frame Rate (Hz) | # of Included Frames | Step Size | # of Skipped Frames |
|---|---|---|---|
| 60 | 1000 | 1 | 0 |
| 30 | 500 | 2 | 1 |
| 20 | 333 | 3 | 2 |
| 15 | 250 | 4 | 3 |
| 12 | 200 | 5 | 4 |
| 10 | 166 | 6 | 5 |
| 8.57 | 143 | 7 | 6 |
| 7.5 | 125 | 8 | 7 |
| 6.67 | 111 | 9 | 8 |
| 6.0 | 100 | 10 | 9 |
| 5.45 | 91 | 11 | 10 |
| 5.0 | 83 | 12 | 11 |
| 4.0 | 66 | 15 | 14 |
| 3.0 | 50 | 20 | 19 |
| 2.5 | 41 | 24 | 23 |
| 2.0 | 33 | 30 | 29 |
| 1.5 | 25 | 40 | 39 |
| 1.0 | 16 | 60 | 59 |

#### 5.7.1.6 Likelihood of Detection

In order to develop a likelihood of detection for the degraded dataset, it is necessary to compare the degraded spectral angles to the spectral angle of the non-degraded data. This was done by normalizing degraded spectral angles by the non-degraded spectral angle; depicted analytically by

$$\theta_{Normalize} = \frac{\theta_{Degraded}}{\theta_{Non-Degraded}} \tag{5.11}$$

where $\theta$ represents the spectral angle of the data. If the non-degraded spectral angle is indicative of an object exchange, then spectral angles that deviate from this angle, either positive or negative deviations, present situations where it is less likely that an object exchange will be detected. For those values below one the values are left as they

are, however those values greater than one are reduced by their overage. This depicted analytically in the following example.

$$\begin{aligned} \theta_{Normalized} &= 1.3 \\ \theta_{Overage} &= 1.3 - 1 \\ \theta_{Overage} &= 0.3 \\ \theta_{Normalized\ Remapped} &= \theta_{Normalized} - 2 \cdot \theta_{Overage} \\ \theta_{Normalized\ Remapped} &= 0.7 \end{aligned} \tag{5.12}$$

Rather then simply multiplying the spatial and temporal likelihoods together to develop a likelihood surface, it was decided that developing each point independently would be best. In order to develop this activity-based likelihood surface, each of the spatial degradations was temporally degraded and the spectral analysis accomplished. The temporal values were first normalized independently before applying the spatial normalizations the entire dataset. Since every spatial and temporal degradation will provide a separate spectral angle, these values can be placed into a matrix for comparison. The columns will represent the spatial degradations and the rows will represent the temporal degradations. The following uses a notional matrix of spectral angles to depict this point

1. Notional matrix for spectral angles (non-degraded data included)

$$\begin{array}{c} \\ 60\text{Hz} \\ 30\text{Hz} \\ 20\text{Hz} \end{array} \begin{array}{c} \begin{array}{ccc} \text{5cm} & \text{10cm} & \text{15cm} \end{array} \\ \begin{pmatrix} 9 & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{pmatrix} \end{array} \tag{5.13}$$

2. Degrade the data spatially

$$\begin{array}{c} \\ 60\text{Hz} \\ 30\text{Hz} \\ 20\text{Hz} \end{array} \begin{array}{c} \begin{array}{ccc} \text{5cm} & \text{10cm} & \text{15cm} \end{array} \\ \begin{pmatrix} 9 & 7 & 3 \\ ? & ? & ? \\ ? & ? & ? \end{pmatrix} \end{array} \tag{5.14}$$

3. Degrade the data temporally

$$\begin{array}{c} \\ 60\text{Hz} \\ 30\text{Hz} \\ 20\text{Hz} \end{array} \begin{array}{c} \begin{array}{ccc} \text{5cm} & \text{10cm} & \text{15cm} \end{array} \\ \begin{pmatrix} 9 & 7 & 3 \\ 8 & 5 & 2 \\ 7 & 3 & 1 \end{pmatrix} \end{array} \tag{5.15}$$

4. Normalize row one from step two

$$\begin{array}{c} \\ 60\text{Hz} \\ 30\text{Hz} \\ 20\text{Hz} \end{array} \begin{array}{c} \begin{array}{ccc} \text{5cm} & \text{10cm} & \text{15cm} \end{array} \\ \begin{pmatrix} 1 & 0.78 & 0.33 \\ \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \end{pmatrix} \end{array} \tag{5.16}$$

5. Normalize each column in step three independently

$$\begin{array}{c} \\ 60\text{Hz} \\ 30\text{Hz} \\ 20\text{Hz} \end{array} \begin{array}{c} \begin{array}{ccc} \text{5cm} & \text{10cm} & \text{15cm} \end{array} \\ \left(\begin{array}{c|c|c} 1.0 & 1.0 & 1.0 \\ 0.89 & 0.71 & 0.67 \\ 0.78 & 0.43 & 0.33 \end{array}\right) \end{array} \tag{5.17}$$

6. Multiply each column in step five by the normalized value in step four

$$\begin{array}{c} \\ 60\text{Hz} \\ 30\text{Hz} \\ 20\text{Hz} \end{array} \begin{array}{c} \begin{array}{ccc} \text{5cm} & \text{10cm} & \text{15cm} \end{array} \\ \begin{pmatrix} 1.0 \cdot 1 & 1.0 \cdot 0.78 & 1.0 \cdot 0.33 \\ 0.89 \cdot 1 & 0.71 \cdot 0.78 & ... \\ 0.78 \cdot 1 & ... & ... \end{pmatrix} \end{array} \tag{5.18}$$

7. Normalized data

$$\begin{array}{c} \\ 60\text{Hz} \\ 30\text{Hz} \\ 20\text{Hz} \end{array} \begin{array}{c} \begin{array}{ccc} \text{5cm} & \text{10cm} & \text{15cm} \end{array} \\ \begin{pmatrix} 1 & 0.78 & 0.33 \\ 0.89 & 0.55 & 0.22 \\ 0.78 & 0.33 & 0.11 \end{pmatrix} \end{array} \tag{5.19}$$

This normalized data would then be plotted to depict detection graphs similar to those in Figures 2.1 and 2.2.

### 5.7.2 Detection of Highly Polarized Objects

Some activities include objects that can be highly polarized or depict a high DoLP. One such activity of interest is the movement and use of an RPG [86]. A method of detecting such preparations for launch are to look for objects with a high DoLP moving throughout the scene. For this activity, Person 2 in Figure 5.38 was given the PVC pipe described in Table 4.12 and told to execute a series of movements within the scene. These movements involved transitioning from one location to another, lifting the pipe onto their shoulder, and moving to a final location. Figures 5.38 through 5.40 depict and abbreviated portion of the sequence. Figure E.4 in Appendix E displays the full set of directions.

The specific activity methodology used to detect a polarimetric object is depicted in Figure 5.51. This activity recognition technique is simply searching the scene for a moving object with a high DoLP, tagging that object as interesting, and cueing another sensor for further investigation. The benefit of this technique is that the polarimetric

sensor only needs to be tipped once for the algorithm to cue an adjacent sensor for further evaluation. Thus, this subsection will attempt to prove that a polarimetric object exists and that it is possible to transfer that information to another sensor. Note, each time the polarimetric nature of the object is depicted, it is done in a different orientation. That is done to intentionally depict the orientation invariant nature of this technique against the cylindrical object.

The first step in the process was to determine of the chosen object produced a high DoLP relative to its surroundings. This is done by following the procedure in Section 5.5.3. The following sections detail the methods necessary to confirm an object has a high DoLP and how this DoLP is viewed in the field.

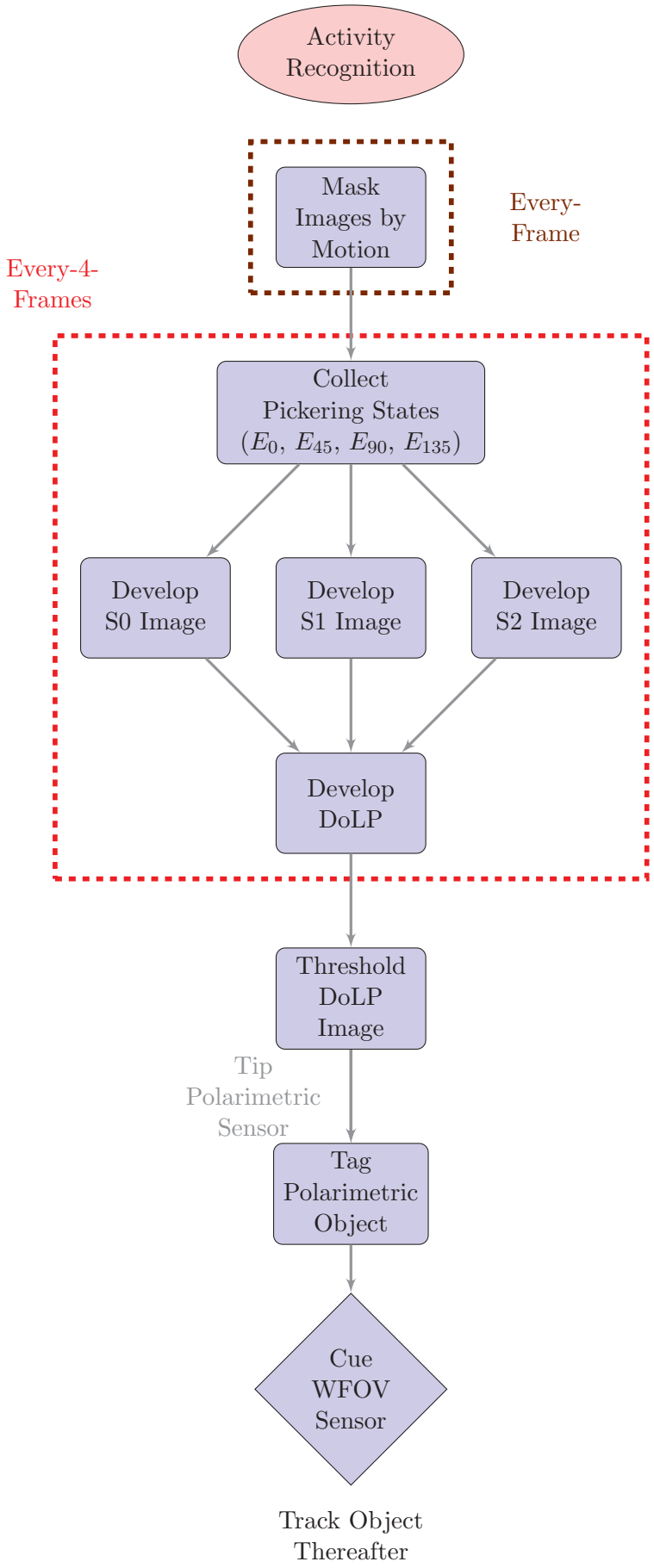
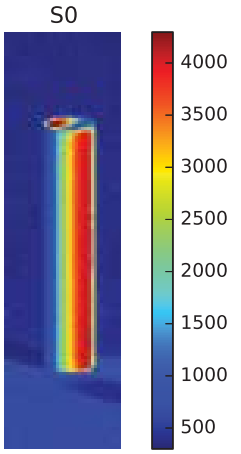


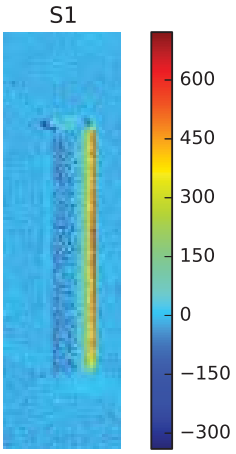

FIGURE 5.51: Polarimetric Tipping and Cueing Flow Diagram

#### 5.7.2.1 Stationary In-Scene Stokes Vector

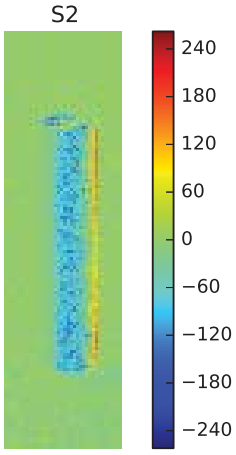
The same PVC pipe was placed in the scene standing on end and evaluated to determine if it still depicted high DoLP. Figures 5.52 depicts the S0, S1, S2, and DoLP results respectively of the stationary in-scene object. As the two tests indicate, the object does have a detectable DoLP.



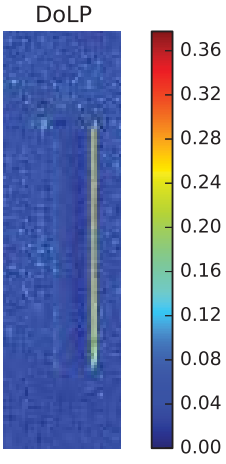

(A) S0 Component




(B) S1 Component




(C) S2 Component




(D) Degree of Linear Polarization

FIGURE 5.52: Stationary Polarimetric In-Scene Results of Object

#### 5.7.2.2 Moving In-Scene Masks

Due to the motion of the objects, full Stokes vectors could not be produced for the entire area of the object. However, portions of the object overlap during adjacent time steps, thus allowing evaluation of the polarization states. To ensure these portions are compared, a motion mask was developed and implemented to remove areas of non-overlap. Figures 5.53 and 5.54 depict the original and masked polarized images of the moving in-scene object. The masks were created by retaining radiance values greater than or equal to 45% of max value in each image. There are four images representing the four polarization states.


(A) 0 Degree Polar Image


(B) 0 Degree Masked Polar Image


(C) 45 Degree Polar Image


(D) 45 Degree Masked Polar Image

FIGURE 5.53: 0 and 45 Degree Original and Masked Polar Image


(A) 90 Degree Polar Image


(B) 90 Degree Masked Polar Image

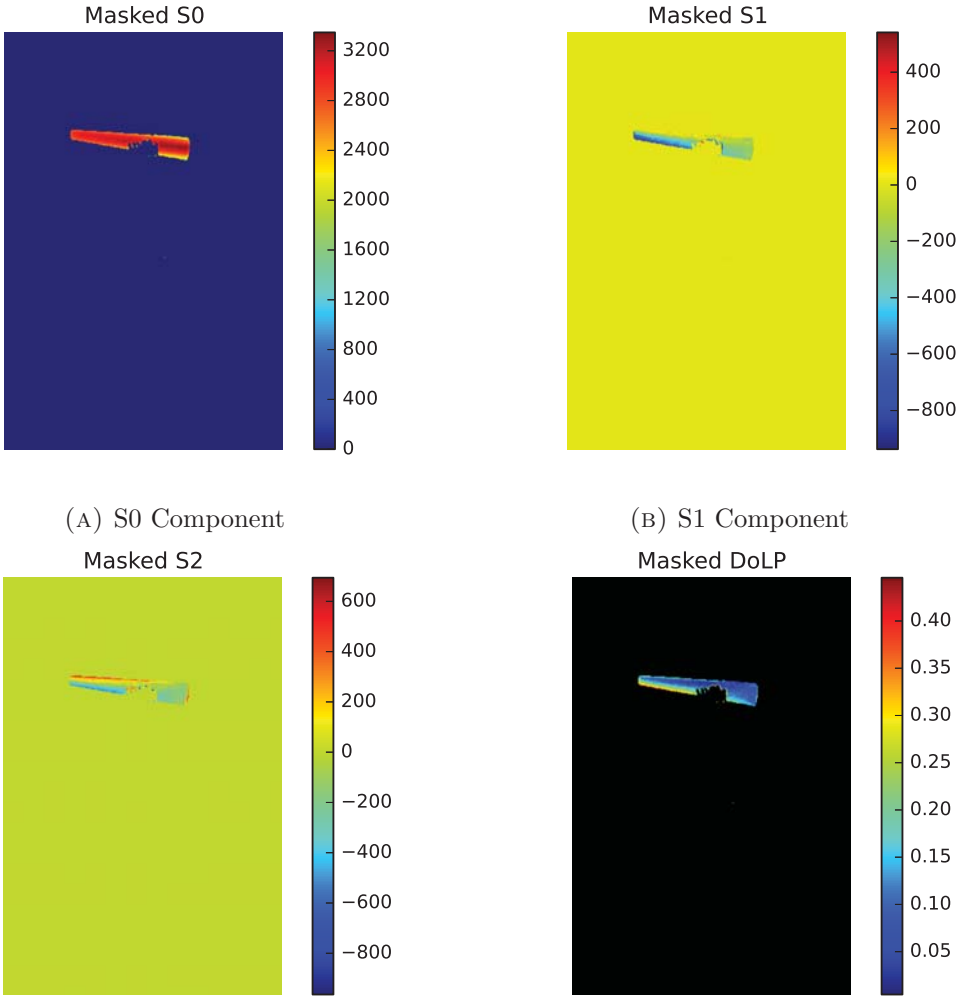

(C) 135 Degree Polar Image


(D) 135 Degree Masked Polar Image

FIGURE 5.54: 90 and 135 Degree Original and Masked Polar Image

#### 5.7.2.3 Moving In-Scene Stokes Vector

As stated in the previous section, due to the movement of the object within the scene, it is not possible to develop a Stokes vector for each unique position depicted in the individual images. Each of the previous masks were multiplied together to produce a single mask covering the extent of the four frames under consideration. This mask was then applied to each image and the remaining images were used to form the Stokes vector described in Section 5.5.3. Figure 5.55 depicts the S0, S1, S2, and DoLP results respectively of the moving in-scene object.




(A) S0 Component

(B) S1 Component

(C) S2 Component

(D) Degree of Linear Polarization

FIGURE 5.55: Polarimetric Stationary In-Scene Results of Object

#### 5.7.2.4 Track Association Between Sensors

Since the polarimetric data is not spatially registered to the data as the other modalities, a single frame manual comparison is done to associate the high DoLP object in the MAPPS sequence, with the person holding the object in the GoPro sequence. Once this match is made, the person in the GoPro data can be watched for further analysis.

A method of automatically matching people in spatially unregistered data could involve correlating track data between the imagers. In order to compare tracking data, a tracking algorithm would need to be run on each imager separately. Only the objects depicting a higher-than-background DoLP would be retained in the polarimetric imagery, whereas all people and objects would be retained in the GoPro imagery. Following this, the track data would need to be normalized by the size of the imager to place them in a common spatial basis. After normalization, a two-dimensional correlation could be used to compare the track locations of the object in the polarimetric imagery to all the objects in the GoPro imagery. Due to the narrow FOV and the positioning of the MAPPS sensor, each of its images is a subset of the much larger GoPro image, as depicted in Figure 4.21. Thus, it follows that every moving object in the MAPPS image is also in the GoPro imagery. The final step would be to associate the object or person with the highest correlation between the track data. The data degradations and likelihood of detection will be discussed in the results section.

# Chapter 6

# Results

Two activities were selected for evaluation within this research. The first was an object exchange activity, involving two individuals passing one another and exchanging a briefcase-like object. After the exchange each individual continued along their original paths. This analysis concentrated on the spectral nature of the briefcase object and the individuals that exchanged this object.

The second was a simulated RPG activity which followed the steps of an individual walking to a field and raising a PVC pipe onto their shoulder. This activity was accomplished by analyzing the polarimetric characteristics of a PVC pipe as it moves throughout the scene.

## 6.1 Object Exchange

In this research, an object was exchanged between two individuals walking in the scene. A spectral signature was calculated for each person walking in the scene and a spectral angle was calculated for the baseline signature and the signatures of each frame thereafter. Figure 6.1 depicts the output of the spectral angles of each person for each frame and It is interesting to note that if these data were evaluated alone, a case could be made that Person 2 and Person 3 must have been the two exchanging the object. This is due to the abrupt drop in spectral angle from Person 2 at the exact point of joint possession and the variation in spectral angle of Person 3 after the object exchange has occurred.

The drop in spectral angle for Person 2 is actually completely coincidental. As Person 2 was walking into position the PVC was occluded by their body (frames 250-400), but as the exchange began to occur, the PVC pipe returned from occlusion as remained unobstructed throughout the remaining section of the video sequence. By strictly making a decision based on these figures, that information would have been lost and an inaccurate analysis developed. It was by filtering the people by their spatial distance that allowed Figure 6.1 to be reduced to the people involved in the object exchange.

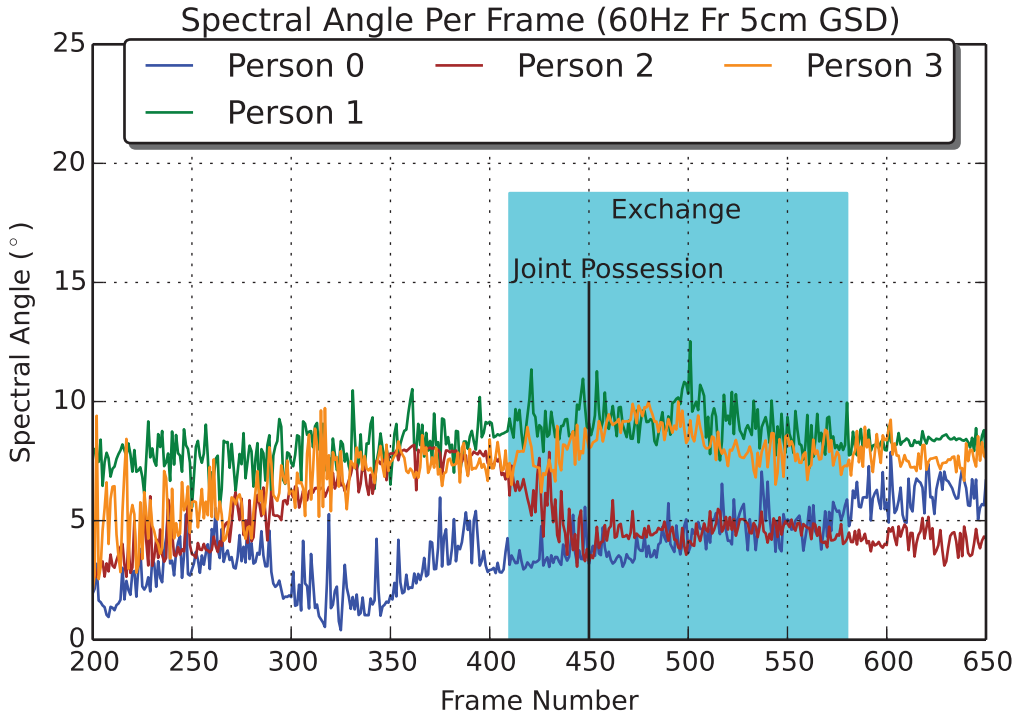



FIGURE 6.1: Spectral Angle of All Filtered People

#### 6.1.0.5 Filter People by Distance

Figure 6.2 shows the angles of the two individuals involved in the exchange. This distance was empirically determined to be 30 pixels in the lateral direction and 15 pixels in the longitudinal over several frames. The number of frames in dependent on the frame rate of the data being evaluated. A one second period of data was determined to be adequate. Therefore, the people should be within the elliptical bounds for at least one second to be included in the object exchange. For example, at 60Hz they should be within the

elliptical distance for 60 consecutive frames; at 20Hz they should be within the elliptical distance for 20 consecutive frames.

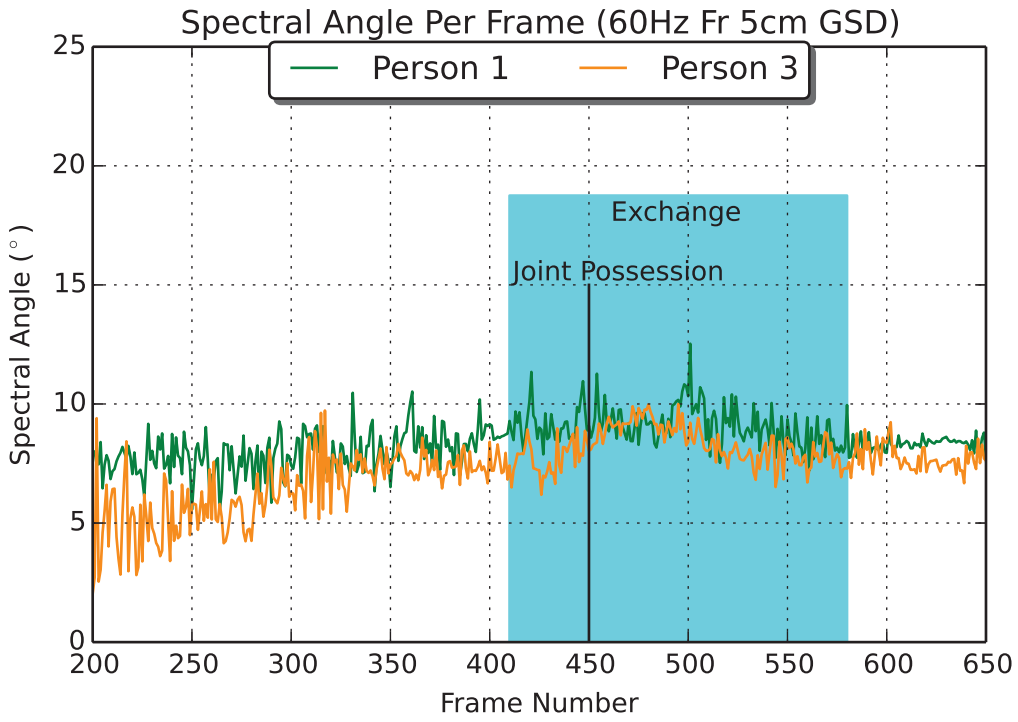



FIGURE 6.2: Spectral Angle of Spatially Filtered People

#### 6.1.0.6 Threshold Analysis

Figure 6.3 highlights the pre-exchange portion of the video sequence. By taking this mean angle the value indicative of an exchange can be calculated as

$$\begin{aligned} \theta_{Post-exchange\ mean} &\geq 1.1 \cdot 7.96 \\ &\geq 8.36 \end{aligned}$$

where $\theta$ represents the spectral angle of the data. Figure 6.4 highlights the post exchange frames and depicts the mean spectral angle. A value of 8.968 degrees was determined to be the post exchange mean spectral angle. This is roughly an 18% difference in mean spectral angle before and after the exchange.

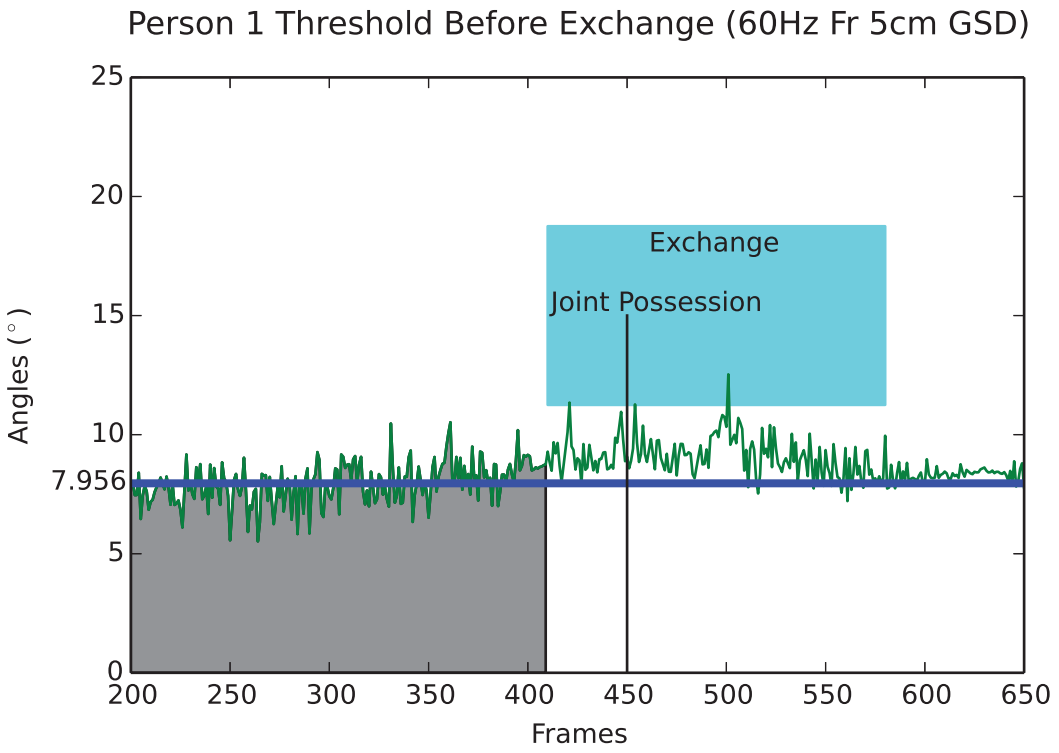



FIGURE 6.3: Person 1 Threshold Spectral Angle Before Exchange

After confirming an exchange has occurred, via the criteria stated above, the post exchange mean spectral angle is taken as the post exchange spectral angle indicative of an object exchange. This value will be used as the threshold for developing a "likelihood of detection" in the spatially and temporally degraded data. It should be noted that before the mean was used a comparison of the standard deviations was performed. The mean of the data before the exchange was 7.856 with a standard deviation of 0.8682. After the exchange, the mean was 8.968 with a standard deviation of 0.8243. The change in standard deviation after the exchange represents a 5.32% difference from the standard deviation before the exchange. The almost identical values of the standard deviations did not afford enough of a change to be considered useful for evaluation.

Throughout this section there have been a several thresholds and ad hoc restrictions used to evaluate the data at hand. However, it important to understand that this research is designed to develop a performance assessment methodology capable of characterizing the utility of a particular system given a specific method of detecting an activity. Since there existed no objective activity analysis methodology, notional activity and analysis

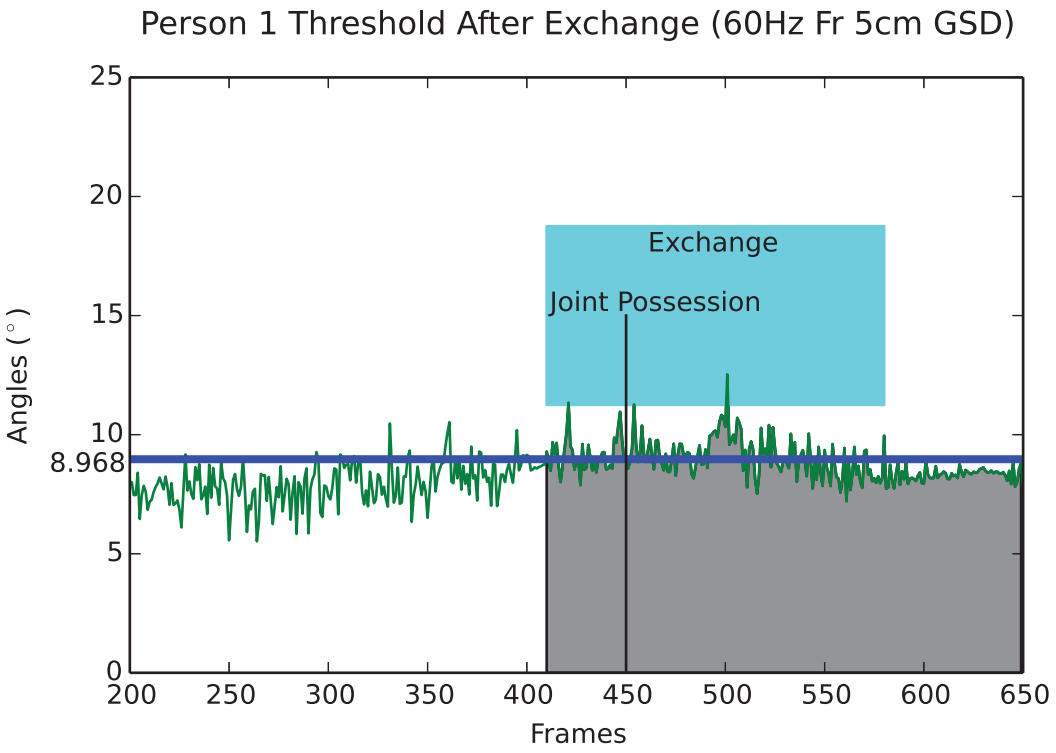



FIGURE 6.4: Person 1 Threshold Spectral Angle After Exchange

schema were developed to depict this point.

**Assessing the Noise within the Data** As each of the individuals moves throughout the scene, it is expected that their spectral angle will fluctuate around some mean value. It is natural to have this measure of variability within the data because the object never repeats the exact orientation, perspective, or sun-target-sensor geometry within the sequence. If this variation is considered noise, then it becomes possible to estimate the SNR using a statistical analysis defined by

$$SNR = \frac{\mu}{\sigma} \tag{6.1}$$

where this Signal-to-Noise Ratio is the ratio of the mean ($\mu$) of the signal over the standard deviation ($\sigma$) of the signal. Using the values depicted above and those of the dataset as a whole, Table 6.1 depicts the signal, noise, and SNR calculations of the data for Person 1.

TABLE 6.1: Signal-to-Noise of Spectral Angle Data

| Portion of Dataset | Mean (degrees) | Standard Deviation (degrees) | SNR |
|---|---|---|---|
| Before Exchange | 7.856 | 0.8682 | 9.049 |
| After Exchange | 8.968 | 0.8243 | 10.88 |
| Before & After Combined | 8.412 | 2.124 | 3.961 |

As this table indicates, the SNR of the separate datasets is greater than 9.0, but drops closer to 4.0 when both sets are merged. The combined SNR will continue to decrease as the spectral angle before and after the exchange increases. The converse of this statement provides an interesting and possibly useful method of evaluating the presence of an object exchange. In order for the combined SNR to increase to the levels of the before and after SNRs, requires that the angular mean remains relatively constant throughout the video sequence. Assuming the standard deviation will remain the same, then the low combined SNR can be associated with the increase in standard deviation of the data. This increase in standard deviation is directly related to the shift in data after the exchange. Thus, as the angular disparity describing the exchange reduces, it stands to reason that the combined SNR will increase. This low SNR presents another method of determining the existence of an exchange in the dataset.

#### 6.1.0.7 Alternate Methods of Assessing Spectral Angle Data

During the development of this activity recognition technique, the author noted that the data could have been evaluated in several different methods This section is included to briefly state each of those methods for evaluation in future research.

**Method of Proportions** In a real world situation the exact time of an exchange may not be known a priori. One option for detecting this point would be to compare x percentage of the first portion of the data, to 1-x percentage of the latter part of the data. This would be done in an iterative method whereby the first 10% and the latter 90% would be compared, then 20% to 80%, 30% to 70% and so forth.

**Method of Angular Difference** In the research above, it was decided that the basis for an object exchange will be the mean spectral angle after an exchange has

occurred. Another method for evaluating the data is to use the difference in spectral angle before and after the exchange had occurred. Utilizing this method of evaluation, would afford researchers a method of evaluating how the angular difference changes as the spatio-temporal degradations occur.

**Method of Sliding Window** Another option would be to create a sliding window that compares the current frame to all prior frames in the window. Figure 6.5 depicts a notional outcome of this type analysis. For example, using a window size of 20 frames, the current frame in the analysis is compared to the prior 19 and a relative difference can be annotated. Normalizing by the maximum difference would point to the frame where the maximum change in spectral mean is located. Note that this figure is simply a Gaussian distribution depicting an ideal example of this sliding window concept.

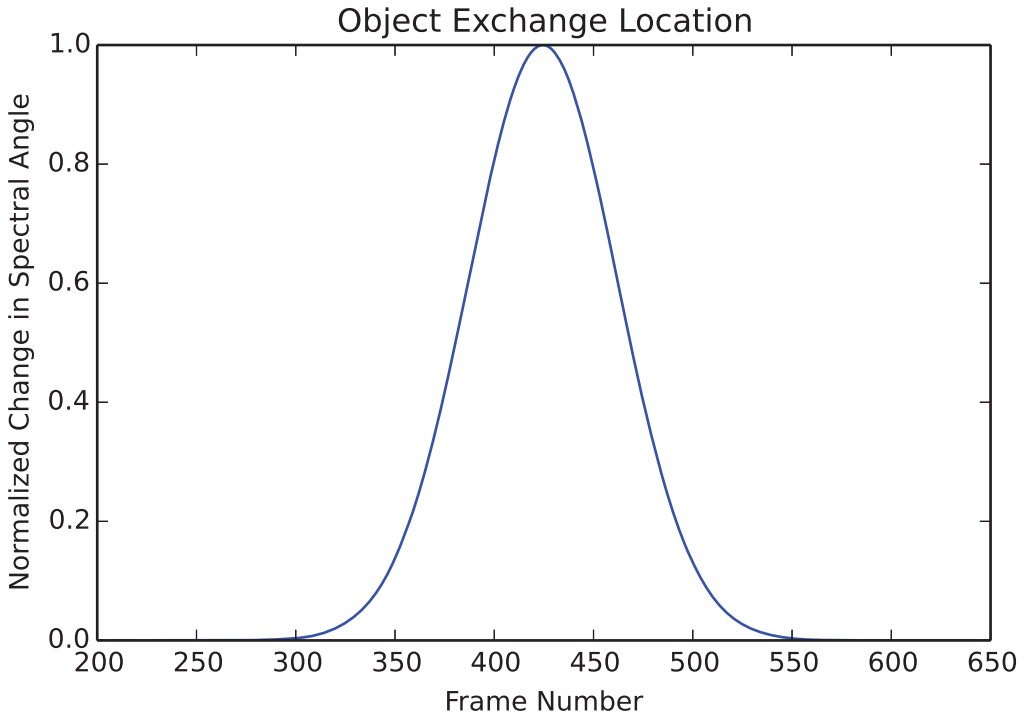



FIGURE 6.5: Sliding Analysis of Spectral Means

**Method of Standard Deviations** Another option for assessing the spectral angle of the data would be to perform a comparison of the standard deviations before and after the exchange. The mean of the data before the exchange was 7.856 with a standard deviation of 0.8682. After the exchange the mean was 8.968 with a standard deviation

of 0.8243. The change in standard deviations after the exchange represents a 5.32% difference from the standard deviation before the exchange. This minor change in standard deviations is within the noise of the data and thus not considered significant.

### 6.1.1 Spatial Analysis

Figure 6.6 depicts the spectral angles of the participants as the data is spatially degraded. By normalizing the spatial degradations a likelihood of detecting the exchange can be developed. Figure 6.7 depicts this likelihood of detecting as a function of GRD. Figures 6.8 and 6.9 filter the spatial degradation data by only retaining the two individuals involved in the object exchange.

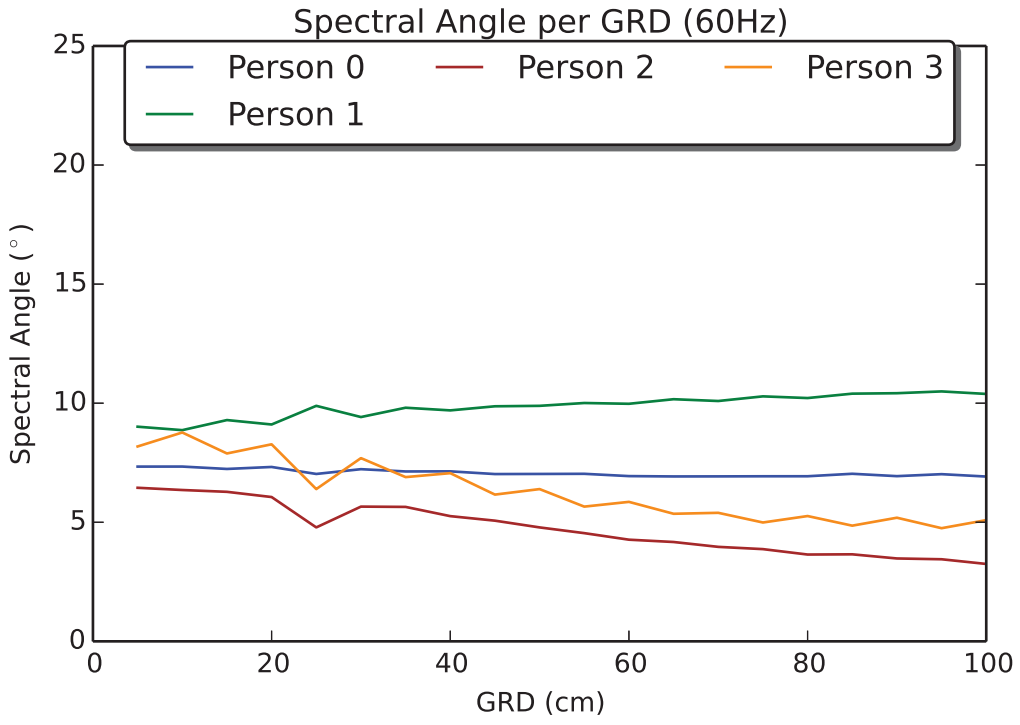



FIGURE 6.6: Spectral Angle per GRD (60Hz)

As the data is spatially degraded, the foreground data becomes more similar to the background data. This has an overall effect of decreasing each participants spectral angle. As the uniqueness of the foreground data is reduced, it becomes more difficult to identify the exchange of a small object from one person to another. In this particular data set, these spatial degradations have caused some of the detection likelihoods to

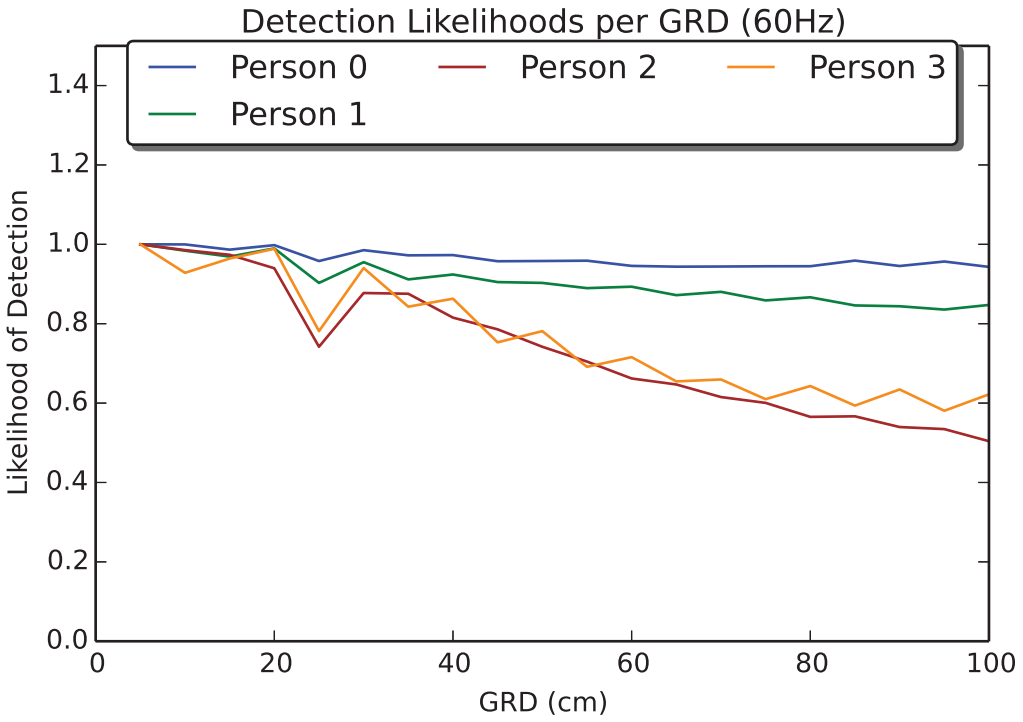



FIGURE 6.7: Detection Likelihood per GRD (60Hz)

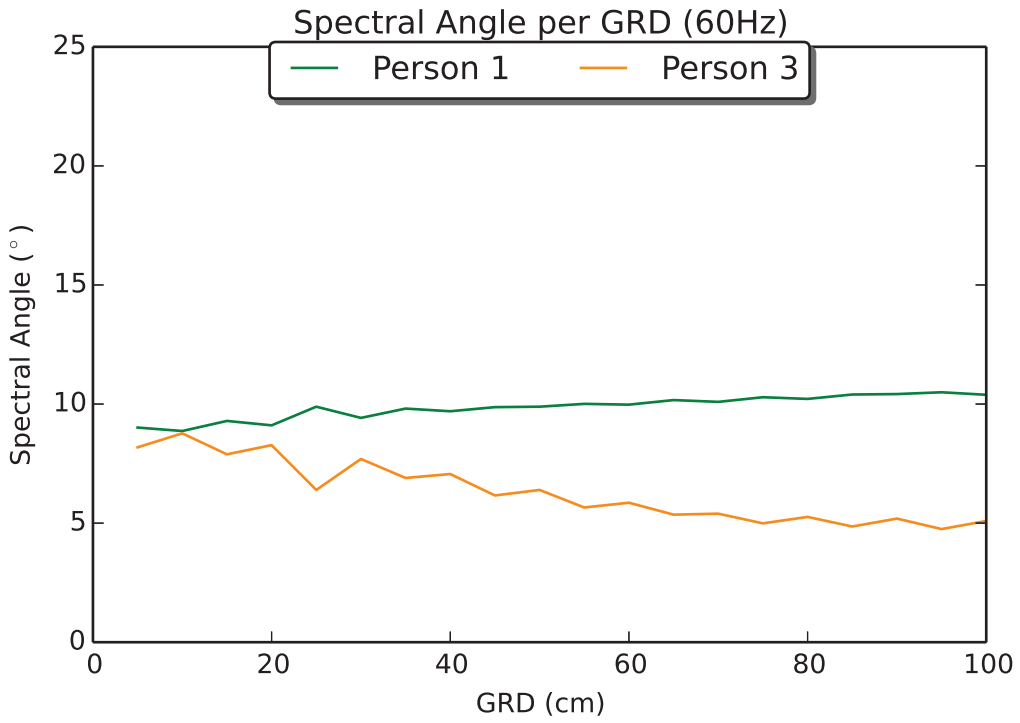



FIGURE 6.8: Spectral Angle per GRD (60Hz) of Individuals in Object Exchange

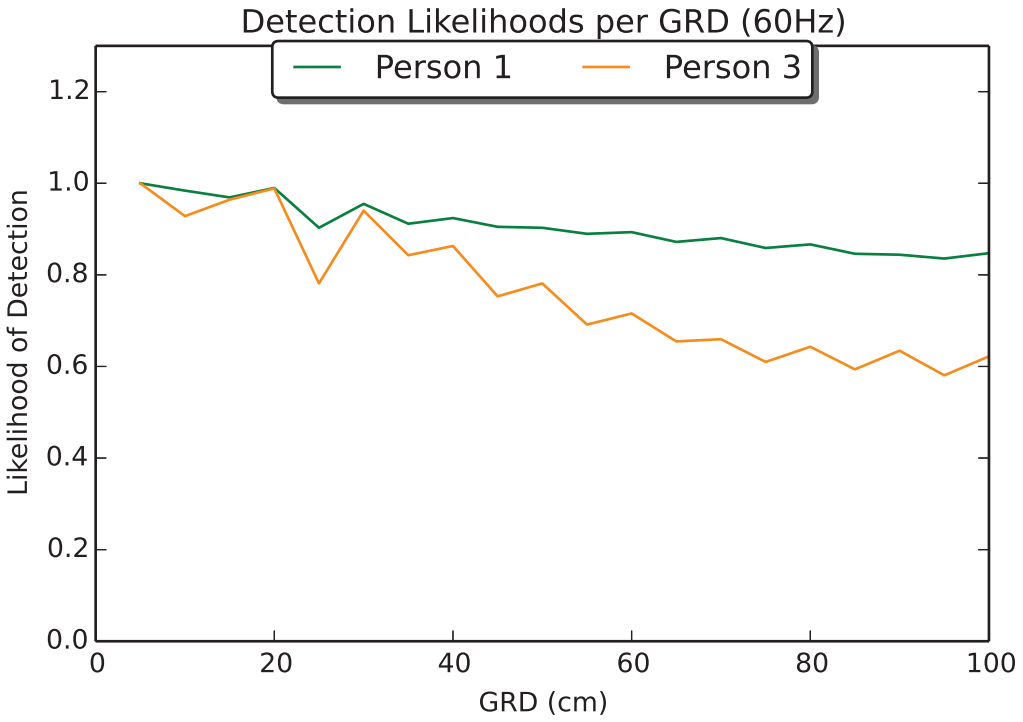



FIGURE 6.9: Detection Likelihood per GRD (60Hz) of Individuals in Object Exchange

decrease by over 40%, as seen in Figure 6.7. The differences in the drastic nature of the decrease can likely be attributed to the unique spatial characteristics of each of the people involved. Person 2, carrying the PVC pipe, is noted to have the pipe move in and out of occlusion through the scenario as the participant moves into position. Once there, the pipe is moved from a vertical position hanging down at the waist to a horizontal position on top of the shoulder. This drastic change in positioning, coupled with the occlusions throughout the movement, have left its spectral signal changing quite drastically throughout the dataset. Person 3 also depicted a high change in likelihood of detection as the GRD increased. This may be attributed to the size of the person and their gait. Person 3 had a thin stature and long stride, as seen in Figure 5.47a, as they moved throughout the scenario. The long length of the stride increased the width of the bounding box, which meant more non-people pixels would be included in the spectral signature of the individual. This individual was also moving through the busiest portions of the scene, affording more opportunities to be around other people, shadows, and foreground objects. This likely differs from Person 1, also involved in the exchange, due to the difference in their spatial extent. Person 1 is wearing a book bag through the

scenario, thus extending their spatial coverage within the scene. This extended coverage affords a more stable and unique spectral signature and reduces the remaining space within the bounding box that could be taken by unwanted data. As previously stated, the large size of the bounding box allows the entire person to be captured at each frame, but also allows adjacent foreground clutter to be included as well. All individuals in the experiment experienced a sharp decrease in likelihood of detection as a result the increase ordecrease in spectral angle after the 20cm degradation. Further evaluation of the data is needed to determine the exact cause of the simultaneous decrease in spectral angles for persons 0, 2, and 3, and an increase in Person 1.

### 6.1.2 Temporal Analysis

The overall effect of the temporal degradation was an increase in the spectral angular difference over the course of the video sequence. As the number of frames is reduced in the sequence, the number of frames included in the spectral baseline of each person is also reduced. This stipulates that each reference signature is more likely to reflect a frame-unique signature of the person rather than a time-averaged signature. As the person moves throughout the scene, their sun-target-sensor geometry changed, producing a different radiance at the aperture. Along with the change in sun-target-sensor geometry, the perspective of each individual and orientation of their clothes changed in each successive frame. By averaging more frames, the effect that each of these factors had on the baseline signature was reduced and thus less important overall. However, as the number of frames incorporated into the baseline signature was reduced, each of these effects became more prominent.

Figures 6.10 and 6.12, respectively, depict the spectral angle as a function of frame rate for all the participants and those engaged in the object exchange scenario. Figures 6.11 and 6.13, respectively, depict the detection likelihood graphs for all the participants and those engaged in the object exchange scenario.

It was expected that the overall effect of reducing the frame rate would be an increase in the angular difference and a decrease in the likelihood of detection. Figures 6.10 through 6.13 only show frame rates down to 1Hz, because degradations beyond that point resulted in drastic decreases in the likelihood of detection. This is due to the

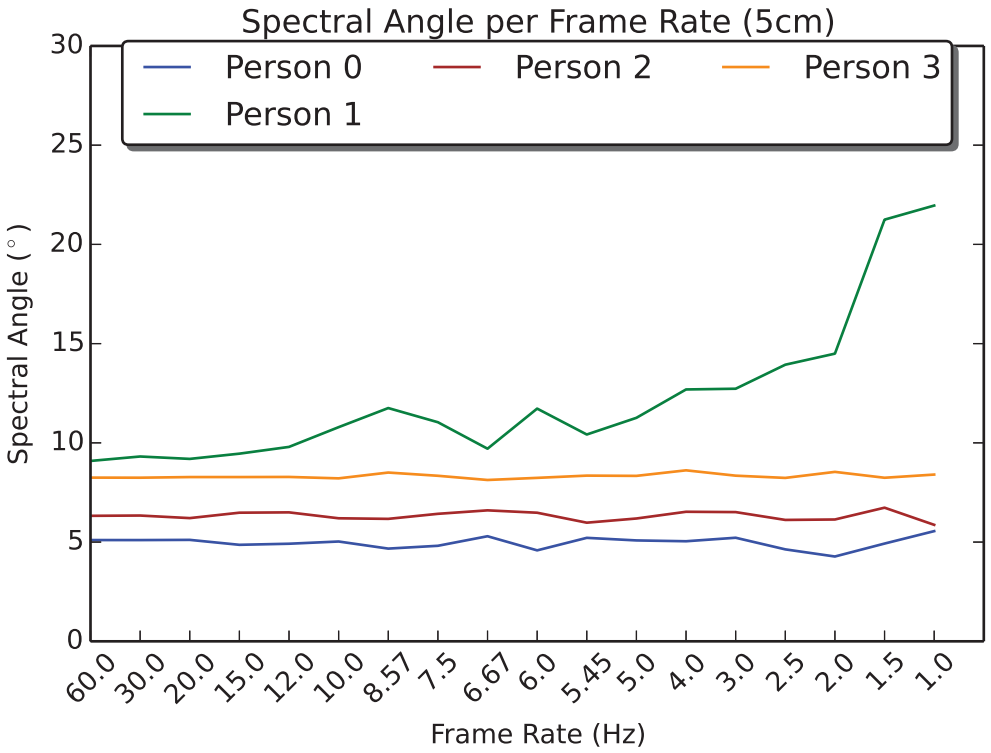



FIGURE 6.10: Spectral Angle per GRD (5cm)

limited number (single digit) of frames included in the spectral baseline. Reviewing the degradations presented, it can be seen that Person 1 has a prominent change in spectral angle from 15Hz down to 1Hz. This is likely due to the large spatial extent of Person 1 relative to the other individuals in the scene. Person 1 was wearing a book bag which increased their spatial extent throughout the collection. The book bag was a non-rigid object attached in a non-rigid manner that allowed the object to freely move on the person's back. This movement allowed it to change perspective and orientation both with, and independent of, the person carrying the bag. A quick visual inspection of the bag, during the capture, showed that it was made of a material with a higher reflectance than the clothes Person 1 was wearing. This reflective property appeared more specular than Lambertian indicating that a change in sun-target-sensor geometry would provide large differences in the at-aperture radiance values. For these reasons, decreasing the number of frames in the baseline signature provided, significant changes in the spectral angle derived for each frame rate. This in turn led to drastically decreased likelihood of detection. It is noted that while the PVC pipe is also highly specular in nature, it

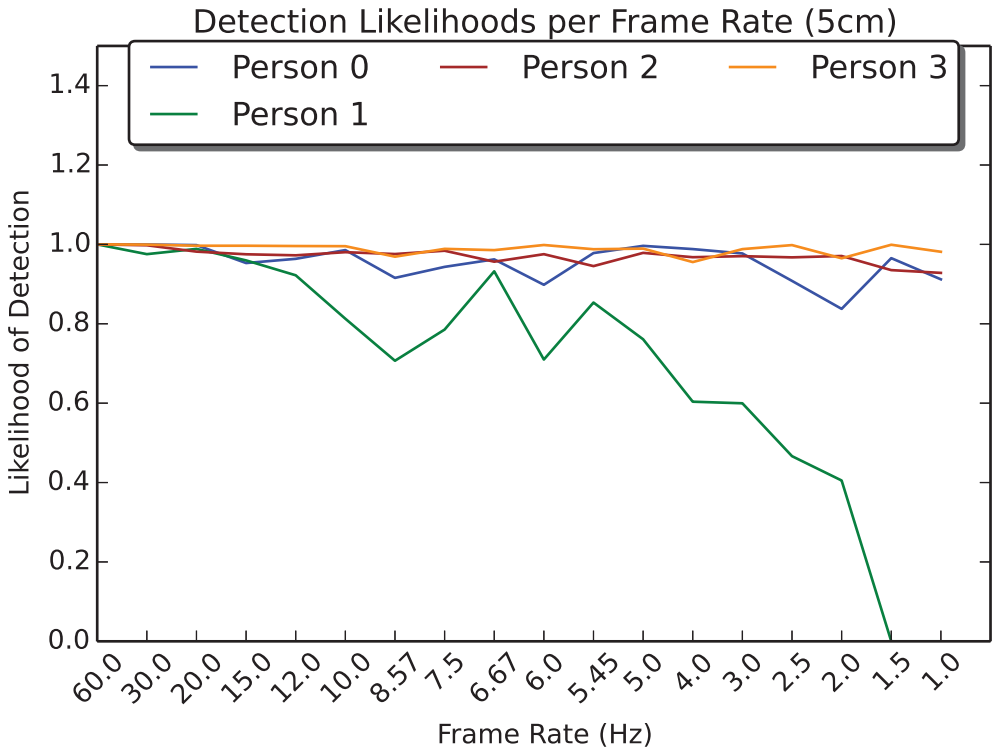



FIGURE 6.11: Likelihood of Detection per Frame Rate (5cm)

differs in that it is a cylindrical object with a reflectivity close to one. Thus regardless of orientation, there will always be a strong reflection coming back toward the sensor.

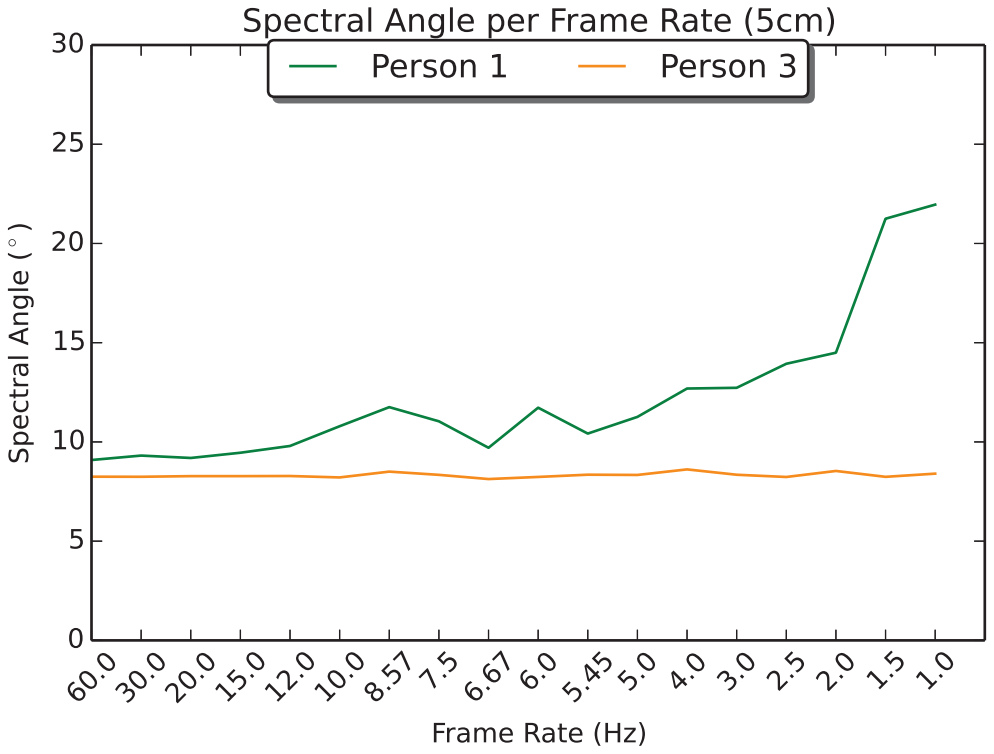



FIGURE 6.12: Spectral Angle per Frame Rate (5cm)

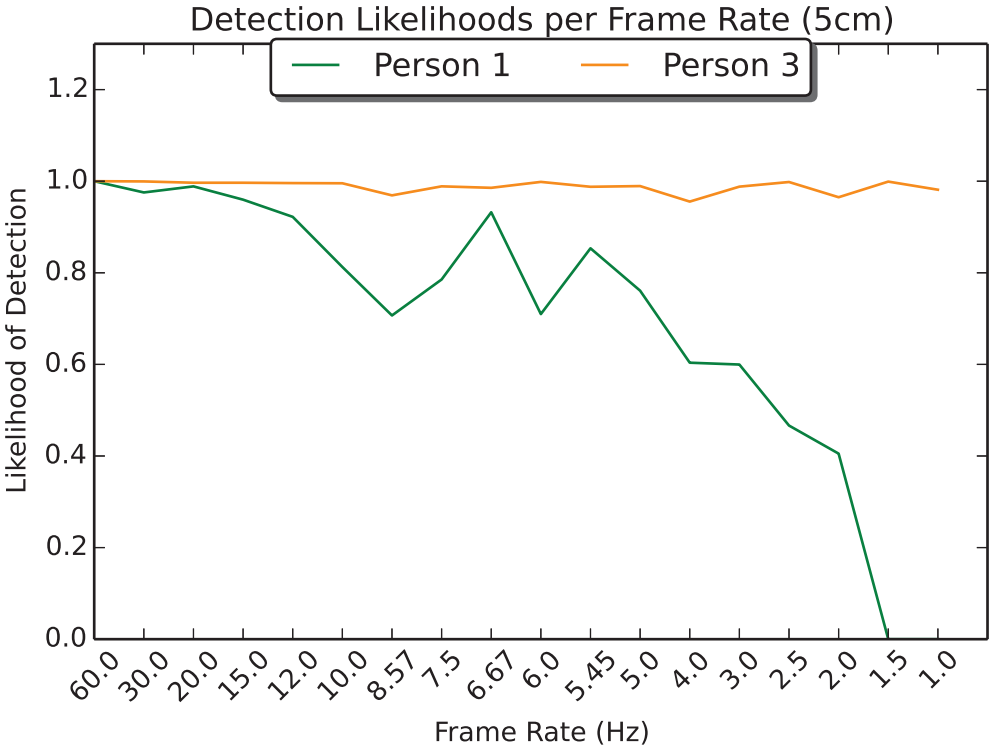



FIGURE 6.13: Likelihood of Detection per Frame Rate (5cm)

### 6.1.3 Likelihood Surface

Figures 6.15 through 6.17 depict the likelihood of detection surfaces for each of the people within the scene. As only Person 1 and Person 3 were involved in the object exchange, the detection surfaces are only valid for these two. However, the remaining two surfaces are included so that general trends of the spatio-temporal degradations can be evaluated.

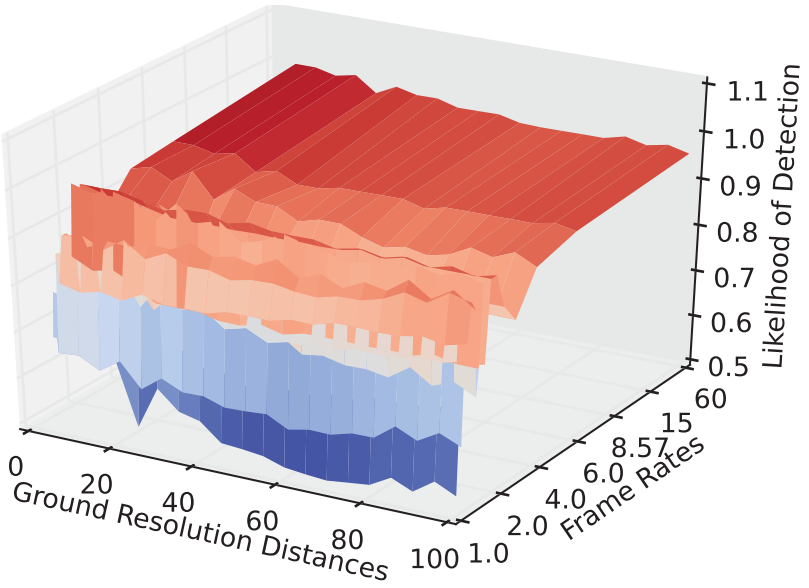



FIGURE 6.14: Likelihood Surface - Person 0 (No activity)

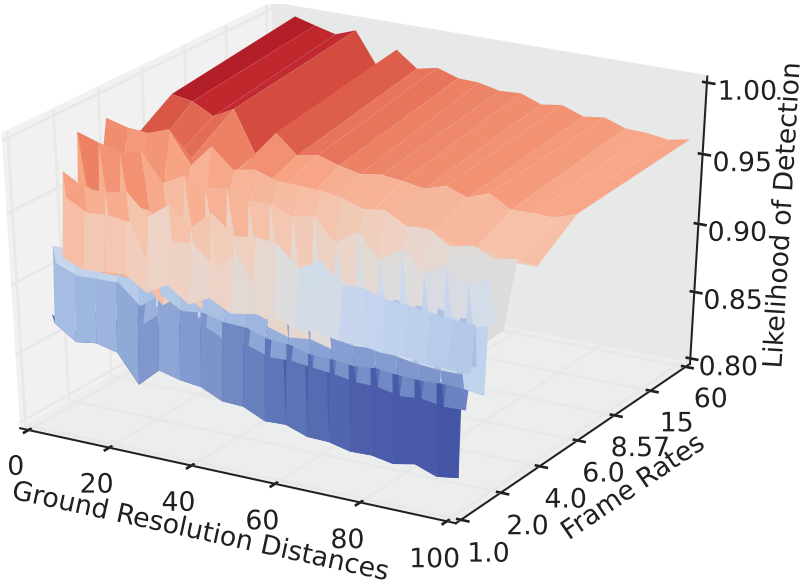



FIGURE 6.15: Likelihood Surface - Person 1 (Object Exchange)

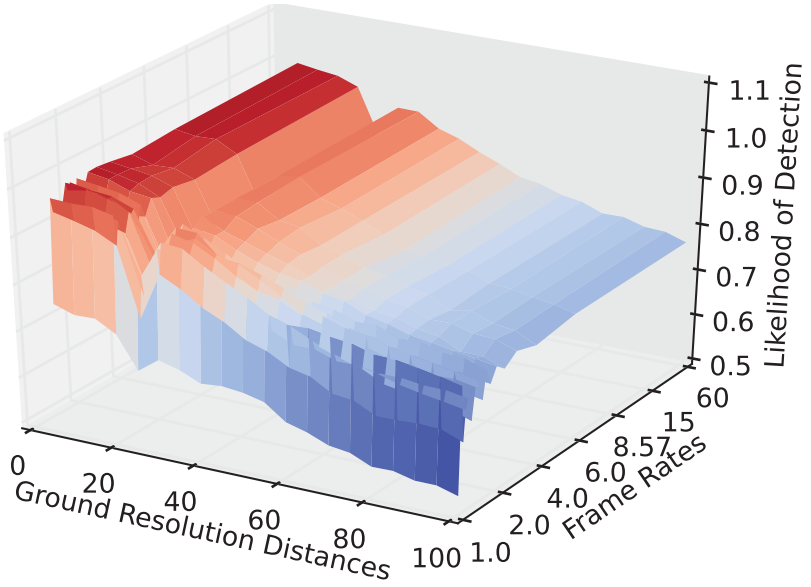



FIGURE 6.16: Likelihood Surface - Person 2 (PVC Pipe)

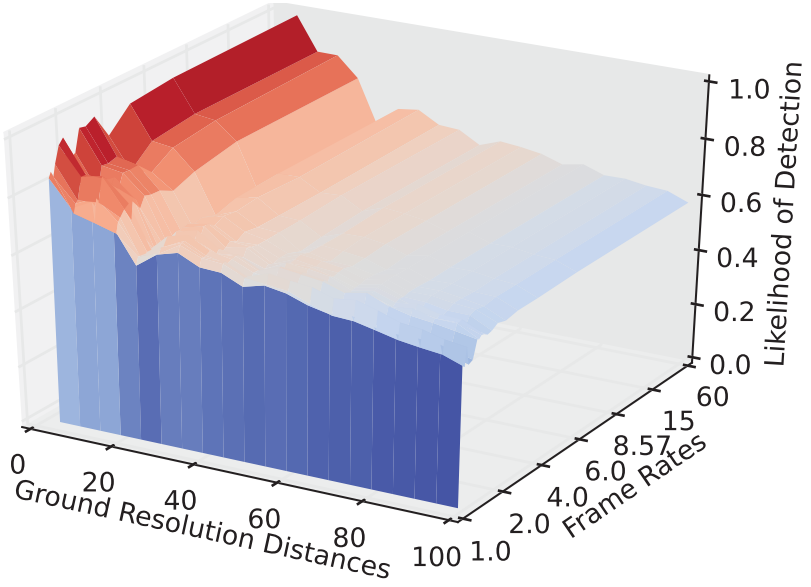



FIGURE 6.17: Likelihood Surface - Person 3 (Object Exchange)

For half of the participants, the spatial degradations up to 100cm were not very impactful, while for the others they were. All the participates maintained the sharp decline in detection likelihood around the 20cm spatial resolution that was discussed earlier. All participant surfaces also experienced the temporal cliff that occurs at at frame rates of 1Hz and below. Comparing the temporal and spatial degradations, it would appear that the impact of GRD can be quite large, whereas all participant surfaces experienced a temporal decline which reached a point of futility.

Person 0 depicted an almost negligible change in detection likelihood over all the spatial and temporal resolutions. This means the spectral signature of the person did not change much throughout the sequence. This makes a bit of sense, as this individual walked around in a small circle for most of the video sequence, neither carrying anything nor exchanging objects with others. Person 1 depicted a 5% decrease in detection likelihoods over the spatial degradation and sloping decrease in likelihood over the temporal degradations out to the cliff at 1Hz. This states that it is possible to detect an object exchange over a wide range of spatio-temporal resolutions with a high likelihood of detection. Person 2 exhibited greater impacts from the spatial degradations than the temporal degradations, likely due to the object occlusions from the moving PVC pipe. Even these only brought the likelihood of detection down by about 20% along the spatial extent. Person 3's likelihood surface had an opposite trend of the prior three surfaces in that the temporal degradation caused more loss of likelihood than the spatial degradation. Person 3 was the other individual engaged in the object exchange. This individual appeared have been impacted the most by GRD, with a decrease in likelihood of 40% over the course of the degradations. As stated in the spatial section above, this could be due to the small stature of the individual relative to the space designated by the bounding box.

Since two people were involved in the object exchange, we have two detection surfaces worth of data for this activity. The case could be made that you only need one surface to represent this activity, since there was in-fact only one activity occurring. If that were the case and each surface indicates an independent likelihood of detecting the exchange, then either surface could be considered a valid representation of detecting the object exchange event. Of course, the surfaces are not truly independent because they are partially derived from the same spectral data that represents the object. However, for the

purpose of this discussion the independence of the surfaces allows one to choose a surface to represent the object exchange activity in a future ABI lookup table.. That being the said, the surface associated with Person 1 with the overall likelihood of detection greater than 90% will be surface to represent this exchange.

## 6.2 Polarimetric Tipping and Cueing

Sections 5.7.2 and 5.7.2.4 depicted the steps necessary to accurately identify an object with a high DoLP and associate tracks between disparate imagers. This section serves to combine the two concepts by showing it is possible to tip the polarimetric sensor to an object of interest within the scene, and then cue an adjacent sensor to follow that object. Figure 6.18 shows the MAPPS frame in which a high DoLP was detected in the sequence. Figure 6.19 depicts the DoLP image of this frame and Figure 6.20 depicts a close-up of the region with a high DoLP. Notice that there are points greater than one in this image. That is due to the motion induced by the person moving the object, which is why the motion masks are necessary. This motion-induced false DoLP "signature" is seen throughout the scene in the leaves and people moving between images. One of the in-scene fiducials is also seen to have a moderately high DoLP. However, since this fiducial does not move, it was masked out with the other background data.

After masking the imagery to remove non-overlapping pixels, Figure 6.21 remains. This is the same image that was depicted in Section 5.7.2.3. This particular object had a value as high as 0.4 with a preponderance of the data hovering closer to 0.1. This is compared to an environment with an average DoLP no higher than 0.05, as depicted in Figure 6.19. After a high DoLP object was been detected, a tag was placed over the object within the polarimetric image. Using a manual association, the GoPro imager is the cued to track the person holding the object using its wider FOV lens. Figures 6.22 and 6.23 depict the tipping and cueing in the polarimetric and RGB data respectively. Note that each tracking system had its own numbering scheme, which is why the DoLP text is over Person 2 in the GoPro image and Person 0 in the MAPPS image.

FIGURE 6.18: First frame in DoLP Sequence

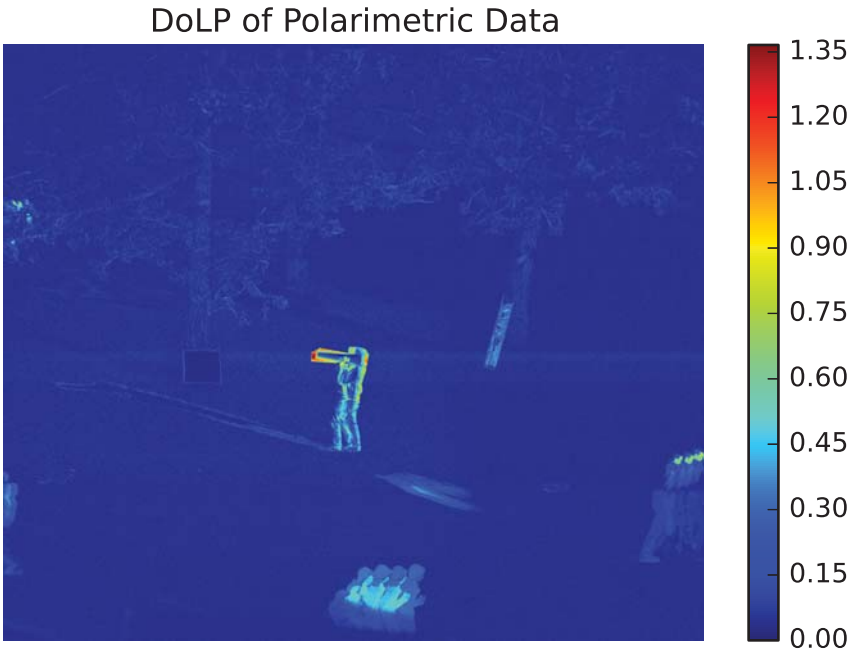



FIGURE 6.19: Full DoLP Image

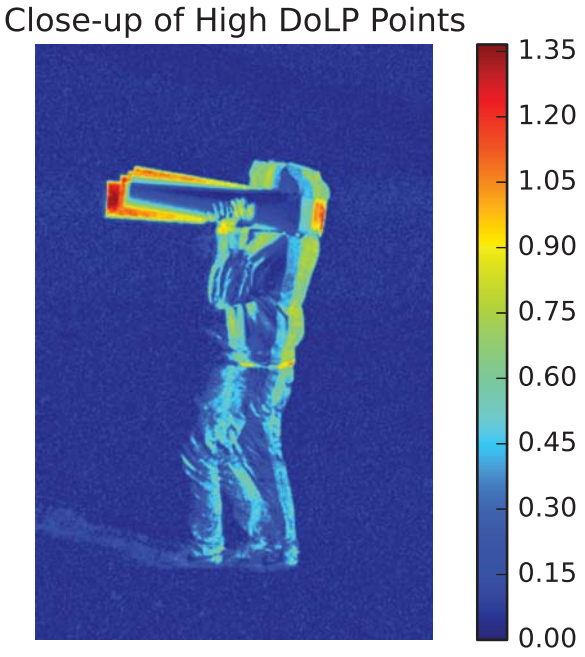



FIGURE 6.20: Close-up of High DoLP Region

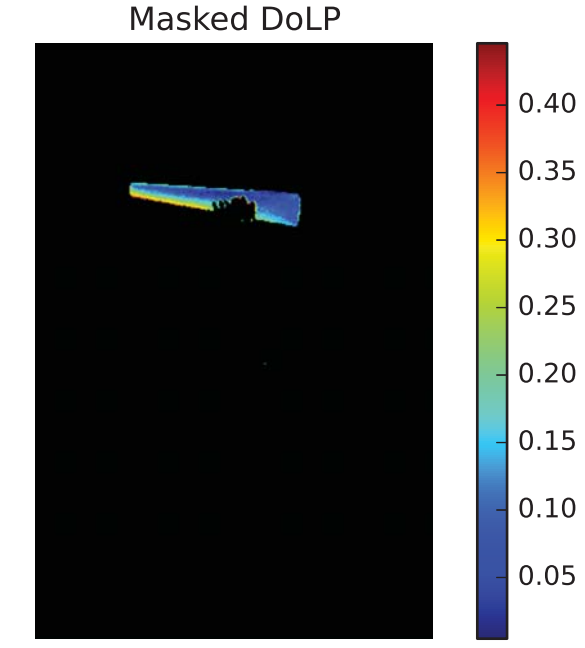



FIGURE 6.21: Masked Close-up of High DoLP Region

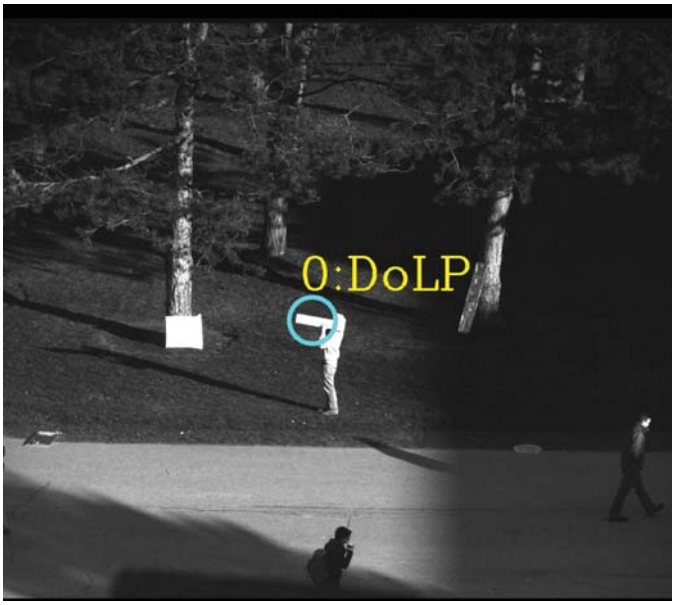



FIGURE 6.22: Polarimetric Tip in MAPPS Imagery

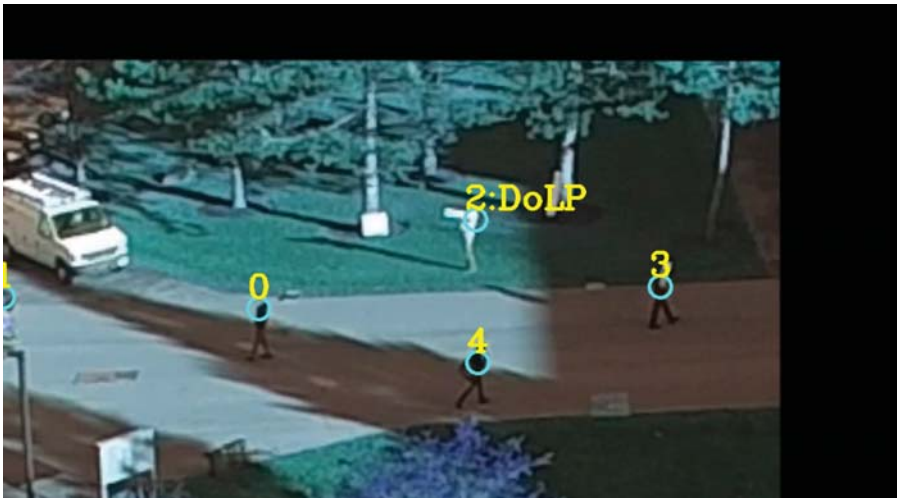



FIGURE 6.23: GoPro Imagery with DoLP Cue

### 6.2.1 Polarimetric Data Degradations and Likelihood of Detection

Two primary differences exist between the polarimetric data in this data set and the associated spectral data. First, in order to perform DoLP evaluations, four sequential polarimetric images are necessary. As described in Section 4.2.2, MAPPS has a unique spinning wheel configuration, that cycles through the polarimetric filters in the following order: 0, 45, 90, 135, 135, 90, 45, 0, 0, 45, etc. At a 6Hz configuration the 0 degree image is captured, then recaptured 1.167 seconds later, then recaptured 0.167 seconds later. That means any interpolation would have to be done between these intervals, as it would not be correct to interpolate between differing polarimetric images. The low frame rate and odd temporal filter wheel configuration of MAPPS did not easily allow for temporal degradation of this data. Thus, this step is left to future researchers. A likelihood of detection is determined by the ability to detect a polarimetric object within the given scene. Since no spatio-temporal degradations were performed on this data, and a high DoLP object was properly detected, a detection likelihood of 1.0 is assigned to this scenario.

## 6.3 Summary

Two activity recognition techniques were successfully implemented in this research. The first was able to detect an object exchange that occurred within the dataset. As the data were spatially and temporally degraded the likelihood of detecting the exchange decreased. The temporal degradations provided only gradual decreases as the frame rate was decreased from 60Hz to 1Hz. At 1Hz and lower there was a drastic drop in the likelihood of detecting the activity. For two of the people the spatial degradations provided a 5% reduction in likelihood of detection, whereas the other two resulted in 20% to 40% reductions in likelihoods. This was attributed to the spatial extents of each person and the stride of their gait.

The second technique involved identifying a simulated RPG activity and using that data in a tipping and cueing scenario. The simulated RPG activity was successfully identified by detecting a high DoLP from the PVC pipe. This information was then used to cue the GoPro imager to track the person holding the pipe outside the FOV of MAPPS. No

temporal or spatial degradations were performed on this dataset due to the low frame rate and odd filter wheel configuration of the MAPPS imager.

# Chapter 7

# Conclusion

## 7.1 Problem Statement and Research Objectives

Two questions drove this research: Is it possible to utilize a series of multimodal sensors in a semi- or fully- automated fashion to develop intelligence based on the activities within a given scene? If so, could an objective performance assessment be developed to determine if a sensor is capable of detecting specific AoIs in motion imagery? Based on the work in this research, the answer to both of these questions is yes.

To address the first question, two AoIs were analyzed. First, an object exchange AoI was imaged by a series of multispectral sensors. SAM was used to automatically determine if an exchange had occurred in the motion imagery dataset. The second AoI was a simulated RPG activity imaged by a polarimetric and RGB sensor. By evaluating the polarimetric data, it was possible to detect the simulated RPG by identifying its high DoLP signature, relative to the background. Once detected in the narrow FOV polarimetric imager, an algorithm cued the wide FOV GoPro imager to continue tracking the object across the scene.

The second question was related to developing an objective performance assessment methodology. Two reasons were cited for developing this methodology: Assessing the requirements for developing tomorrow's imaging platforms and assessing the performance of current platforms in detecting specific AoIs. As mentioned in the introduction, both the military and commercial sector have been continually improving the spatial and

temporal resolutions of imaging systems with little regard for the analysts' objective use of the data. This "more is better" mindset has led to improved imaging systems but not necessarily an increase in the amount of information analysts obtain from these systems. This ABI performance assessment methodology uses the inherent characteristics of specific AoIs to develop performance measures for detecting those AoIs. It is this link to the activity which provides system designers and analysts with a credible set of requirements for use in baseline systems and tasking assets.

In this research, two notional graphs were suggested in Section 2.2. One would analyze the spatio-temporal tradespace associated with detecting an activity while the other analyzes the multimodal tradespace. This research produced the former, but left the multimodal tradespace to future researchers. For the object exchange activity, a spatio-temproral tradespace was developed by producing a likelihood of detection surface using a SAM based algorithm. This likelihood of detection surface provides an objective measure of assessing how this algorithm would perform under a range of spatio-temporal resolutions. It is expected that in the future, a lookup table similar to that of Figure 2.3 will be used to compare a broad list of algorithms capable of detecting the AoI.

## 7.2 Research Tasks

Several tasks were designated in Section 2.3 that needing to be accomplished in order to complete this research. The design and implementation of an experiment was a big step forward in developing this fairly new research area called Activity Based Intelligence. This dataset included several activities with varying spatial and temporal extents along with a few activities with unique spectral, polarimetric, and thermal characteristics. The rich multimodal nature of this dataset allows future researcher's to evaluate several ABI algorithms across a wide spread of activities using a broad range of multimodal sensors. The co-temporal nature of the activities also allows future researches to perform cross AoI analysis to determine how sensors handle multiple AoIs in one scene.

After developing the experiment, seven steps were evaluated.

1. Camera Calibration

2. Video Stabilization
3. Registration
4. Data Fusion
5. Tracking
6. Activity Recognition
7. Tradespace Development

Of those, as previously mentioned, the camera calibration and video stabilization steps were not used on this dataset. Regarding the tracking step, the target detection was done manually and the track association was automatic. While each step was evaluated by a particular method listed in this research, these are not the only ways of addressing these tasks. In fact, this list of tasks can be thought of as the spanning tree depicted in Figure 7.1. Beginning with the raw data, each branch represents a method of accomplishing a particular task. The second level presents two options for registration, the SURF method used in this research or a Maximization of Mutual Information (MMI) technique discussed in Section 3.4.1.2. This figure shows the nearly limitless combinations that could be evaluated by swapping out techniques in this sequence. It is presumed that each change will have some effect on the final detection surface, which would require evaluation.

## 7.3 Contributions to the Field

Four contributions to the field were described in Section 2.4:

- Development of a multimodal ABI dataset
- An end-to-end ABI evaluation of one activity
- Development of a limited multimodal ABI trade space
- Setting the foundation for an ABI lookup table

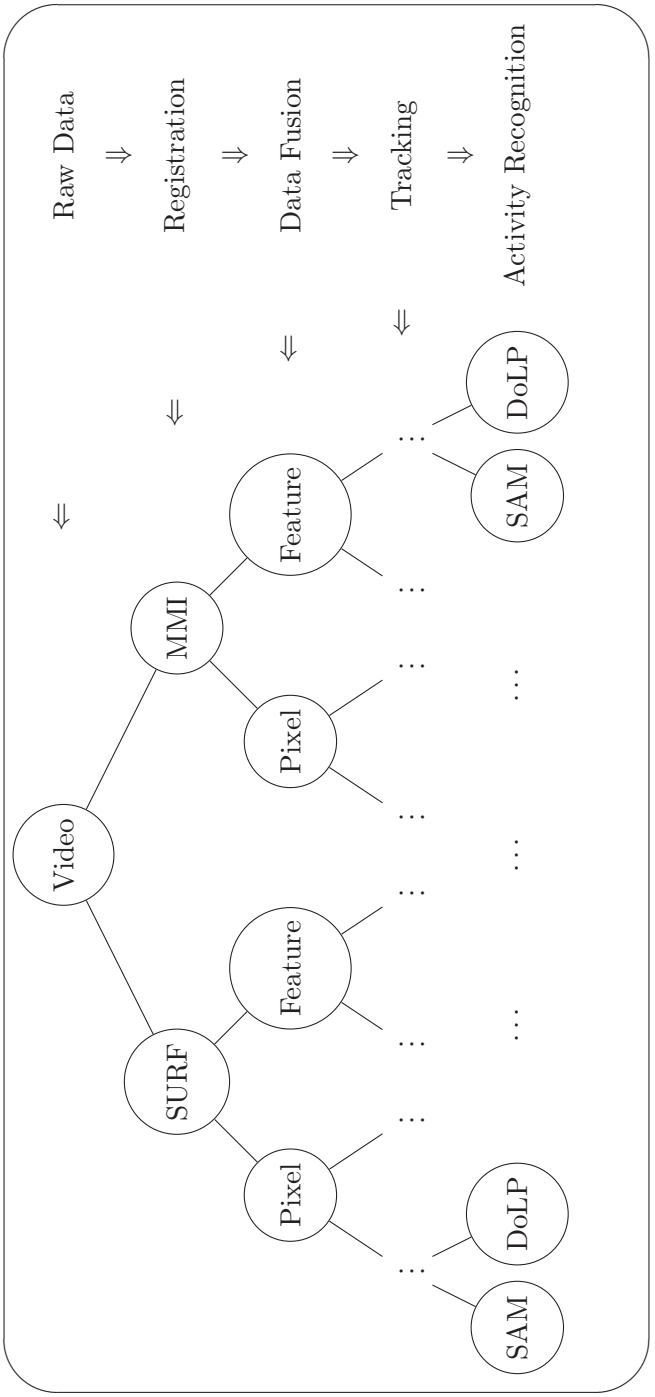



FIGURE 7.1: Task Options Spanning Tree

A multimodal ABI dataset was developed an used to complete the objectives of this research. Nine multimodal imagers captured several concurrent activities being executed in a real-world environment. This data is open and available for distribution in its raw from for interested future researchers.

An end-to-end ABI evaluation of one activity was depicted in both abbreviated and detailed forms in Figures 5.1 and 5.2 respectively. These two figures list the steps necessary to transform raw data into a set of detection surfaces for analysis. It is within the registration section of this evaluation that the object exchange analysis was performed.

The multimodal ABI tradespace was developed after a series of spatio-temporal degradations were performed on the results of the object exchange dataset. This led to developing several detection surfaces which can be used to make associations between AoIs and the sensor parameters needed capture these AoIs. Figures 6.15 through 6.17 depict these results.

Finally, the foundation for an ABI lookup table has been set. The activity recognition technique was successful in detecting an object exchange within this dataset, and thus will be included in the lookup table as a baseline for future work. Granted this work was completed on a limited dataset with a healthy dose of supervision. The purpose of its inclusion is to cite the novelty of the work; in hopes that someone will find it interesting enough to replace it with something better. Figure 7.2 depicts the object exchange lookup table with this technique included.

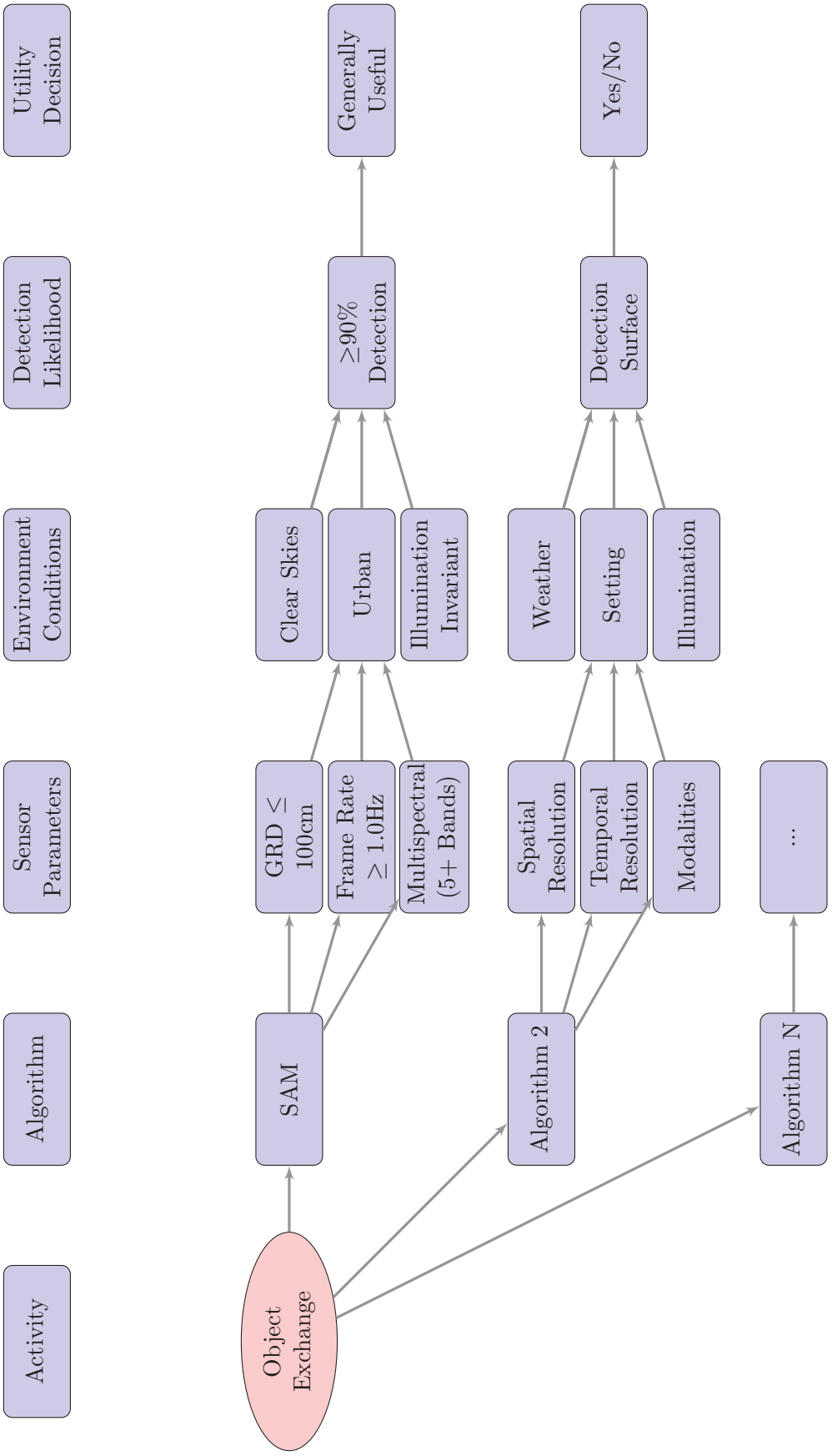



FIGURE 7.2: Object Exchange Lookup Table

# Chapter 8

# Future Work

Throughout the research, several areas of future work were noted and left for further discussion. Below are a few of these areas as well as some guidance regarding a possible direction for future research in this area.

1. Analysis of Other Activities in Dataset
2. Activity-Based Feature Space
3. Bounding Box Sensitivity Study
4. Time to Activity Analysis
5. Temporal Sensitivity Study
6. End-to-End Error Analysis
7. Alternate Methods of Assessing Spectral Angle Data

**Analysis of Other Activities in Dataset** Due to the limited time of this research, it was not possible to evaluate all of the activities that occurred within this dataset. That being an area of continued interest, analysis of these activities will be left to future researchers interested in continuing this work.

**Activity-Based Feature Space** Due to the requirements defined in Section 5.4.1, it is known that there are specific qualities of the in-scene activities that drive the requirements of the sensors needed to view these activities. As such, it may be possible to develop a specific feature space whereby certain activities are exclusively identified by their locality within this space.

**Bounding Box Sensitivity Study** In Section 5.7.1.3, we talked about using a predefined bounding box to isolate the object target pixels from those of other foreground pixels. In this research, an empirical value was used with objective study of how it actually affects the spectral angle of each frame.

**Time to Activity Analysis** While the likelihood surfaces in this scenario were developed to address specific sensor characteristics, other types of surfaces can be useful in detecting activities. An important quality that came up during the temporal degradations was the need to have more frames in the baseline spectrum. However, there was no analysis done to determine how far before the activity the baseline needed to be developed. If you have a sensor capturing imagery 30s ahead of the activity, can you reduce the frame rate of the sensor and still achieve a high likelihood of detection? What if the sensor were only able to capture 1s to 2s before the activity, but had a frame rate of 120Hz? Would this be enough to characterize the activity as it occurred? Figure 8.1 depicts a notional graphic of this concept.

Completing this type of analysis would allow future tipping and cueing scenarios where specialized sensors with exotic spatio-temporal characteristics could be cued by a generic sensor before the activity occurs. Very high frame rate systems with high spatial resolutions need large quantities of storage to retain the data they collect. Being able to minimize the time in which they are collecting could make using one of these sensors viable in activity recognition scenarios.

**Temporal Sensitivity Study** When the temporal degradations were performed on the dataset, they always began with the first frame in the sequence and skipped an integer number of frames thereafter. When the frame rates reach 1Hz and below the likelihood of detecting the object exchange dropped drastically. One of the questions

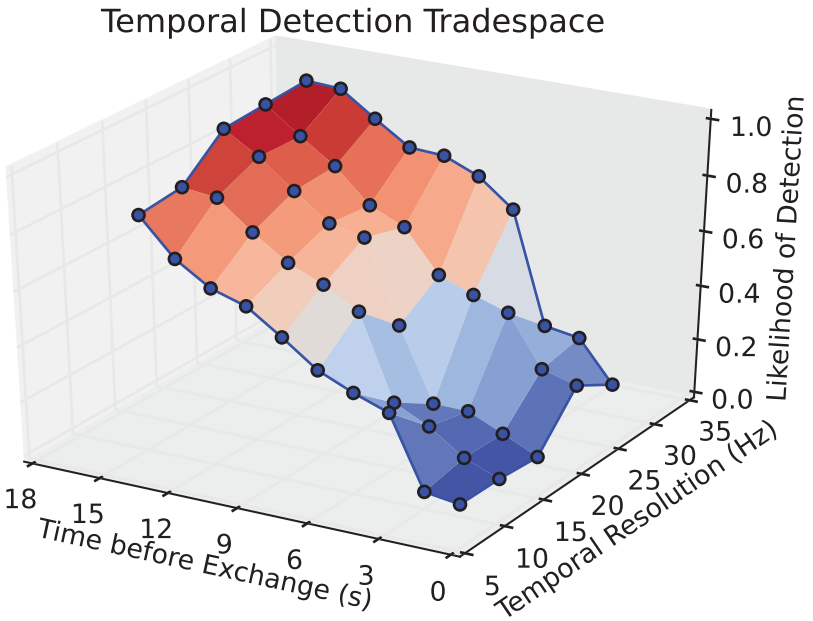



FIGURE 8.1: Time to Activity Tradespace

that occurred was, what if the degradations did not begin with the first frame in the sequence? How sensitive is the likelihood surface to the specific frames in the sequence begins with?

**End-to-End Error Analysis** Throughout this methodology, it was possible to perform an error propagation analysis to determine how systematic errors could effect the final result, but the work was not completed. An interesting study would be to develop an end-to-end analysis and determine how each step in figure 5.2 affects the final likelihood of detecting an activity. This would provide future analysts with a set of tolerances for each step in the process, thereby expanding the tradespace to include the software component of this process.

**Alternate Methods of Assessing Spectral Angle Data** Section 6.1.0.7 lists several additional methods that can be used to evaluate the spectral data in the object exchange scenario. Each is only a slightly different method of evaluation, but may prove useful to future researchers continuing this work.

In addition to the methods described in Section 6.1.0.7 one could evaluate the spectral angle data by using a change detection algorithm based on a mean-shift and outlier-distance. Zollweg et al [93] discuss this method for automatically detecting change in sequences based on these two principles.

# Bibliography


[1] "Kodak time lapse image (webpage)." http://www.kodak.com/eknec/PageQuerier.jhtml?pq-path=10948&pq-locale=en_US&_requestid=4363, June 2014.

[2] "Motosport time lapse image (webpage)." https://farm2.staticflickr.com/1149/850611221_6ab3b2597c.jpg, 2014 June.

[3] J. C. Leachtenauer, W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, "General image-quality equation: Giqe," *Appl. Opt.* **36**, pp. 8322–8328, Nov 1997.

[4] D. Young, J. Yen, F. Petitti, T. Bakir, M. Brennan, and R. Butto, Jr., "Video national imagery interpretability rating scale criteria survey results," 2009.

[5] "Focal length vs fov image demonstration." http://en.wikipedia.org/wiki/Angle_of_view#mediaviewer/File:Focal_length.jpg, February 2008.

[6] L. . DIRS, "Wasp lite sensor system," *Center for Imaging Science, Rochester Institute of Technology* , September 2013.

[7] J. Richardson, Michael; Faulring, "Wasp-lite design review." April 2005.

[8] J. A. Herweg, J. P. Kerekes, and M. T. Eismann, "Hyperspectral measurements of natural signatures: pedestrians," 2012.

[9] J. A. Herweg, J. P. Kerekes, O. Weatherbee, D. Messinger, J. van Aardt, E. Ientilucci, Z. Ninkov, J. Faulring, N. Raqueño, and J. Meola, "Spectir hyperspectral airborne rochester experiment data collection campaign," 2012.

[10] J. A. Herweg, J. P. Kerekes, E. J. Ientilucci, and M. T. Eismann, "Spectral variations in hsi signatures of thin fabrics for detecting and tracking of pedestrians," 2011.

[11] B. D. Bartlett, J. F. Faulring, and C. Salvaggio, "System characterization and analysis of the multispectral aerial passive polarimeter system (mapps)," *Proc. SPIE* **8160**, pp. 81600A–81600A–8, 2011.

[12] Amazon, "Go pro hero3: Black edition." `http://www.amazon.com/GoPro-CHDHX-301-HERO3-Black-Edition/dp/B009TCD8V8`, February 2014.

[13] "Google maps imagery." `www.google.com/maps`.

[14] B. Bartlett, "Fisheye camera calibration fisheye camera calibration fisheye camera calibration." `http://twiki.cis.rit.edu/twiki/bin/view/Main/FisheyeCameraCalibration`, January 2007.

[15] I. R. Assessment, R. Assessment, R. I. R. Assessment, and R. S. I. Committee, "Multispectral imagery interpretability rating scale: Reference guide." February 1995.

[16] J. C. Leachtenauer, W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, "General image-quality equation for infrared imagery," *Appl. Opt.* **39**, pp. 4826–4828, Sep 2000.

[17] L. for Imaging Algorithms and Systems', "Low altitude multi-spectral mapping system (lamms) (aka wasp-lite) specification and interface control document." October 2007.

[18] J. Group, "JAI products: BM-500GE webpage." `http://www.jai.com/en/products/bm-500ge`, 2013.

[19] G. Pro, "Hero 3 camera FAQs webpage." `http://gopro.com/support/articles/hero3-faqs`, February 2014.

[20] "Go pro hero3: Black edition field of view information." `http://gopro.com/support/articles/hero3-field-of-view-fov-information`, 2014.

[21] "Go pro hero3: Black edition technical specs." `http://gopro.com/cameras/hd-hero3-black-edition#technical-specs`, 2014.

[22] C. Drew, "Military is awash in data from drones," *The New York Times* , January 2010.

[23] S. Magnuson, "Military 'swimming in sensors and drowning in data'," *National Defense* , January 2010.

[24] A. Corrin, "Sensory overload: Military is dealing with a data deluge," *Defense Systems* , p. 2, February 2010.

[25] "Too much information: Taming the UAV data explosion," *Defense Industry Daily* , May 2010.

[26] The Economist, "A special report on managing information: Data, data everywhere," *The Economist* , February 2010.

[27] T. Shanker and M. Richtel, "In new military, data overload can be deadly," *The New York Times* , January 2011.

[28] Z. Sutton, "How can we kill the megapixel war?," *Fstoppers* , July 2013.

[29] C. Chan, "Here's a good reason why the camera megapixel wars needs to stop," *Gizmodo* , December 2012.

[30] S. Dent, "The next mobile imaging war won't be waged over megapixels," *Engadget* , February 2014.

[31] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, "Simple and complex activity recognition through smart phones," tech. rep., Washington State University.

[32] J. James R. Clapper, "Geoint 2012 - opening keynote, director of national intelligence." `http://geointv.com/archive/geoint-2012-tuesday-keynote-james-r-clapper-jr-director-of.../`, October 2012.

[33] "Center for urban science + progress." `http://cusp.nyu.edu/`.

[34] J. P. G. E. Havig, Paul R.; McIntire and F. Mohd-Zaid, "Why social network analysis is important to air force applications," *Proc. SPIE* **8389**, 2012.

[35] T. Economist, "Nanosats are go!," *The Economist* , June 2014.

[36] D. P., "Htc one and the ultrapixels: the end of the megapixel war," *Phonarena* , 2013.

[37] D. L. Young, J. Ruszczyk, and T. Bakir, "Video quality and interpretability study using samviq and video-niirs," 2011.

[38] J. M. Irvine, D. Cannon, J. Miller, J. Bartolucci, G. O'Brien, L. Gibson, C. Fenimore, J. Roberts, I. Aviles, M. Brennan, A. Bozell, L. Simon, and S. A. Israel, "Methodology study for development of a motion imagery quality metric," 2006.

[39] J. M. Irvine, A. I. Aviles, D. M. Cannon, C. Fenimore, D. S. Haverkamp, S. A. Israel, G. O'Brien, and J. Roberts, "Developing an interpretability scale for motion imagery," *Optical Engineering* **46**(11), pp. 117401–117401–12, 2007.

[40] "Motion imagery standards board - frequently asked questions." `http://www.gwg.nga.mil/misb/faq.html`.

[41] B. D. Bartlett and M. D. Rodriguez, "Snapshot spectral and polarimetric imaging; target identification with multispectral video," *Proc. SPIE* **8743**, pp. 87430R–1:87430R–8, 2013.

[42] J. R. Schott, *Fundamentals of Polarimetric Remote Sensing*, vol. TT81, SPIE Tutorial Texts in Optical Engineering, SPIE Press, Bellingham, Washington USA, 2009.

[43] S. Tominaga, H. Kadoi, K. Hirai, and T. Horiuchi, "Metal-dielectric object classification by combining polarization property and surface spectral reflectance," 2013.

[44] D. Nilosek, S. Sun, and C. Salvaggio, "Geo-accurate model extraction from three-dimensional image-derived point clouds," 2012.

[45] E. Ontiveros, C. Salvaggio, D. Nilosek, N. Raqueño, and J. Faulring, "Evaluation of image collection requirements for 3d reconstruction using phototourism techniques on sparse overhead data," 2012.

[46] X. Hu, L. Ye, X. Li, J. Zhu, and H. Long, "Fusion of airborne lidar point cloud and imagery captured from integrated sensor system," 2011.

[47] C. J. Miller, "Fusion of high-resolution lidar elevation data with hyperspectral data to characterize tree canopies," 2001.

[48] J. S. J. Peri, "Lidar characteristics for detecting and tracking high-speed bullets," 2011.

[49] S. Sato, M. Hashimoto, M. Takita, K. Takagi, and T. Ogawa, "Multilayer lidar-based pedestrian tracking in urban environments," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, pp. 849–854, June 2010.

[50] J. Douglas, M. Burke, and G. J. Ettinger, "High-resolution sar atr performance analysis," 2004.

[51] G. S. Goley, B. Rigling, and A. R. Nolan, "Sar based classification of ground moving targets to assist vehicle tracking," 2013.

[52] Z. Zhao, K. Ji, X. Xing, and H. Zou, "Adaptive cfar detection of ship targets in high resolution sar imagery," 2013.

[53] G. E. Newstadt, E. Zelnio, L. Gorham, and A. O. Hero III, "Detection/tracking of moving targets with synthetic aperture radars," 2010.

[54] J. R. Schott, *Remote Sensing: The Image Chain Approach*, Oxford University Press, USA, 2nd ed., 2007.

[55] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.* **110**, pp. 346–359, June 2008.

[56] X. Fan, H. Rhody, and E. Saber, "A spatial-feature-enhanced mmi algorithm for multimodal airborne image registration," *Geoscience and Remote Sensing, IEEE Transactions on* **48**, pp. 2580–2589, June 2010.

[57] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall Professional Technical Reference, 2002.

[58] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer-Verlag New York, Inc., New York, NY, USA, 1st ed., 2010.

[59] J. E. Solem, *Programming computer vision with Python*, O'Reilly, Beijing; Cambridge; Sebastopol [etc.], 2012.

[60] T. Zhang, "Multiple-target tracking using spectropolarimetric imagery," Master's thesis, Rochester Institute of Technology, Chester F. Carlson Center for Imaging Science, April 2013.

[61] K. T. Ausfeld, "Tracking of various targets in the infrared and issues encountered," Master's thesis, Rochester Institute of Technology, Chester F. Carlson Center for Imaging Science, October 2012.

[62] S. S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*, Artech House radar library, Artech House, Boston, 1999.

[63] H. Ling, Y. Wu, E. Blasch, G. Chen, H. Lang, and L. Bai, "Evaluation of visual tracking in extremely low frame rate wide area motion imagery," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pp. 1–8, 2011.

[64] J. Gao, H. Ling, E. Blasch, K. Pham, Z. Wang, and G. Chen, "Pattern of life from wami objects tracking based on visual context-aware tracking and infusion network models," 2013.

[65] M. Jamali and H. Abolhassani, "Different aspects of social network analysis," in *Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference on*, pp. 66–72, Dec 2006.

[66] P. R. Havig, J. P. McIntire, E. Geiselman, and F. Mohd-Zaid, "Why social network analysis is important to air force applications," 2012.

[67] S. Kant, "Activity-based exploitation of full motion video (fmv)," 2012.

[68] T. D. Lash, "Uses of motion imagery in activity-based intelligence," 2013.

[69] I. Laptev and T. Lindeberg, "Space-time interest points," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 432–439 vol.1, Oct 2003.

[70] F. Yuan, G.-S. Xia, H. Sahbi, and V. Prinet, "Spatio-temporal interest points chain (stipc) for activity recognition," in *Pattern Recognition (ACPR), 2011 First Asian Conference on*, pp. 22–26, Nov 2011.

[71] C.-M. Zhai, Y.-L. Guo, and J. xiang Du, "Event recognition based on bag of local space-time interest points' features," in *Advanced Computational Intelligence (IWACI), 2011 Fourth International Workshop on*, pp. 477–481, Oct 2011.

[72] I. Bellamine and H. Tairi, "Detecting motion using the structure-texture image decomposition and space-time interest points," in *Intelligent Systems: Theories and Applications (SITA), 2013 8th International Conference on*, pp. 1–7, May 2013.

[73] M.-C. Chang, N. Krahnstoever, S. Lim, and T. Yu, "Group level activity recognition in crowded environments across multiple cameras," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pp. 56–63, Aug 2010.

[74] A. Yazidi, O.-C. Granmo, and B. Oommen, "Learning-automaton-based online discovery and tracking of spatiotemporal event patterns," *Cybernetics, IEEE Transactions on* **43**, pp. 1118–1130, June 2013.

[75] A. Cavent and N. Cinbis, "Recognition of complex human activities by using sequential pyramid matching," in *Signal Processing and Communications Applications Conference (SIU), 2012 20th*, pp. 1–4, April 2012.

[76] W. Raetz, "A new approach to graph analysis for activity based intelligence," in *Applied Imagery Pattern Recognition Workshop (AIPR), 2012 IEEE*, pp. 1–8, Oct 2012.

[77] F. Bunyak, *Moving object detection and tracking for event-based video analysis*. PhD thesis, University of Missouri-Rolla, 2005.

[78] C. M. Lewis, D. Messinger, and B. Neuberger, "Multimodal sensor experiment for evaluating motion imagery in activity-based intelligence," 2014.

[79] C. M. Lewis, D. Messinger, and M. G. Gartley, "Activity-based intelligence tipping and cueing using polarimetric sensors," 2014.

[80] A. J. Lingg and B. D. Rigling, "Foreground estimation in motion imagery using multi-frame change detection techniques," 2013.

[81] G. Levchuk, M. Jacobsen, C. Furjanic, and A. Bobick, "Learning and detecting coordinated multi-entity activities from persistent surveillance," 2013.

[82] *2009 ASHRAE Handbook - Fundamentals (SI Edition)*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 1791 Tullie Circle, NE, Atlanta, GA, 2009.

[83] T. E. ToolBox, "Radiation emissivity of some common materials: Radiation emissivity of some common materials - water, ice, snow, grass ..." `http://www.engineeringtoolbox.com/radiation-heat-emissivity-d_432.html`.

[84] T. E. ToolBox, "Radiation emissivity of some common materials: The radiation heat transfer emissivity coefficient of some common materials as aluminum, brass, glass and many more." `http://www.engineeringtoolbox.com/emissivity-coefficients-d_447.html`.

[85] R. J. F. H. Hanlon, Joseph F; Kelsey, *Handbook of Package Engineering*, Technomic Publishing Co., Lancaster, Pa, 3rd ed., 1998.

[86] C. N. Tran, R. Innocenti, G. Kirose, K. I. Ranney, and G. Smith, "Inverse synthetic aperture radar imagery of a man with a rocket propelled grenade launcher," 2004.

[87] Photometrix, "Australis guide." Users Manual, Photometrix Pty Ltd, PO Box 3023, Kew, Vic 3101, Australia, January 2004.

[88] A. L. Robinson, B. Miller, S. Moyer, and C. Ra, "Low- to mid-altitude temporal/spatial tracking requirements," *Optical Engineering* **46**(9), pp. 096401–096401–8, 2007.

[89] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following.," *Computer Vision, Graphics, and Image Processing* **30**(1), pp. 32–46, 1985.

[90] H. W. Kuhn, "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly* **2**, pp. 83–97, March 1955.

[91] J. R. Munkres, "Algorithms for the Assignment and Transportation Problems," *Journal of the Society for Industrial and Applied Mathematics* **5**, pp. 32–38, March 1957.

[92] M. T. Eismann, *Hyperspectral Remote Sensing*, vol. SPIE PM210, SPIE Press, Bellingham, Washington USA, 1st ed., 2012.

[93] D. B. G. Joshua Zollwega, Ariel Schlamm and D. Messinger, *Change detection using mean-shift and outlier-distance metrics*. PhD thesis, Rochester Institute of Technology.

# Appendix A

# IR and Multispectral National Image Interpretability Rating Scales

Figure A.1 depicts a small sample of the Multispectral NIIRS. Due to the large tradespace including in multispectral data, the current rating system is neither all inclusive nor complete.

Figure A.2 depicts the IR NIIRS.

**MS IIRS Level 4 (continued)**
Detect landslide or rocicslide large enough to obstruct a single-lane road.
Detect small boats ( 15-20 feet in length) in open water.
Identify areas suitable for use as light fixed-wing aircraft (e.g., Cessna, Piper Cub, Beechcraft) landing strips.

**MS IIRS Level 5**
Detect automobile in a parking lot.
Identify beach terrain suitable for amphibious landing operation.
Detect ditch irrigation of beet fields.
Detect disruptive or deceptive use of paints or coatings on buildings/structures at a ground forces installation.
Detect raw construction materials in ground forces deployment areas (e.g., timber, sand, gravel).

**MS IIRS Level 6**
Detect summer woodland camouflage netting large enough to cover a tank against a scattered tree background.
Detect foot trail through tall grass.
Detect navigational channel markers and mooring buoys in water.
Detect livestock in open but fenced areas.
Detect recently installed minefields in ground forces deployment area based on a regular pattern of disturbed earth
Count individual dwellings in subsistence housing areas (e.g., squatter settlements, refugee camps).

**MS IIRS Level 7**
Distinguish between tanks and three-dimensional tank decoys.
Identify individual 55-gallon drums.
Detect small marine mammals (e.g., harbor seals) on sand/gravel beaches.
Detect underwater pier footings.
Detect foxholes by ring of spoil outlining hole.
Distinguish individual rows of truck crops.

FIGURE A.1: NIIRS Rating Scale [15]

**Rating Level 0**
Interpretability of the imagery is precluded by obscuration, degradation, or very poor resolution.
**Rating Level 1**
Distinguish between runways and taxiways on the basis of size, configuration, or pattern at a large airfield.
Detect a large (e.g., greater than 1 $km^2$) cleared area in dense forest.
Detect large ocean-going vessels (e.g., aircraft carrier, supertanker, KIROV) in open water.
Detect large areas (e.g., greater than 1 $km^2$) of marsh/swamp.
**Rating Level 2**
Detect large aircraft (e.g., C-141, 707, BEAR, CANDID, CLASSIC).
Detect individual large buildings (e.g., hospitals, factories) in an urban area.
Distinguish between densely wooded, sparsely wooded, and open fields.
Identify an SS-25 base by the pattern of buildings and roads.
Distinguish between naval and commercial port facilities based on type and configuration of large functional areas.
**Rating Level 3**
Distinguish between large (e.g., C-141, 707, BEAR, A-300 AIRBUS) and small aircraft (e.g., A-4, FISHBED, L-39).
Identify individual thermally active flues running between the boiler hall and smoke stacks at a thermal power plant.
Detect a large air warning radar site based on the presence of mounds, revetments, and security fencing.
Detect a driver training track at a ground forces garrison.
Identify individual functional areas (e.g., launch sites, electronics area, support area, missile handling area) of an SA-5 launch complex.
Distinguish between large (e.g., greater than 200 m) freighters and tankers.
**Rating Level 4**
Identify the wing configuration of small fighter aircraft (e.g., FROGFOOT, F-16, FISHBED).
Detect a small (e.g., 50 $m^2$) electrical transformer yard in an urban area.
Detect large (e.g., greater than 10-m diameter) environmental domes at an electronics facility.
Detect individual thermally active vehicles in garrison.
Detect thermally active SS-25 MSV's in garrison.
Identify individual closed cargo hold hatches on large merchant ships.
**Rating Level 5**
Distinguish between single-tail (e.g., FLOGGER, F-16, TORNADO) and twin-tailed (e.g., F-15, FLANKER, FOXBAT) fighters.
Identify outdoor tennis courts.
Identify the metal lattice structure of large (e.g., approximately 75 m) radio relay towers.
Detect armored vehicles in a revetment.
Detect a deployed TET (transportable electronics tower) at an SA-10 site.
Identify the stack shape (e.g., square, round, oval) on large (e.g greater than 200 m) merchant ships.
**Rating Level 6**
Detect wing-mounted stores (i.e., ASM, bombs) protruding from the wings of large bombers (e.g., B-52, BEAR, BADGER).
Identify individual thermally active engine vents atop diesel locomotives.
Distinguish between a FIX FOUR and FIX SIX site based on antenna pattern and spacing.
Distinguish between thermally active tanks and APC's.
Distinguish between a two-rail and four-rail SA-3 launcher.
Identify missile tube hatches on submarines.
**Rating Level 7**
Distinguish between ground attack and interceptor versions of the MIG-23 FLOGGER based on the shape of the nose.
Identify automobiles as sedans or station wagons.
Identify antenna dishes (less than 3-m diameter) on a radio relay tower.
Identify the missile transfer crane on a SA-6 transloader.
Distinguish between an SA-2/CSA-1 and a SCUD-B missile transporter when missiles are not loaded.
Detect mooring cleats or bollards on piers.
**Rating Level 8**
Identify the RAM airscoop on the dorsal spine of FISHBED J/K/L.
Identify limbs (e.g., arms, legs) on an individual.
Identify individual horizontal and vertical ribs on a radar antenna.
Detect closed hatches on a tank turret.
Distinguish between fuel and oxidizer Multi-System Propellant Transporters based on twin or single fitments on the front of the semi-trailer.
Identify individual posts and rails on deck edge life rails.
**Rating Level 9**
Identify access panels on fighter aircraft.
Identify cargo (e.g., shovels, rakes, ladders) in an open-bed, light-duty truck.
Distinguish between BIRDS EYE and BELL LACE antennas based on the presence or absence of small dipole elements.
Identify turret hatch hinges on armored vehicles.
Identify individual command guidance strip antennas on an SA-2/CSA-1 missile.
Identify individual rungs on a bulkhead mounted ladder.

FIGURE A.2: IR NIIRS [16]

# Appendix B

# Spatial Registration Results

Figures B.1, B.2, B.3, B.4 depict the registration results for multistep blur and SURF feature extraction process. Note, the left contains the entire image from both imagers, whereas the right masks out non-overlapping portions of imagery. The Red and Blue channel were filled with the panchromatic image and the Green channel was filled with the greyscale registered Go Pro Image. The titles of each image indicate the blur kernel size and amount of Sum Square Error (SSE).

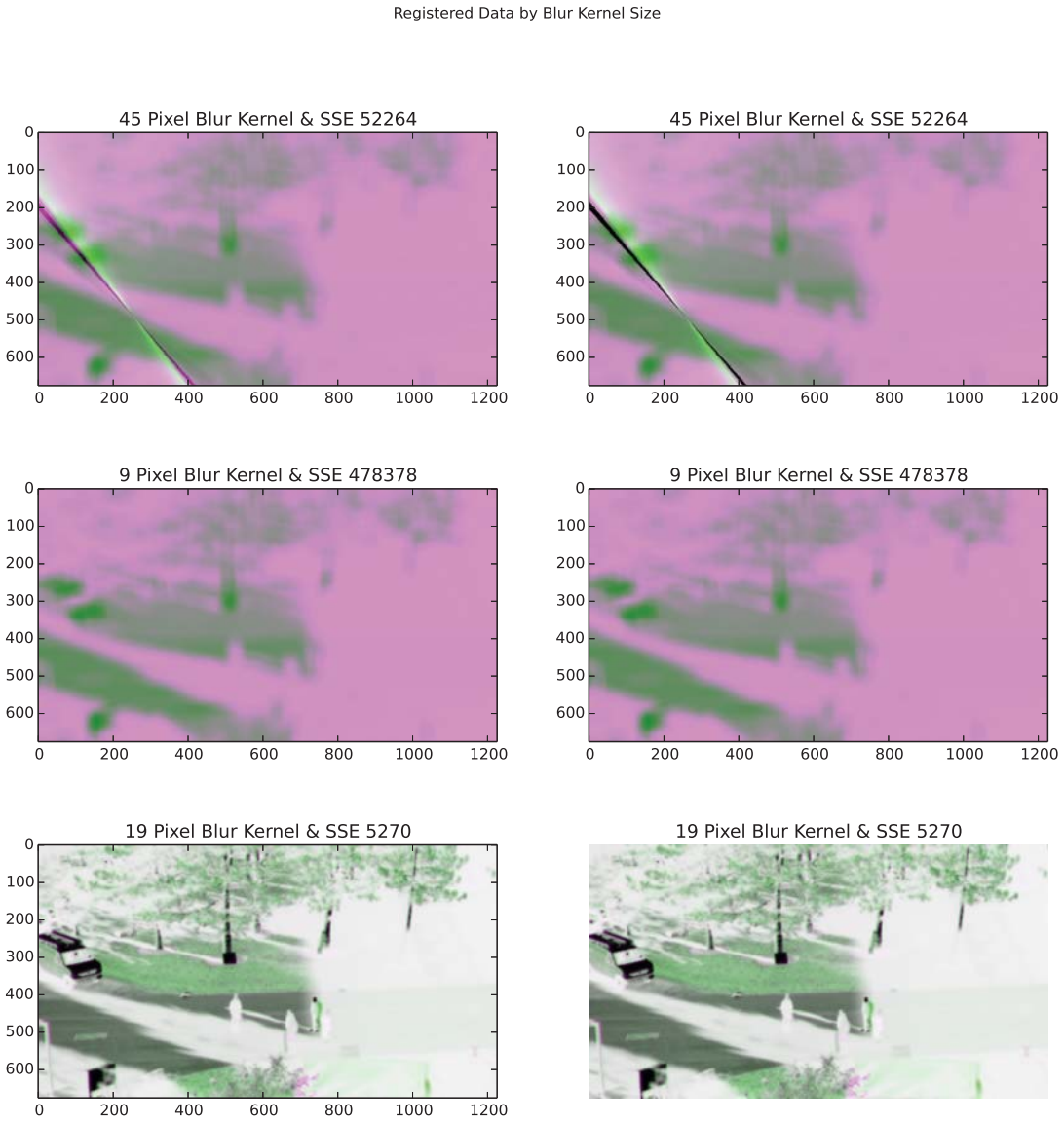



FIGURE B.1: Multispectral Filter 1

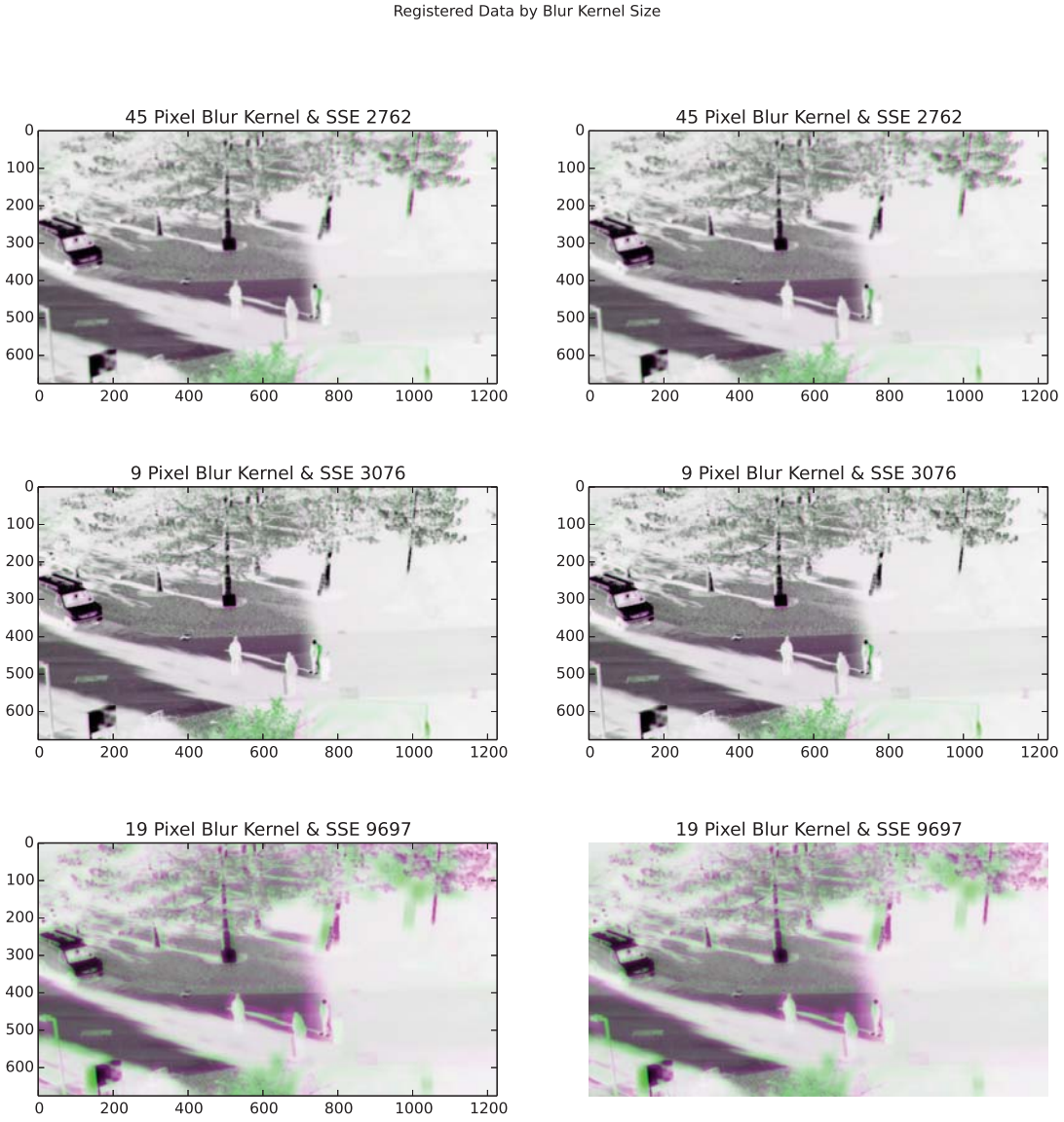



FIGURE B.2: Multispectral Filter 2

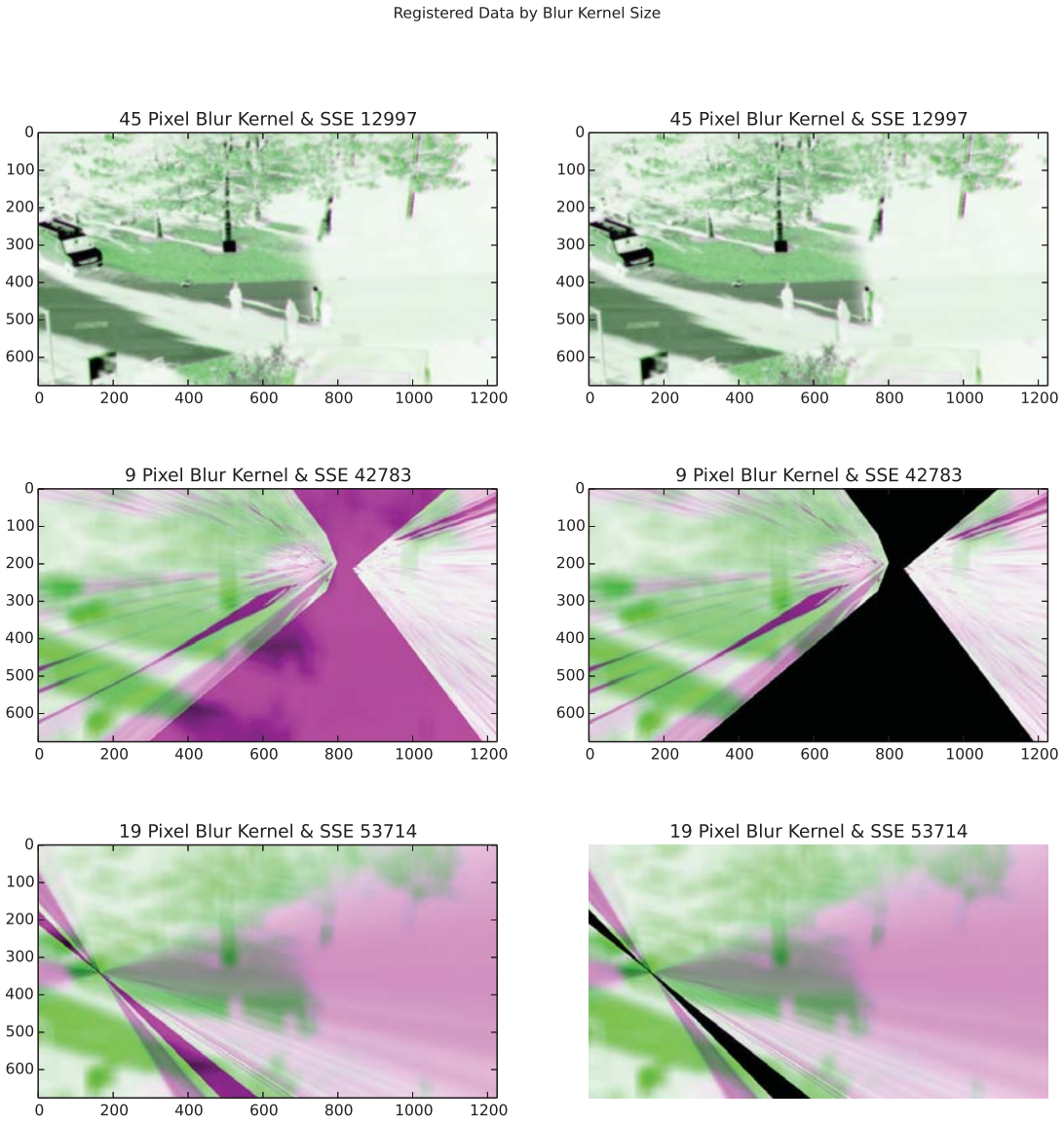



FIGURE B.3: Multispectral Filter 4

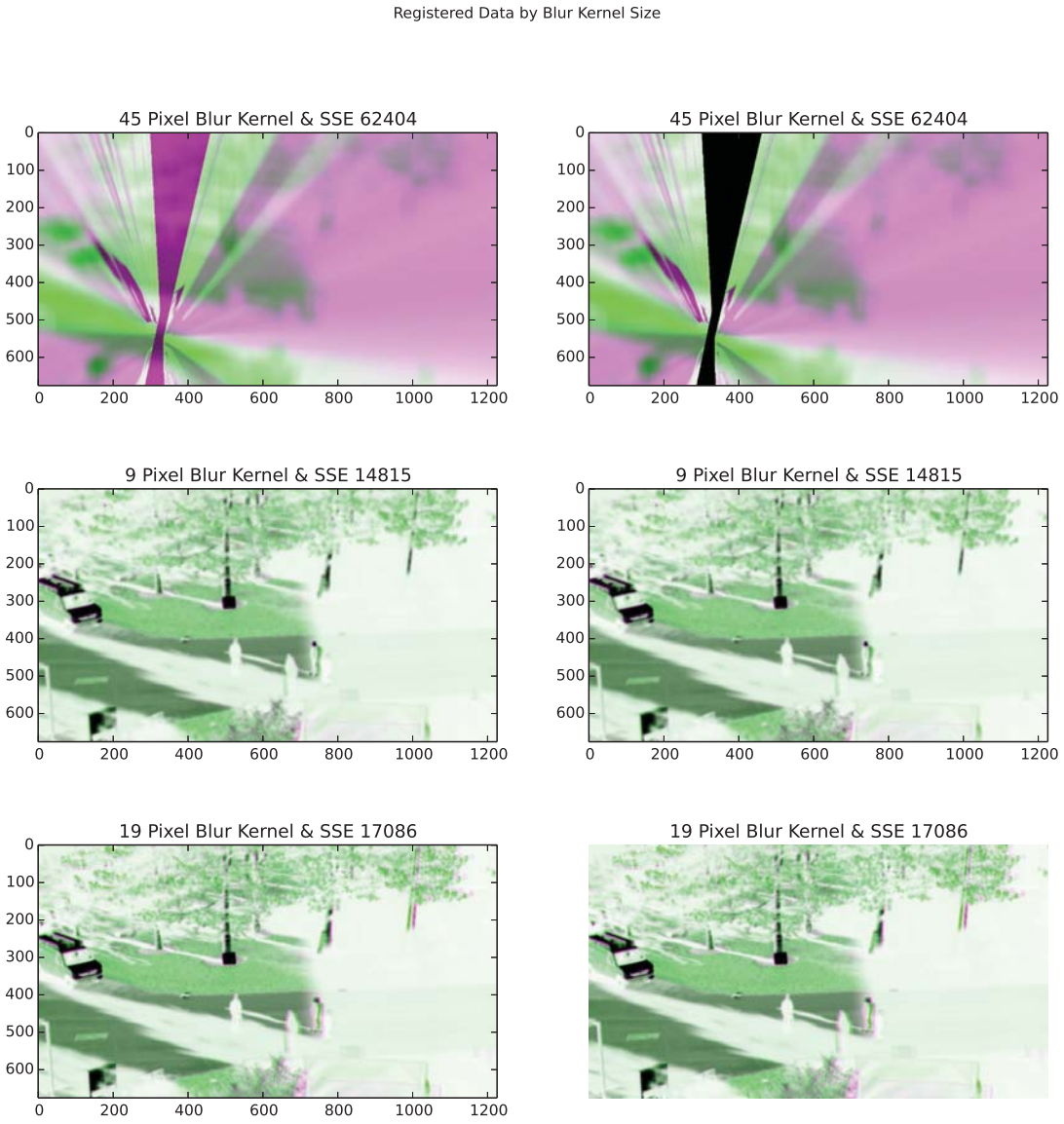



FIGURE B.4: Multispectral Filter 5

# Appendix C

# Experimental Setup Imagery

Figures C.1, C.2, C.3, C.4, C.5 depict the setup of the equipment for the experiment.

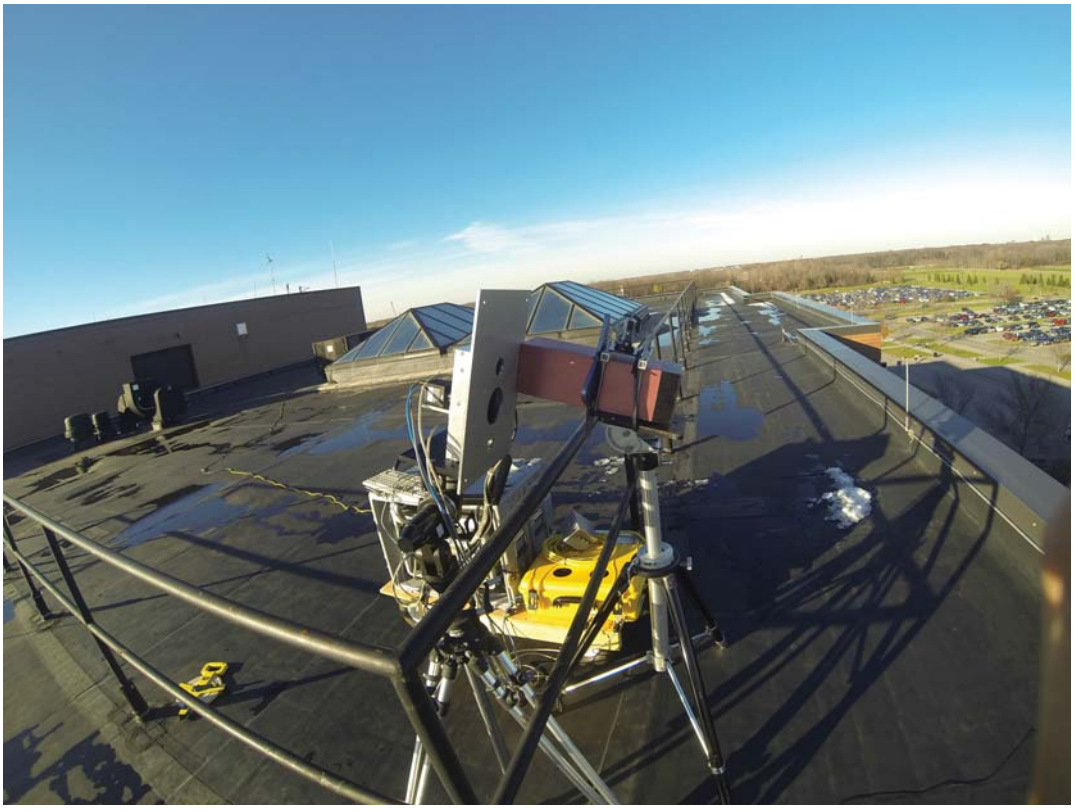


FIGURE C.1: Experimental Setup Image 2

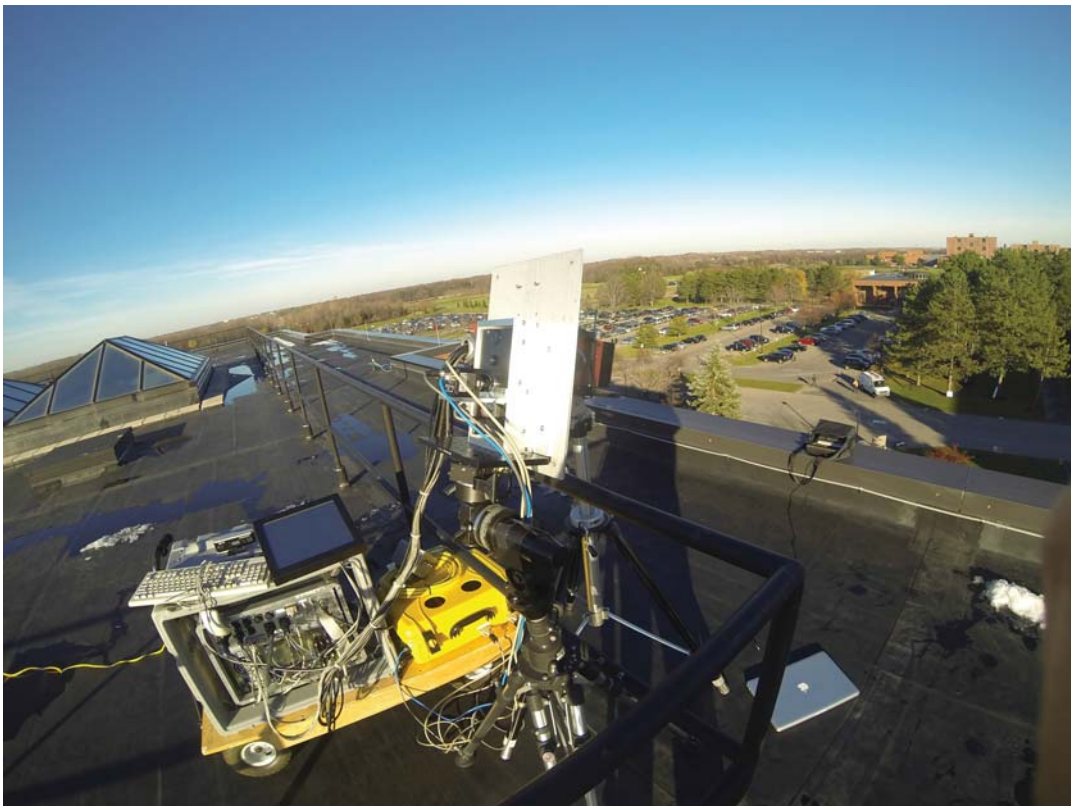


FIGURE C.2: Experimental Setup Image 3

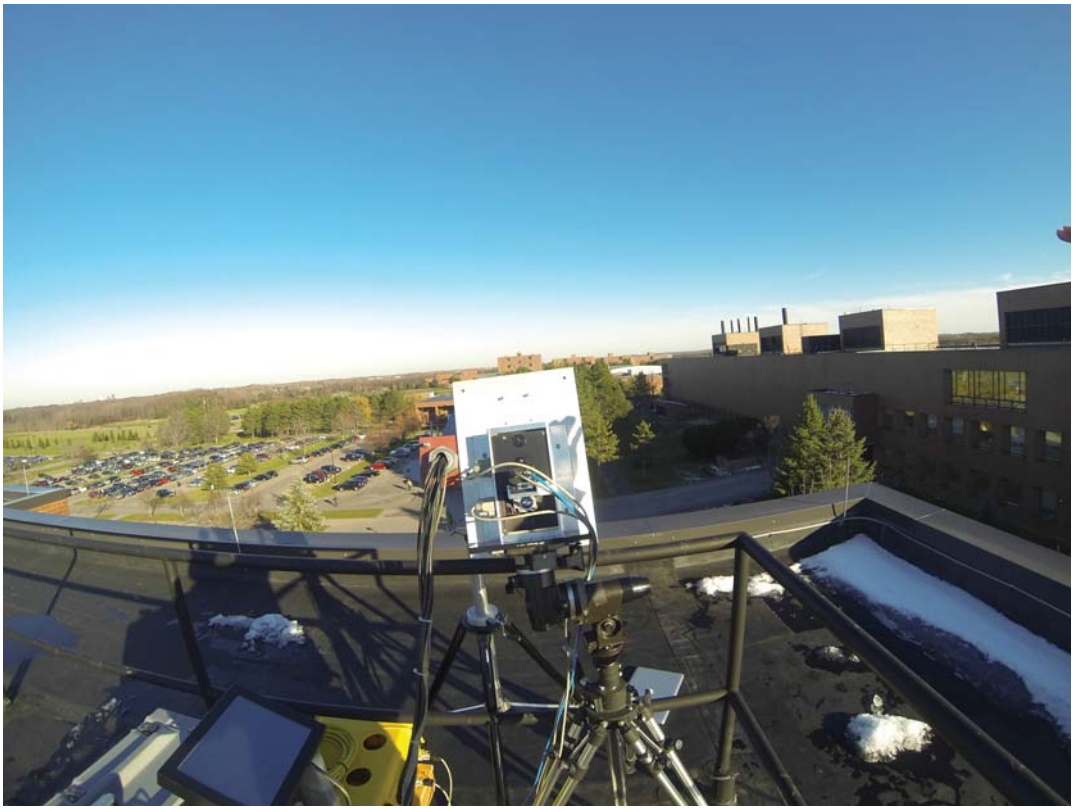


FIGURE C.3: Experimental Setup Image 4

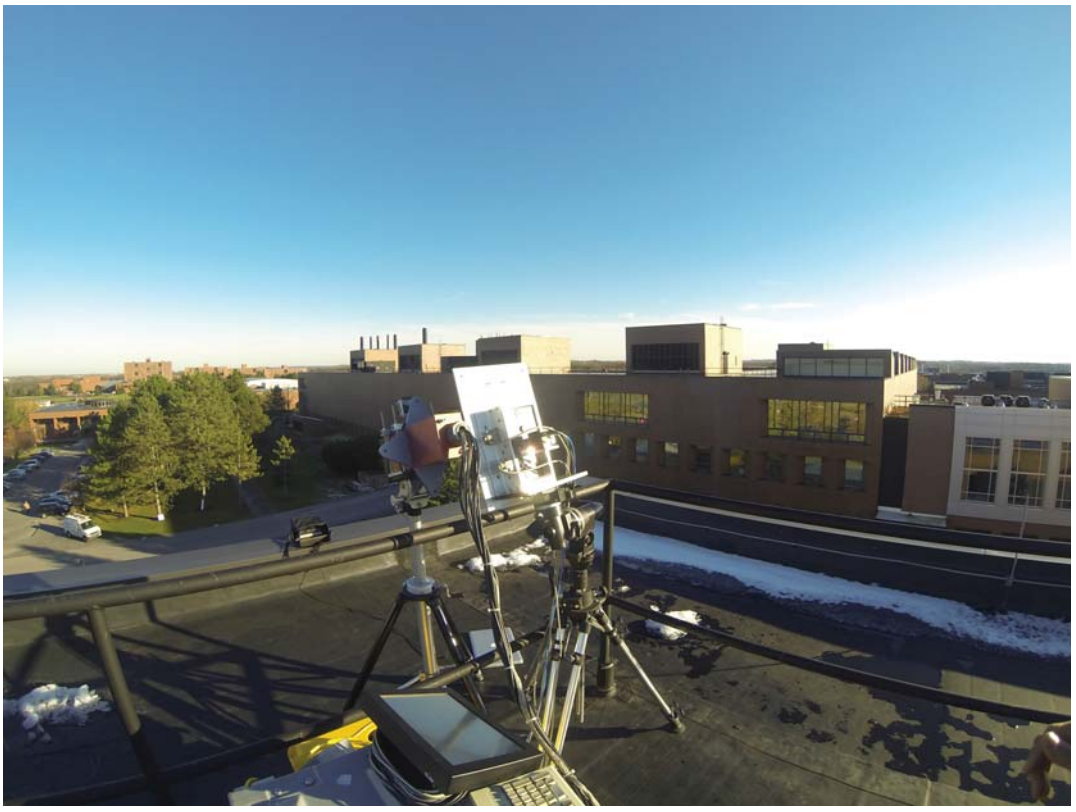


FIGURE C.4: Experimental Setup Image 5

FIGURE C.5: Experimental Setup Image 8

# Appendix D

# Experimental Fiducials

Figures D.2, D.1, D.3, D.4, D.5, D.6, D.7, D.8, D.9, D.10, depict the fiducials used in this experiment. The order begins with Fiducial B to save white space on this page.

Fiducials J and K, depicted in Figures D.9 D.10 respectively, were large pieces of plexiglass with cardboard layered behind them. The thermally reflective qualities of the plexiglass allowed for distinct cold space-based emissions to be directed at the sensor, portraying a well defined object relative to its surroundings. The cardboard was used to outline the general shape of the plexiglass for detection in the visible regime. The two figures above were early, labeled, iterations of the fiducials. In the final implementation the cardboard was completely covering the backside of the plexiglass without any overhang.

FIGURE D.1: Fiducial B

FIGURE D.2: Fiducial A

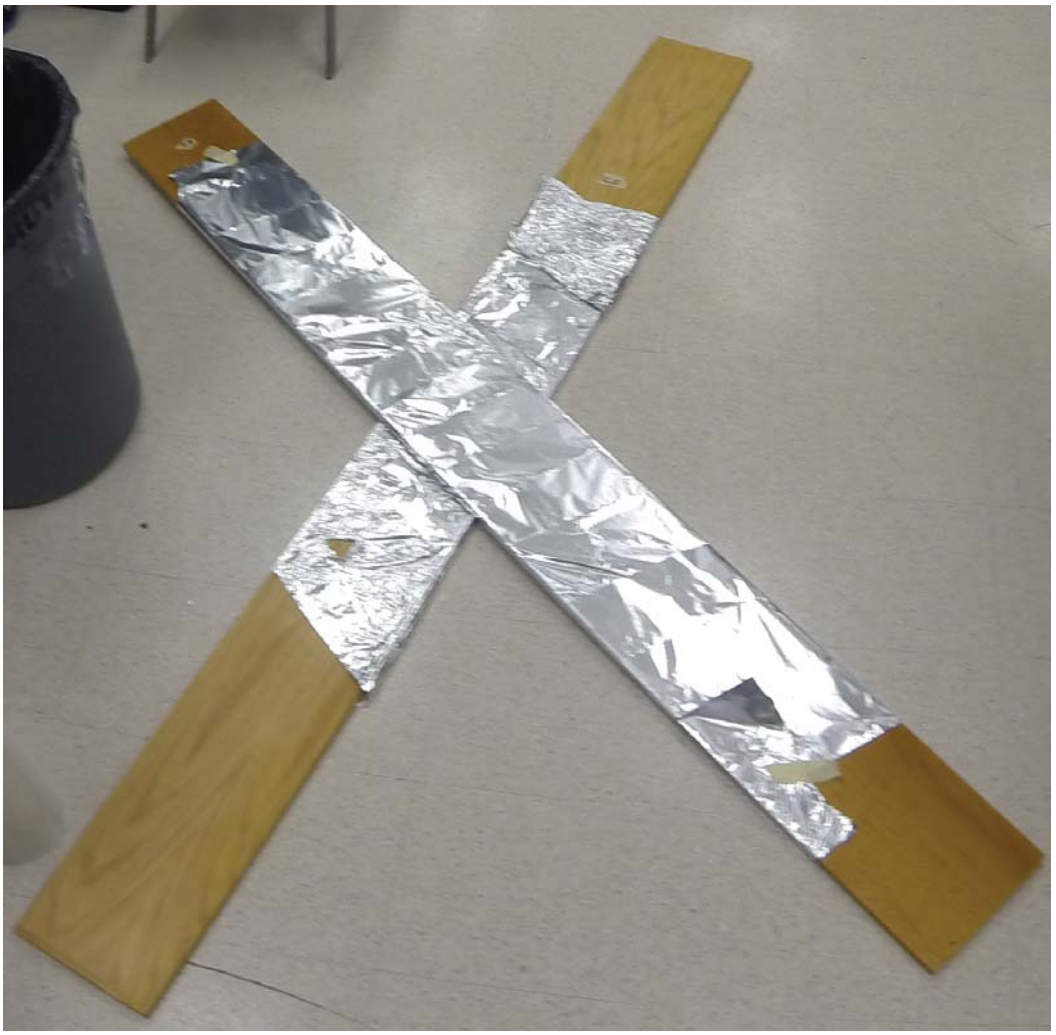


FIGURE D.3: Fiducial C


FIGURE D.4: Fiducial D

FIGURE D.5: Fiducial F


FIGURE D.6: Fiducial G


FIGURE D.7: Fiducial H


FIGURE D.8: Fiducial I


FIGURE D.9: Fiducial J

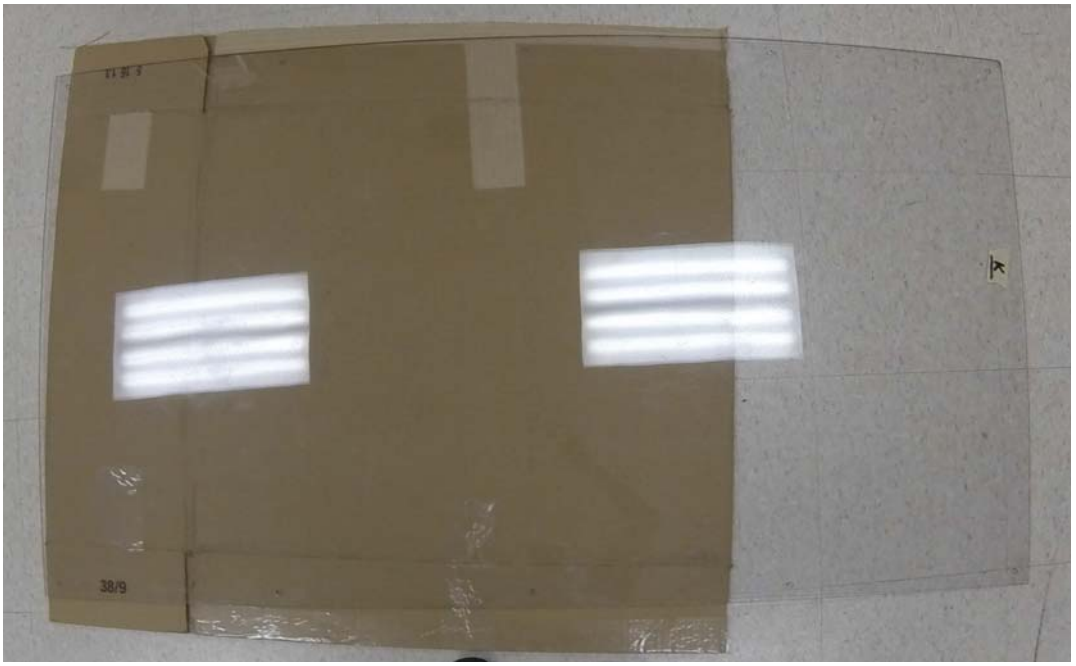
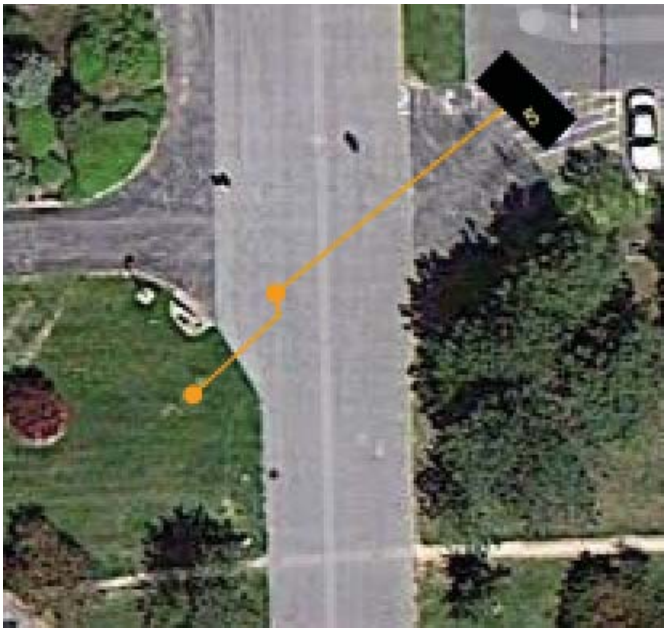



FIGURE D.10: Fiducial K

# Appendix E

# Participant Directions

Figures E.1, E.2, E.3, E.4, E.5, E.6, E.7, E.8, depict the directions given to each of the participants in this experiment. We begin with page three so as not to waste the white space on this page.

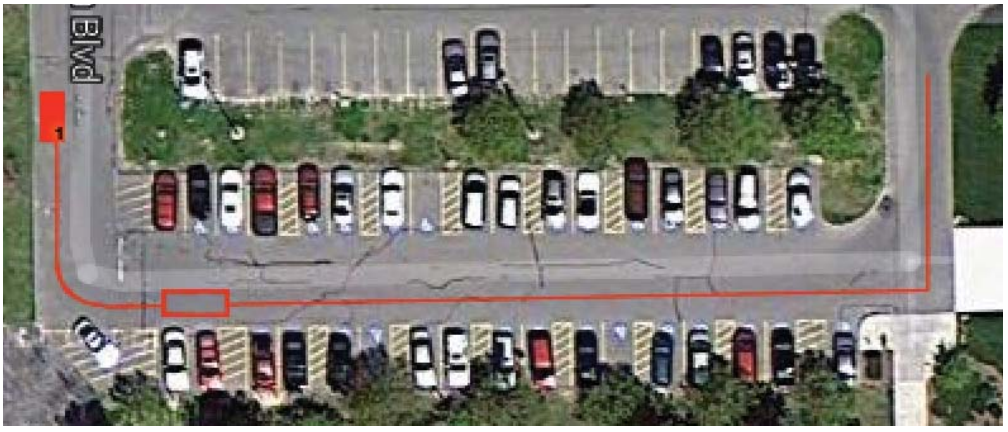


5. Begin in passenger seat of car. Exit the vehicle and walk towards Carlson to meet a subject at the corner of the field. Hand-off object. Continue walking onto the field and join other subjects in a larger group.

**Begins with subject 2

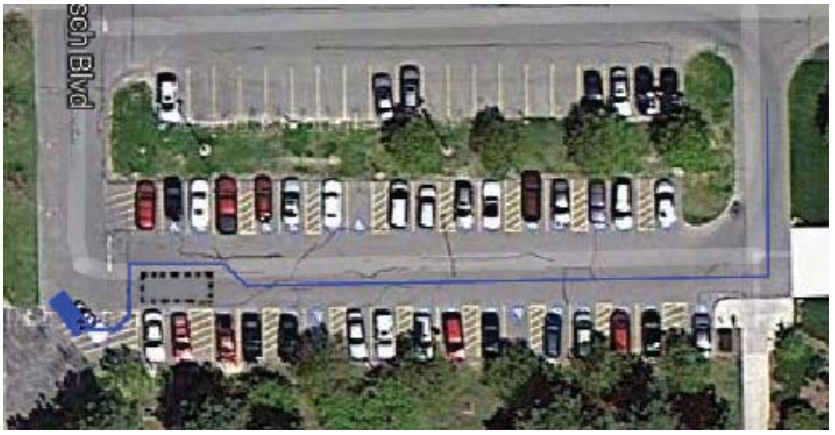
FIGURE E.1: Directions Page 3

Directions:




1. Begin at Bausch Blvd loop. Drive forward and turn into the parking lot. Pull over after passing a few cars and wait to pick up subject. Once subject is in the car, continue driving out of the loop.

**Links with subject 6




2. Begin parked at Bausch Blvd loop at the turn. Drop off subject then pull out and drive out of the loop, going around the parked car.

**Begins with subject 5

FIGURE E.2: Directions Page 1

3. Begin at edge of gravel path by Bausch and Lomb Center. Walk across the gravel sidewalk, cross large intersecting walkway, and walk over to the bottom of the field in front of Carlson. Pause and meet up with three subjects then walk together up the field until meeting up with a larger group of subjects.


4. Begin by biking down the sidewalk next to Bausch Blvd. Bike towards the back of James E. Booth. After passing the gravel pathway, turn right onto the field and meet up with three other subjects. Once larger group has begun game, move together to join them.

FIGURE E.3: Directions Page 2

6. Begin outside the overhang in the back of James E. Booth, you will be given a PVC pipe to carry. Walk up towards the parking lot with subject next to you. Upon reaching the gravel pathway, pause and place PVC pipe on shoulder. Remove pipe from shoulder and walk across the field towards the parking lot. Get into car pulled over in parking lot driven by other subject. (Make sure you enter the correct vehicle) **

**Links with subject 1

FIGURE E.4: Directions Page 4


7. Begin outside the overhang in the back of James E. Booth, with backpack in left hand. Walk up towards the parking lot with the subject next to you. Upon reaching the gravel pathway, leave the other subject and walk up diagonally to the left to meet a subject at the corner of the fields in front of Carlson. Leave backpack in middle of walkway. Continue walking together towards the center of the field to join other subjects in a larger group.

**Begins with subject 6

FIGURE E.5: Directions Page 5


9. Begin in middle of large walkway by parking lot. Walk down the path with subject next to you. Pick up backpack in middle of walkway and hold in right hand. A little after crossing the gravel pathway, turn right and walk onto the bottom of the field in front of Carlson to meet up with three other subjects. Once larger group has begun game, move together to join them.

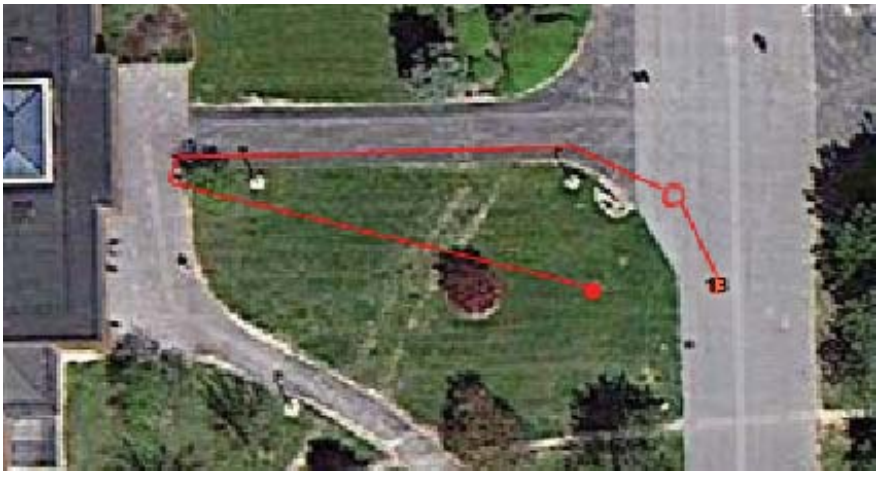
   **Begins with subject 8

FIGURE E.6: Directions Page 7


10. Begin right outside of Carlson. Walk towards parking lot and turn onto the large walkway then go towards the James E. Booth building. At the corner of the field, meet up with a small group of subjects. Walk together onto the field and join other subjects.


11. Begin in group at the corner of the field on the side walk. When three other subjects reach and join your group, walk together to the field.

FIGURE E.7: Directions Page 8




12. Begin in group on the side walk next to the field. Walk up to the corner and meet another subject that just got out of a car. Complete trade-off and continue like you are walking towards Carlson. Ensure object is being held in left hand. Before going into the building, turn around and walk onto the field and join larger group of subjects.

FIGURE E.8: Directions Page 9

# Appendix F

# Activity Analysis Interpolation Results

Picking up from the original spectral angle data in section 5.7.1.3.

From here, the zero values can be removed and actually data points connected; figure F.1

Figure F.2 depicts the two overlaid.

Performing an interpolation between the missing data points provides inter frame values. Figure F.3

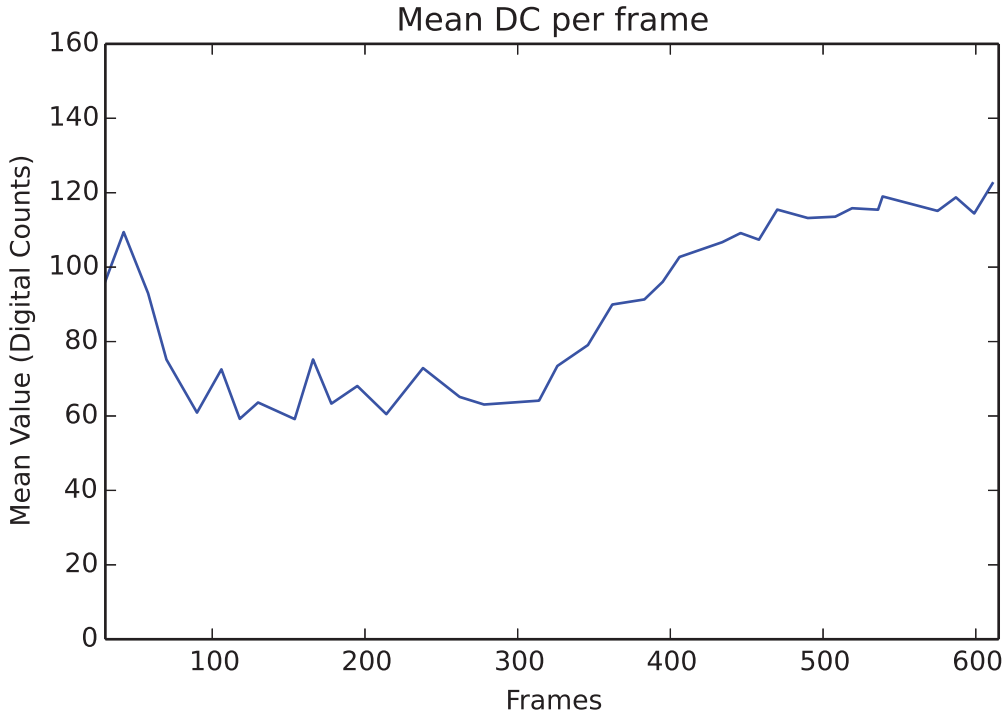



FIGURE F.1: Original Mean Digital Counts per Frame with Zeros Remove

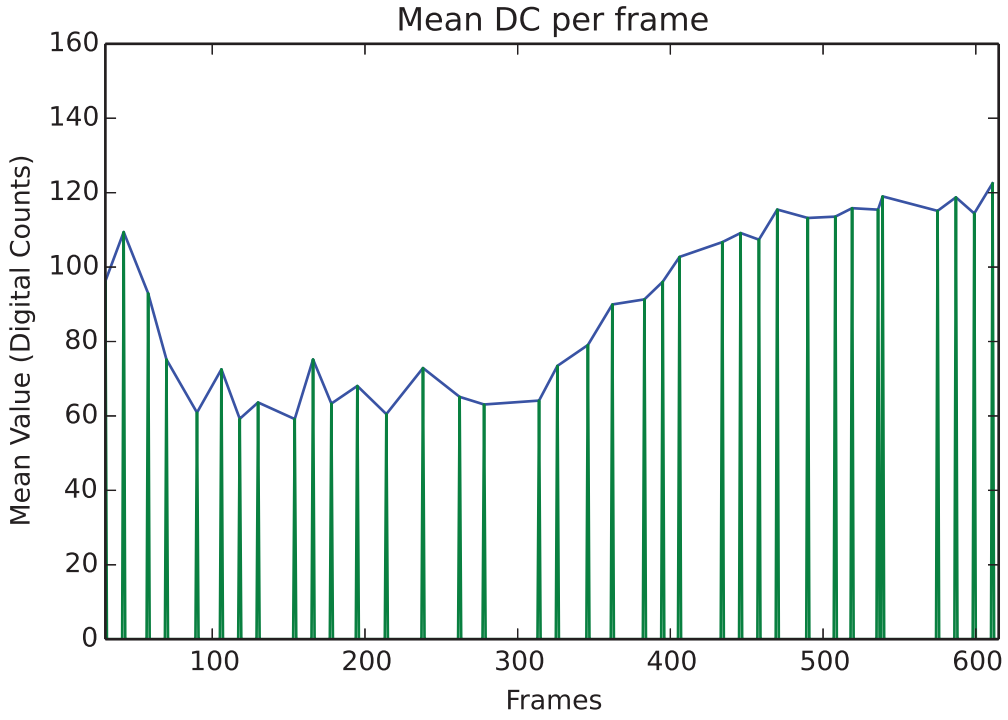



FIGURE F.2: Original Mean Digital Counts per Frame with Zeros Remove

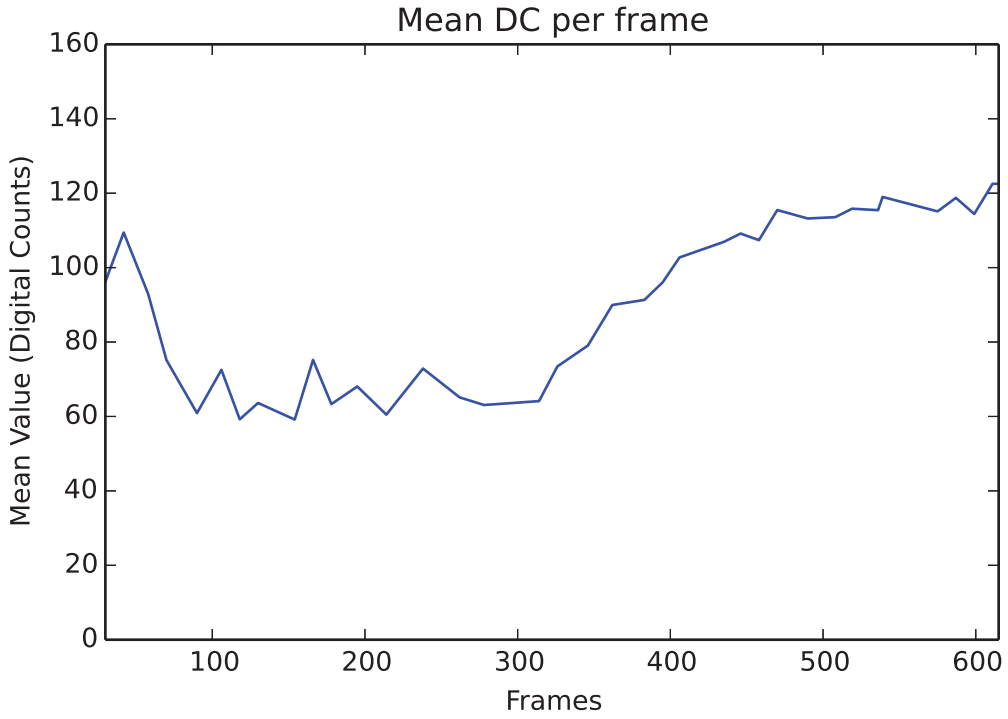



FIGURE F.3: Interpolated Mean Digital Counts per Frame

# Appendix G

# Normalized Data

Figures G.1, G.8 depict the normalized data values for the five participants included in the object exchange dataset.

```
# Array shape: (20, 18, 5)
Spatial Degradation - GSD: 5cm
1.00    1.00    1.00    1.00    1.00
0.97    0.99    0.99    0.86    1.00
0.96    0.97    0.99    0.88    0.98
0.90    0.99    0.99    0.75    0.91
0.96    0.93    1.00    0.77    1.00
0.97    0.96    0.99    0.67    0.98
0.98    0.99    0.95    0.87    0.98
0.87    0.91    0.99    0.89    0.91
0.99    0.95    0.91    0.61    0.90
0.84    0.90    0.99    0.54    0.93
0.84    0.95    0.98    0.89    0.88
0.89    0.97    1.00    0.49    0.98
0.89    0.92    0.95    0.66    0.96
0.80    0.89    0.96    0.88    0.90
0.86    0.91    0.97    0.73    0.86
0.65    0.93    0.95    0.63    0.88
0.79    0.87    0.98    0.95    0.87
0.69    0.88    0.76    0.79    0.00
Spatial Degradation - GSD: 10cm
1.00    1.00    0.99    0.98    0.88
0.97    0.99    0.99    0.82    0.88
0.97    0.96    0.97    0.86    0.86
0.93    0.99    0.99    0.73    0.80
0.94    0.92    0.99    0.74    0.87
0.96    0.95    0.99    0.62    0.86
0.98    0.99    0.94    0.84    0.87
0.84    0.90    0.99    0.87    0.79
0.98    0.94    0.90    0.57    0.82
0.82    0.89    0.98    0.51    0.82
0.84    0.95    0.99    0.83    0.80
0.87    0.96    0.99    0.47    0.86
0.90    0.91    0.93    0.62    0.87
0.79    0.89    0.97    0.83    0.79
0.88    0.91    0.95    0.70    0.78
0.66    0.93    0.94    0.58    0.76
0.78    0.86    0.98    0.96    0.78
0.69    0.87    0.75    0.82    0.00
Spatial Degradation - GSD: 15cm
0.99    0.99    0.99    0.99    0.88
0.96    0.98    0.99    0.84    0.87
0.95    0.96    0.97    0.88    0.85
0.90    0.99    0.98    0.74    0.81
0.95    0.92    0.98    0.75    0.86
0.96    0.95    0.97    0.65    0.85
0.98    0.99    0.93    0.85    0.86
0.85    0.90    0.98    0.88    0.81
0.99    0.94    0.90    0.59    0.86
0.83    0.89    0.98    0.53    0.84
```

```
0.82    0.95    0.98    0.89    0.83
0.89    0.96    0.98    0.48    0.86
0.88    0.91    0.94    0.66    0.88
0.80    0.89    0.95    0.87    0.80
0.85    0.91    0.97    0.71    0.82
0.63    0.92    0.93    0.65    0.74
0.77    0.86    0.98    0.96    0.81
0.65    0.89    0.75    0.76    0.00
Spatial Degradation - GSD: 20cm
1.00    1.00    0.97    1.00    0.81
0.97    0.99    0.96    0.84    0.81
0.98    0.97    0.96    0.87    0.78
0.91    0.99    0.96    0.74    0.76
0.95    0.92    0.96    0.74    0.78
0.94    0.95    0.97    0.63    0.78
0.97    0.99    0.93    0.83    0.79
0.81    0.90    0.96    0.89    0.76
0.98    0.94    0.89    0.56    0.79
0.81    0.90    0.96    0.50    0.81
0.83    0.95    0.96    0.85    0.73
0.89    0.97    0.97    0.46    0.79
0.91    0.91    0.92    0.64    0.78
0.79    0.90    0.94    0.84    0.76
0.86    0.91    0.95    0.70    0.75
0.66    0.93    0.91    0.63    0.71
0.75    0.86    0.97    1.00    0.74
0.67    0.88    0.72    0.81    0.00
Spatial Degradation - GSD: 25cm
0.97    0.98    0.88    0.95    0.65
0.93    0.96    0.87    0.79    0.65
0.92    0.95    0.87    0.79    0.64
0.83    0.97    0.87    0.73    0.63
0.89    0.91    0.87    0.69    0.65
0.93    0.93    0.86    0.59    0.65
0.96    0.97    0.84    0.74    0.64
0.77    0.89    0.85    0.85    0.65
0.95    0.93    0.84    0.53    0.64
0.77    0.89    0.88    0.46    0.63
0.76    0.91    0.87    0.85    0.65
0.90    0.95    0.85    0.47    0.65
0.88    0.90    0.81    0.66    0.65
0.81    0.88    0.80    0.84    0.62
0.79    0.89    0.85    0.64    0.65
0.61    0.91    0.76    0.71    0.65
0.67    0.84    0.83    0.94    0.62
0.53    0.90    0.65    0.66    0.00
Spatial Degradation - GSD: 30cm
0.99    0.99    0.94    0.99    0.73
0.95    0.97    0.93    0.82    0.73
0.96    0.96    0.93    0.85    0.72
```

```
0.89    0.99    0.93    0.74    0.70
0.92    0.92    0.93    0.73    0.72
0.93    0.94    0.92    0.61    0.71
0.97    0.98    0.90    0.80    0.72
0.77    0.90    0.94    0.88    0.72
0.99    0.94    0.88    0.55    0.72
0.77    0.89    0.94    0.48    0.72
0.81    0.94    0.91    0.86    0.68
0.90    0.96    0.93    0.47    0.69
0.90    0.91    0.91    0.66    0.72
0.80    0.89    0.89    0.84    0.69
0.84    0.91    0.93    0.68    0.69
0.64    0.92    0.86    0.66    0.67
0.72    0.85    0.92    0.98    0.67
0.63    0.89    0.69    0.77    0.00
Spatial Degradation - GSD: 35cm
0.98    0.98    0.94    0.96    0.74
0.92    0.96    0.94    0.80    0.73
0.94    0.95    0.93    0.83    0.74
0.87    0.97    0.93    0.71    0.73
0.92    0.92    0.91    0.70    0.74
0.94    0.94    0.91    0.61    0.72
0.97    0.97    0.89    0.78    0.74
0.78    0.89    0.92    0.85    0.74
0.96    0.93    0.87    0.56    0.71
0.77    0.89    0.94    0.47    0.74
0.79    0.94    0.93    0.87    0.71
0.92    0.95    0.92    0.47    0.72
0.88    0.90    0.90    0.68    0.74
0.80    0.89    0.87    0.85    0.72
0.80    0.90    0.92    0.66    0.72
0.62    0.91    0.84    0.71    0.72
0.72    0.85    0.92    0.94    0.69
0.59    0.90    0.67    0.68    0.00
Spatial Degradation - GSD: 40cm
0.98    0.98    0.91    0.97    0.70
0.93    0.96    0.90    0.81    0.69
0.94    0.95    0.90    0.82    0.69
0.86    0.98    0.91    0.73    0.69
0.91    0.91    0.90    0.71    0.69
0.93    0.94    0.89    0.61    0.68
0.96    0.98    0.87    0.77    0.69
0.77    0.89    0.90    0.86    0.68
0.97    0.94    0.86    0.55    0.67
0.76    0.89    0.90    0.47    0.69
0.78    0.93    0.90    0.86    0.67
0.90    0.96    0.90    0.47    0.68
0.88    0.91    0.86    0.67    0.69
0.80    0.89    0.84    0.83    0.69
0.81    0.90    0.89    0.66    0.68
```

```
0.62    0.92    0.81    0.69    0.69
0.69    0.85    0.87    0.97    0.66
0.57    0.89    0.65    0.71    0.00
Spatial Degradation - GSD: 45cm
0.97    0.98    0.90    0.94    0.69
0.93    0.96    0.89    0.78    0.69
0.91    0.94    0.90    0.80    0.68
0.86    0.97    0.89    0.71    0.68
0.90    0.91    0.88    0.68    0.68
0.94    0.93    0.87    0.60    0.68
0.97    0.97    0.85    0.75    0.68
0.77    0.88    0.87    0.84    0.69
0.93    0.93    0.85    0.53    0.66
0.76    0.89    0.89    0.46    0.67
0.77    0.94    0.89    0.87    0.68
0.91    0.95    0.87    0.48    0.68
0.87    0.90    0.84    0.68    0.68
0.81    0.88    0.82    0.85    0.67
0.77    0.90    0.87    0.63    0.68
0.61    0.91    0.79    0.74    0.68
0.69    0.84    0.84    0.92    0.65
0.54    0.90    0.66    0.63    0.00
Spatial Degradation - GSD: 50cm
0.97    0.98    0.88    0.95    0.65
0.93    0.96    0.87    0.79    0.65
0.92    0.95    0.87    0.79    0.64
0.83    0.97    0.87    0.73    0.63
0.89    0.91    0.87    0.69    0.65
0.93    0.93    0.86    0.59    0.65
0.96    0.97    0.84    0.74    0.64
0.77    0.89    0.85    0.85    0.65
0.95    0.93    0.84    0.53    0.64
0.77    0.89    0.88    0.46    0.63
0.76    0.91    0.87    0.85    0.65
0.90    0.95    0.85    0.47    0.65
0.88    0.90    0.81    0.66    0.65
0.81    0.88    0.80    0.84    0.62
0.79    0.89    0.85    0.64    0.65
0.61    0.91    0.76    0.71    0.65
0.67    0.84    0.83    0.94    0.62
0.53    0.90    0.65    0.66    0.00
Spatial Degradation - GSD: 55cm
0.97    0.97    0.86    0.93    0.68
0.94    0.96    0.85    0.76    0.68
0.92    0.94    0.86    0.77    0.67
0.84    0.97    0.86    0.71    0.67
0.88    0.91    0.85    0.67    0.67
0.94    0.93    0.85    0.60    0.68
0.96    0.96    0.82    0.72    0.67
0.77    0.88    0.85    0.83    0.67
```

```
0.97    0.93    0.82    0.52    0.68
0.76    0.89    0.86    0.44    0.66
0.76    0.93    0.84    0.86    0.67
0.92    0.95    0.84    0.48    0.67
0.88    0.90    0.80    0.67    0.64
0.81    0.88    0.78    0.85    0.64
0.76    0.89    0.82    0.62    0.67
0.61    0.90    0.75    0.74    0.64
0.68    0.83    0.82    0.88    0.64
0.53    0.91    0.64    0.60    0.00
Spatial Degradation - GSD: 60cm
0.96    0.97    0.84    0.93    0.67
0.94    0.96    0.82    0.77    0.67
0.90    0.94    0.84    0.78    0.67
0.83    0.97    0.84    0.71    0.66
0.89    0.90    0.84    0.68    0.65
0.93    0.92    0.84    0.62    0.67
0.96    0.97    0.79    0.70    0.66
0.77    0.88    0.83    0.84    0.67
0.96    0.93    0.79    0.52    0.67
0.77    0.89    0.84    0.45    0.64
0.73    0.90    0.81    0.85    0.66
0.89    0.94    0.82    0.48    0.66
0.87    0.90    0.77    0.66    0.63
0.82    0.88    0.76    0.84    0.63
0.77    0.89    0.81    0.62    0.66
0.59    0.91    0.72    0.72    0.63
0.68    0.83    0.81    0.91    0.63
0.51    0.91    0.61    0.62    0.00
Spatial Degradation - GSD: 65cm
0.96    0.97    0.83    0.92    0.66
0.92    0.95    0.82    0.74    0.66
0.89    0.94    0.83    0.78    0.66
0.83    0.96    0.83    0.69    0.65
0.88    0.90    0.82    0.67    0.65
0.94    0.92    0.83    0.63    0.66
0.95    0.95    0.79    0.70    0.65
0.77    0.88    0.82    0.83    0.65
0.95    0.92    0.78    0.51    0.66
0.77    0.88    0.82    0.44    0.64
0.74    0.91    0.80    0.85    0.65
0.91    0.95    0.82    0.48    0.65
0.87    0.90    0.76    0.67    0.60
0.82    0.87    0.76    0.85    0.60
0.77    0.89    0.78    0.61    0.65
0.60    0.90    0.71    0.74    0.60
0.68    0.83    0.82    0.86    0.60
0.51    0.92    0.60    0.57    0.00
Spatial Degradation - GSD: 70cm
0.96    0.97    0.82    0.92    0.63
```

```
0.93    0.96    0.81    0.75    0.63
0.90    0.94    0.82    0.77    0.63
0.83    0.97    0.82    0.71    0.62
0.88    0.90    0.81    0.68    0.63
0.94    0.92    0.81    0.63    0.63
0.96    0.96    0.77    0.70    0.63
0.76    0.88    0.81    0.84    0.63
0.93    0.93    0.76    0.51    0.62
0.77    0.88    0.80    0.45    0.61
0.73    0.90    0.78    0.84    0.63
0.89    0.93    0.81    0.49    0.63
0.88    0.90    0.75    0.67    0.58
0.83    0.87    0.75    0.85    0.58
0.78    0.88    0.77    0.61    0.63
0.60    0.90    0.68    0.74    0.58
0.69    0.83    0.81    0.89    0.59
0.50    0.92    0.58    0.59    0.00
Spatial Degradation - GSD: 75cm
0.96    0.96    0.81    0.91    0.65
0.94    0.95    0.81    0.73    0.65
0.89    0.94    0.81    0.77    0.65
0.83    0.96    0.81    0.70    0.65
0.88    0.90    0.80    0.67    0.65
0.94    0.91    0.80    0.63    0.65
0.94    0.95    0.76    0.70    0.64
0.76    0.88    0.81    0.84    0.64
0.95    0.92    0.76    0.52    0.64
0.77    0.88    0.79    0.46    0.63
0.73    0.90    0.77    0.84    0.64
0.91    0.93    0.80    0.50    0.65
0.88    0.90    0.73    0.67    0.57
0.83    0.87    0.75    0.85    0.58
0.79    0.88    0.75    0.60    0.64
0.61    0.90    0.68    0.76    0.57
0.69    0.82    0.80    0.84    0.59
0.51    0.93    0.57    0.55    0.00
Spatial Degradation - GSD: 80cm
0.96    0.97    0.79    0.91    0.62
0.94    0.96    0.79    0.74    0.62
0.90    0.94    0.78    0.77    0.62
0.84    0.96    0.79    0.71    0.62
0.88    0.90    0.79    0.68    0.62
0.95    0.92    0.78    0.63    0.62
0.96    0.96    0.75    0.70    0.61
0.76    0.88    0.79    0.84    0.62
0.94    0.93    0.74    0.52    0.61
0.77    0.88    0.78    0.46    0.60
0.72    0.90    0.76    0.84    0.61
0.90    0.93    0.79    0.51    0.61
0.89    0.89    0.71    0.67    0.56
```

```
0.83    0.87    0.74    0.85    0.56
0.79    0.88    0.74    0.60    0.62
0.61    0.90    0.66    0.76    0.56
0.70    0.82    0.79    0.87    0.57
0.51    0.93    0.55    0.58    0.00
Spatial Degradation - GSD: 85cm
0.97    0.96    0.79    0.90    0.62
0.94    0.95    0.79    0.72    0.62
0.93    0.94    0.79    0.76    0.62
0.83    0.95    0.79    0.70    0.62
0.87    0.89    0.78    0.67    0.62
0.94    0.91    0.78    0.62    0.62
0.95    0.94    0.75    0.70    0.61
0.76    0.87    0.79    0.84    0.59
0.96    0.92    0.73    0.53    0.62
0.77    0.88    0.77    0.46    0.60
0.75    0.89    0.75    0.84    0.62
0.94    0.93    0.78    0.51    0.60
0.91    0.89    0.70    0.68    0.54
0.83    0.87    0.74    0.86    0.53
0.79    0.88    0.73    0.59    0.59
0.65    0.90    0.66    0.77    0.54
0.70    0.82    0.79    0.82    0.56
0.54    0.93    0.55    0.54    0.00
Spatial Degradation - GSD: 90cm
0.96    0.96    0.78    0.91    0.61
0.94    0.95    0.78    0.73    0.61
0.91    0.94    0.77    0.75    0.61
0.83    0.95    0.78    0.71    0.61
0.88    0.90    0.76    0.68    0.61
0.95    0.91    0.77    0.62    0.61
0.95    0.95    0.74    0.70    0.60
0.76    0.87    0.78    0.85    0.59
0.95    0.93    0.72    0.53    0.60
0.78    0.88    0.76    0.47    0.59
0.72    0.89    0.74    0.84    0.60
0.90    0.92    0.77    0.52    0.59
0.90    0.89    0.69    0.68    0.53
0.84    0.87    0.73    0.85    0.54
0.80    0.88    0.72    0.60    0.58
0.62    0.89    0.66    0.77    0.53
0.70    0.82    0.78    0.86    0.56
0.52    0.93    0.54    0.57    0.00
Spatial Degradation - GSD: 95cm
0.97    0.96    0.78    0.90    0.61
0.94    0.95    0.77    0.72    0.61
0.93    0.94    0.77    0.75    0.60
0.83    0.95    0.78    0.70    0.60
0.87    0.89    0.76    0.67    0.60
0.93    0.91    0.76    0.61    0.60
```

```
0.95    0.94    0.73    0.70    0.60
0.76    0.87    0.78    0.85    0.57
0.96    0.92    0.71    0.54    0.60
0.77    0.88    0.76    0.47    0.59
0.74    0.89    0.74    0.84    0.60
0.93    0.92    0.76    0.52    0.57
0.93    0.89    0.69    0.69    0.52
0.83    0.87    0.73    0.86    0.52
0.80    0.88    0.71    0.60    0.57
0.65    0.89    0.66    0.78    0.52
0.71    0.82    0.77    0.80    0.56
0.55    0.93    0.54    0.55    0.00
Spatial Degradation - GSD: 100cm
0.96    0.96    0.76    0.91    0.58
0.94    0.96    0.76    0.73    0.58
0.91    0.94    0.75    0.76    0.57
0.83    0.95    0.76    0.71    0.57
0.87    0.91    0.74    0.68    0.56
0.95    0.91    0.76    0.61    0.57
0.95    0.95    0.73    0.71    0.56
0.77    0.87    0.76    0.87    0.55
0.95    0.92    0.70    0.53    0.56
0.77    0.88    0.74    0.48    0.57
0.72    0.89    0.73    0.84    0.56
0.90    0.92    0.76    0.54    0.56
0.92    0.89    0.67    0.69    0.50
0.84    0.87    0.73    0.85    0.49
0.81    0.88    0.70    0.62    0.56
0.63    0.89    0.64    0.78    0.49
0.71    0.82    0.76    0.84    0.54
0.53    0.93    0.53    0.58    0.00
```

# Appendix H

# SAM Code

Figurea H.1 through H.8 depict the spectral angle mapper code used to detect the object exchange in the motion imagery.

```
import os, cv2, pickle, time, copy
import numpy as np
import matplotlib.pylab as plt
import find_targets_v2 as ft
#Run using the command below this line
#normalized_data = spatio_temporal_degradation_range(temporal_kernel
#spatial_kernel_range=20,grab=None,plotting = 'no') #111 , plotting=

def sam(array1, array2):
    """
    This function calculates the spectral angle between two arrays
    """
    num = array1.T.dot( array2 )
    denom = np.sqrt( array1.T.dot( array1 ) * array2.T.dot( array2 )
    spectral_angle_mapper = np.arccos( num / denom ) * 180. / np.pi
    return spectral_angle_mapper

def tgt_spectral_mean(img,mask,coord):
    """
    Returns the mean of the pixels defined by a box
    """
    #Determine how many spectral components exist
    row,col,dim = img.shape
    #Given the target location, develop the bounding box.
    row_lower = max(coord[1]-50,0) #50
    row_upper = min(coord[1]+50, row) #80
    col_lower = max(coord[0]-30, 0) #40
    col_upper = min(coord[0]+30,col) #40
    #Mask out the specific portions
    idx=(mask==0) #Develop an index of values where the mask is zero
    img[idx]=0
    #Set all image locations to zero.
    a = img[row_lower:row_upper,col_lower:col_upper,...]
    #Block off that portion of the array indicated by the bounding b
    ##Build the spectral signature
    container = []
    #Create a container for the signatures
    for x in range(dim):
        #For each dimension of the data
        container.append( ( a[...,x].mean(), a[...,x].std() ) )
        #Place the mean and standard deviation into the container
    return container
```



FIGURE H.1: Spectral Angle Mapper Code Page 1

```
def interpolate(mean_values_of_data):
    """
    Takes the predefined mean_values and interpolates over the lengt
    sequence.
    Input:mean_values_of_data = a list of the mean data values
    Output:mean_values_of_data_interp = a list of the interpolated m
    """
    mean_values_of_data_interp = []
    #Create a holder for the interpolated values
    holder = [x for x in copy.copy(mean_values_of_data) if len(x)>0]
    #Removed any lists that don't have mean values in them
    for num, obj in enumerate(holder):
        #Cycle through each object
        if len(np.array(obj).shape)==2:
            obj = np.array(obj)[30:,...]#[:,:]
        else:
            obj = np.array(obj)[30:,:,0]#[:,:,0]
        #We only want to deal with the mean values right now. The or
        #has both mean and standard deviation
        frames, spectrum = obj[...,:8].shape
        #Grab the number of frames and spectral dimensions; in this
        #only have eight, but may have holders with nine
        length = range(frames)
        #Create a list of numbers counting off the frames
        for lens in range(spectrum):
            #Cycle through each spectrum of the object
            b = [x for x in zip( length, obj[:,lens] ) if x[1]>0]
            #Zip the frame numbers and their corresponding data toge
            #Only retain data that has a value greater than zero. Ma
            #without WASP data have placeholders of zero.
            if np.array(b).shape[0]==0:
                #If no data exists in this band, move to the next ba
                continue
            else:
                yinterp = np.interp(length, np.array(b)[:,0], np.arr
                #Interpolate the missing values
                obj[:,lens] = yinterp
                #Replace the original object data with the interpola
        mean_values_of_data_interp.append( obj )
        #Add this object to the list of interpolated data
    return mean_values_of_data_interp

def sam_from_spectral(mean_values_of_data_interp):
```



FIGURE H.2: Spectral Angle Mapper Code Page 2

```
        """
        Composes the Spectral Angle for each object on a frame-by-frame
        """
        #Average the first :xxx spectral signatures
        spectral_values = []
        for x in (mean_values_of_data_interp):
            #For each object
            mini_spec = []
            #Create a holder for the frame-by-frame SAM values
            top = len(x) * 0.1
            first = (np.array(x)[:top,:]).mean(axis=0)
            #Populate a spectral reference by averaging the first: xxx s
            for y in x:
                #Evaluate over each frame of interpolated mean value dat
                y = np.array(y)
                #Create an array of the list
                frame_SAM = sam(first,y)
                #Calculate the spectral angle for each frame
                #Note: the multiplication simply remaps from 0-1 to 0-10
                if np.isnan(frame_SAM):
                    frame_SAM = 0
                mini_spec.append( frame_SAM )
                #Add this objects frame SAM value to a list for later
            spectral_values.append( mini_spec )
            #Compile all this objects' SAM values into a list
        return spectral_values

    def mean_values_spatial_degrade(data,flist_raw_full,avg,blur_kernel)
        """
        Calculate the mean spectral vectors of each object for each fram
        Inputs
        ------
        data = locations of targets within each frame in a dictionary wi
        dictionary format. The top dictionary has keys associated with
        the objects, and each value is another dictionary. The second
        dictionary uses the the frame numbers as keys indicated as:
        "frame_#". The values are tuple pair of (x,y) coordinates of the
        target.
        flist_raw_full = a list ofimages with full path lengths
        avg = background image of the sequence
        blur_kernel = size of the blur_kernel in tuple format
        Outputs
        -------
```



FIGURE H.3: Spectral Angle Mapper Code Page 3

```
        mean_values_of_data = list of numpy arrays containing the mean..
        """
        #Grab the mean values of all the data by calculating the masks
        mean_values_of_data = []
        for obj in sorted(data.keys()):
            #Cycle through the objects
            mean_val = []
            for key in sorted(data[obj].keys()):
                #Cycle through the target locations associated with each
                img_num = int((key.split('_'))[1]) - 1000
                #Remap the frame key to the numbering system of the save
                image = np.load( flist_raw_full[ img_num ] )
                #Call the image associated with the frame
                thresh = ft.target_detection(np.uint8(image[...,:3]), av
                plot='no',grab='thresh')
                #Calculated the threshold image using the target detecti
                mask = masking(thresh)
                #Turn the target detection image into a mask for the dat
                coord = data[obj][key]
                #Grab the coordinate of this object within this frame
                blur_image = cv2.blur(image,(blur_kernel,blur_kernel))
                #Apply a blur to the image for reduced spatial resolutic
                spec_mean = tgt_spectral_mean( blur_image, mask, coord[:
                #Send the blurred image, mask, and target coordinates in
                #function to find the band means
                #Note: the image coordinates needed to be reversed
                #(i.e.(x,y)[::-1] = (y,x))
                mean_val.append( spec_mean )
                #Place this frames spectral mean into a list for later
            mean_values_of_data.append( mean_val )
            #Place this objects' frame-by-frame spectral mean into a lis
        mean_values_of_data_keep = [x for x in mean_values_of_data if le
        #If any of the mean value data is an empty list
        return mean_values_of_data_keep

    def AuC_from_SAM(spectral_angles, temporal_blur):
        """
        Taking the spectral angles, this function calculates the thresho
        top down area under the curve encompasses 10% of the number of f
        sequence.
        Inputs
        ------
        spectral_angles = a list of spectral angle values for each obj

                                                                    4
```

FIGURE H.4: Spectral Angle Mapper Code Page 4

```
        Output
        ------
        AuC_value = the angle at which at least 10% of the frames above
        """
        length = len(spectral_angles)
        #Determine how many spectral components there are
        AuC_value = []
        #Create holder for probabilites
        for x in range(length):
            #Cycle through the spectral dimensions
            holder = np.array(spectral_angles[x])
            holder_limit = holder[int(150./temporal_blur):int(650./tempo
            #Only review the portion of the experiment which houses the
            length = len(holder) * 0.1
            #Determine how many frames make up 10% percent of the data
            Thresh = []
            #Create a holder for threshold values
            top =  holder_limit.max() + 1
            if np.isnan(top):
                #If a 'nan' gets through replace it with 1
                top = np.float64(1)
            #Determine the upper limit of the angular disparity
            x_vals = np.linspace(0, top, top*10, endpoint=True)
            #Create a linespace of angular values
            #This determines how accurately you can relate the number of
            #the spectral angle
            for x in x_vals:
                #For each value in the linespace, calculate the area und
                Thresh_calc = len( holder_limit[holder_limit>x] )
                #Calculate the number of frames above the value x in the
                Thresh.append( Thresh_calc )
                #Place this number in a container
            if np.array(Thresh).max()!=0:
                #If the max value is not 0 enter if statement
                pair = zip( x_vals, Thresh )
                #Pair the linespace with the area under the curve calcul
                value=np.array(pair)[::-1][:,0][np.array(pair)[::-1][:,1
            else:
                value = 0
            AuC_value.append( value )
        return AuC_value

    def normalize_probabilities_from_AuC(AuC_ranges):

                                                                    5
```

Figure H.5: Spectral Angle Mapper Code Page 5

```
"""
Normalize  spectral angles by the spectral angle of the fr senso
If greater than one reduce by overage.
Inputs: probabilities = spectral angles of all the the
Outputs: Normalized data
"""
D = np.array( copy.copy(AuC_ranges) )
spatial_norms = 1.0 * D[:,0,:] / D[0,0,:]
#Normalize by spatial dimension

##Correct any normalized values over one by indicating their hig
#probability as a overage to be reduced.
C = spatial_norms.T
#Transpose to use in 'for' operation (Objects are now along rows
#temporal data along columns)
for obj_num, x in enumerate(C):
    #For each object
    x[x>x[0]] = x[0] + x[0] - x[x>x[0]]
    #If there are spectral angles, above the spectal angle at th
    #temporal resolution, subtract the amount above
    #the base amount (i.e. 5 + (5 - 5.6))
    C[obj_num,:] = x
    #Replace in array
spatial_norms = C.T
#Undo previously applied transpose
for spatial_num, C in enumerate(D):
    C = 1.0 * C / C[0,:]
    #Normalize them by highest temporal resolution
    #print "Spatial Degrade {0}:\n{1}\n".format(spatial_num,C)
    C = C.T
    #Transpose to use in 'for' operation (Objects are now along
    #temporal data along columns)
    for obj_num,x in enumerate(C):
        #For each object
        x[x>x[0]] = x[0] + x[0] - x[x>x[0]]
        #If there are spectral angles, above the spectal angle a
        #highest temporal resolution, subtract the amount above
        #the base amount (i.e. 5 + (5 - 5.6))
        x[x<0] = 0
        #If there are negative probabilites, reduce them to zero
        C[obj_num,:] = x
        #Replace in array
    C = C.T
```

6

FIGURE H.6: Spectral Angle Mapper Code Page 6

```
            #Undo previously applied transpose
            D[spatial_num,...] = C
        E = D * spatial_norms[:,np.newaxis,:]
        #Apply the normalized probabilities to the remainder of the data
        normalized_data = E
        return normalized_data

    def spatio_temporal_degradation_range(temporal_kernel_range=1,
    spatial_kernel_range=1, plotting='no', grab=None):
        """
        Calls the above functions
        Inputs
        ------
        temporal_kernel_range = top end of temporal kernel eval range
        spatial_kernel_range = top end of spatial kernel eval range
        plotting = plotting option for end results | default = 'yes'
        grab = return output of intermediate steps
        evaluating the entire function | default = None,
        options = "mean", "interp", "temporal", "SAM", "AuC, and 'Pre-No
        Outputs
        -------
        Displays a plot and saves an eps figure of the results for each
        size
        Spectral_mean_blur_{} file for each set of means developed
        """
        flist_raw_full, avg, data = load_data()
        spatial_AuC_range = []
        for spatial_blur_kernel in xrange(1,spatial_kernel_range+1):
            fname = 'Spectral_mean_blur_{}.p'.format(spatial_blur_kernel
            #Develop the file naming scheme
            if os.path.isfile(fname):
                #If the file exists, open it and use the data in the seq
                with open(fname,'r') as f:
                    mean_values_of_data = pickle.load(f)
                if grab == 'mean':
                    return mean_values_of_data
            else:
                #If the file does not exist, develop it.
                mean_values_of_data = mean_values_spatial_degrade(data,
                flist_raw_full, avg, spatial_blur_kernel)
                #Spatially degrade the data

                #Save the dictionary of dictionaries
```



FIGURE H.7: Spectral Angle Mapper Code Page 7

```
            with open(fname,'w') as f:
                pickle.dump(mean_values_of_data,f)
            if grab == 'mean':
                return mean_values_of_data
        mean_values_of_data_interp = interpolate(mean_values_of_data
        #Interpolate the missing spectral data
        if grab == 'interp':
            return mean_values_of_data_interp
        temporal_AuC_range = []
        Rates = range(1,13)
        for x in [15, 20, 24, 30, 40, 60]:#, 120, 180 ]:
            Rates.append( x )
        for temporal_blur_kernel in Rates:
            #Evaluate the data through the blur ranges suggested
            temporal_mean_values_of_data = [x[::temporal_blur_kernel
            for x in mean_values_of_data_interp]
            #Temporally degrade the data
            if grab == 'temporal':
                return mean_values_of_data_interp
            SAM_values = sam_from_spectral(temporal_mean_values_of_d
            #Assess the per frame spectral angle of the data
            if grab == 'SAM':
                return SAM_values
            if plotting == 'yes':
            #Plot the data
                plotting_spatio_temp_dat(avg, SAM_values, temporal_b
                spatial_blur_kernel)
            AuC = AuC_from_SAM(SAM_values, temporal_blur_kernel)
            #Determine the probability of detecting the exchange
            if grab == 'AuC':
                return AuC
            temporal_AuC_range.append( AuC )
            #print len(temporal_AuC_range)
        spatial_AuC_range.append( temporal_AuC_range )
    if grab == 'Pre-Norm':
        return spatial_AuC_range
    normalized_data = normalize_probabilities_from_AuC( spatial_AuC_
    prob_plotting( avg, normalized_data, Rates, spatial_kernel_range
    return normalized_data
```

8

FIGURE H.8: Spectral Angle Mapper Code Page 8