



**ON THE INTEGRATION OF MEDIUM WAVE INFRARED CAMERAS FOR
VISION-BASED NAVIGATION**

THESIS

Justin S. Tharp, Second Lieutenant, USAF

AFIT-ENG-MS-15-M-063

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

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

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENG-MS-15-M-063

ON THE INTEGRATION OF MEDIUM WAVE INFRARED CAMERAS FOR
VISION-BASED NAVIGATION

THESIS

Presented to the Faculty
Department of Electrical and Computer Engineering
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Electrical Engineering

Justin S. Tharp, B.S.E.E.

Second Lieutenant, USAF

March 2015

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

AFIT-ENG-MS-15-M-063

ON THE INTEGRATION OF MEDIUM WAVE INFRARED CAMERAS FOR
VISION-BASED NAVIGATION

THESIS

Justin S. Tharp, B.S.E.E.
Second Lieutenant, USAF

Committee Membership:

Maj Brian G Woolley, PhD
Chair
John F Raquet, PhD
Member
Kyle J Kauffman, PhD
Member

Abstract

The ubiquitous nature of GPS has fostered its widespread integration of navigation into a variety of applications, both civilian and military. One alternative to ensure continued flight operations in GPS-denied environments is vision-aided navigation, an approach that combines visual cues from a camera with an inertial measurement unit (IMU) to estimate the navigation states of a moving body. The majority of vision-based navigation research has been conducted in the electro-optical (EO) spectrum, which experiences limited operation in certain environments. The aim of this work is to explore how such approaches extend to infrared imaging sensors. In particular, it examines the ability of medium-wave infrared (MWIR) imagery, which is capable of operating at night and with increased vision through smoke, to expand the breadth of operations that can be supported by vision-aided navigation. The experiments presented here are based on the Minor Area Motion Imagery (MAMI) dataset that recorded GPS data, inertial measurements, EO imagery, and MWIR imagery captured during flights over Wright-Patterson Air Force Base. The approach applied here combines inertial measurements with EO position estimates from the structure from motion (SfM) algorithm. Although precision timing was not available for the MWIR imagery, the EO-based results of the scene demonstrate that trajectory estimates from SfM offer a significant increase in navigation accuracy when combined with inertial data over using an IMU alone. Results also demonstrated that MWIR-based positions solutions provide a similar trajectory reconstruction to EO-based solutions for the same scenes. While the MWIR imagery and the IMU could not be combined directly, through comparison to the combined solution using EO data the conclusion here is that MWIR imagery (with its unique phenomenologies) is capable of expanding the operating envelope of vision-aided navigation.

I dedicate this research to my grandfather who spent our time together nurturing and inspiring me to accomplish my dreams.

Acknowledgments

I want to thank my wife for understanding the time it took me to do this research, my parents for always believing in me, and all of my past teachers who have given me the opportunities to undertake greater challenges than I would have ever imagined.

Justin S. Tharp

Table of Contents

	Page
Abstract	iv
Dedication	v
Acknowledgments	vi
Table of Contents	vii
List of Figures	viii
List of Acronyms	ix
I. Introduction	1
1.1 Problem Definition	2
1.2 Research Contributions	4
1.3 Thesis Outline	7
II. Background	8
2.1 Navigation	8
2.1.1 Coordinate Frames	9
2.1.2 World Model	11
2.1.3 Inertial Measurement Unit (IMU)	13
2.1.4 Global Positioning System (GPS)	15
2.2 Integrated Navigation Solutions	16
2.3 Computer Vision	18
2.3.1 Camera Model and Calibration	18
2.3.2 Feature Detection and Extraction	21
2.4 Structure from Motion (SfM)	23
2.4.1 Bundle Adjustment	24
2.4.2 RANSAC	27
2.5 Vision Aided Navigation	27
III. Structure from Motion and Inertial Measurement Unit (IMU) Combined Solution	29
3.1 Dataset	29
3.2 Equipment	30

	Page
3.3 Assumptions	30
3.4 Navigation Solution	31
3.4.1 IMU Simulation from Inertial Navigation System (INS) solution . .	31
3.4.2 Sensor Processing and Inertial Dynamics Error Reduction (SPIDER)	33
3.4.3 Navigation State Estimation	33
3.5 Visual Structure from Motion	35
3.6 SfM Measurements	37
3.6.1 Structure from Motion (SfM) Alignment	37
3.6.2 SfM Trajectory	38
3.6.3 SfM Velocity Measurements	39
3.6.4 Noise Modeling	40
3.7 SfM Aiding Results	43
3.8 Weaknesses of Visible Light Cameras	53
 IV. Medium Wave Infrared (MWIR) Cameras for Navigation	 54
4.1 Equipment	54
4.2 Dataset	55
4.3 Infrared Imaging	55
4.3.1 Construction of Infrared Images	55
4.3.2 Unique Infrared Characteristics	57
4.3.3 Medium Wave Infrared (MWIR)	59
4.4 Contrast Enhancement	60
4.5 Medium Wave Infrared (MWIR) Image Comparisons	61
4.5.1 Vegetation	61
4.5.2 Artificial Structures	64
4.6 SfM Comparisons	64
4.6.1 Point Cloud Alignment	64
4.6.2 Further MWIR Characteristic Analysis	67
4.6.3 SfM Solutions	67
4.6.4 MWIR Night Imagery	71
4.7 MWIR Navigation Viability	75
 V. Conclusions and Future Work	 77
5.1 Conclusions	77
5.2 Future Work	79
 Bibliography	 81

List of Figures

Figure	Page
1.1 IMU Solution Example	4
1.2 2D Navigation Scenario	5
2.1 World Coordinate Frames	10
2.2 DCM Example	12
2.3 IMU Measurements	13
2.4 Pinhole Camera Model	19
2.5 Image Calibration Example	20
2.6 SIFT Feature Example	22
2.7 SfM Position and Scene Reconstruction Example	25
3.1 IMU Solution Over 10 Minutes	34
3.2 SfM Alignment Errors	39
3.3 Alignment Position Error	39
3.4 SfM Alignment Errors	40
3.5 Alignment Velocity Error	40
3.6 SfM Only Position Error	41
3.7 Latitude, Longitude, and Altitude Error of SfM Solution	41
3.8 SfM Velocity Errors	42
3.9 Alignment Velocity Error	42
3.10 Latitude, Longitude, and Altitude Of All Solutions	44
3.11 Position Error with Confidence	45
3.12 Position Error with Confidence Zoomed In	46
3.13 Position Error with Variance over 250 Simulations	47
3.14 Velocity Of All Solutions	48

Figure	Page
3.15 Velocity Error Of All Solutions	49
3.16 Velocity Error Over 250 Simulations with Variance	50
3.17 Average Position Error Over 250 Simulations	51
3.18 Average Position Error Over 250 Simulations Zoomed In	52
4.1 Atmospheric Transmittance	56
4.2 Different Sides of Thermal Crossover	58
4.3 Histogram Spreading Example	62
4.4 Comparison of Vegetation Appearance	63
4.5 Comparison of Artificial Structures	65
4.6 Comparison of Asphalt Appearance	66
4.7 Comparison of Vegetation Point Reconstruction	68
4.8 Comparison of Building Point Reconstruction	69
4.9 Comparison of Road Point Reconstruction	70
4.10 Comparison of Aligned SfM Solutions MSEE 1	72
4.11 Comparison of Aligned SfM Solutions DEBU 2	73
4.12 Comparison of Day and Night MWIR Images	74
4.13 MWIR Night Trajectory	76

List of Acronyms

Acronym	Definition
AFRL	Air Force Research Laboratory
DCM	Direction Cosine Matrix
DTED	Digital Terrain and Elevation Database
ECEF	Earth Centered Earth Fixed
EKF	Extended Kalman Filter
ENU	East North Up
GPS	Global Positioning System
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
LWIR	Long Wave Infrared
MAMI	Minor Area Motion Imagery
MMSE	Minimum Mean Squared Error
MWIR	Medium Wave Infrared
NED	North East Down
NIR	Near Infrared
RANSAC	Random Sample Consensus
SfM	Structure from Motion
SIFT	Scale Invariant Feature Transform
SPIDER	Sensor Processing and Inertial Dynamics Error Reduction
SWIR	Short Wave Infrared
VisualSfM	Visual Structure from Motion
WPAFB	Wright Patterson Air Force Base

ON THE INTEGRATION OF MEDIUM WAVE INFRARED CAMERAS FOR VISION-BASED NAVIGATION

I. Introduction

The ability to navigate is a critical capability for nearly all aspects of military and intelligence operations. In particular, airborne operations require highly accurate position and orientation information which they not only use to navigate, but also to base sensing measurements on. Many current navigation solutions combine Global Positioning System (GPS) signals with an Inertial Measurement Unit (IMU). Interestingly, this pairing is a good fit because GPS provides accurate, if not precise, position information that does not drift over time while the IMU provides very precise acceleration and rotation updates over short periods of time. Without GPS however, solutions using only an IMU to navigate will experience compounding errors causing the position solution to drift away from the truth. Given the ease with which GPS signals can be disrupted, either intentionally or unintentionally, what can be done to ensure robust navigation without external signaling (e.g., from GPS)?

One alternative method of aiding inertial sensors that has gained recent interest is using cameras on board the aircraft to provide additional navigation information from observed motion of ground targets [1]. This method of navigation aiding is self-contained making it highly resistant to both hostile and accidental interference. Current research into vision aided navigation has focused on Electro-Optical (EO) cameras that sense light in the visible spectrum [2] [1]. The aim of this thesis is to explore how current vision aided navigation techniques might perform in other bands. Specifically, this work focuses on Medium Wave

Infrared Cameras (MWIR) as an alternative vision source due to their ability to operate at night and penetrate through smoke, clouds, and fog.

MWIR imaging systems have limits that affect their ability to aid in navigation. Cameras that sense the infrared band tend to have less resolution than similar EO cameras, blurring effects appear when sensing certain objects, and they are much more expensive than their EO counterparts. The strengths of the domain that make it useful as a navigation tool mentioned before are increased visibility in low light and smoke occluded environments, areas that the visible spectrum struggles to cope with.

Outages in vision sensors aiding navigation solutions quickly render a position solution no longer usable for operation. Vision-aided navigation, in the context of this research, only provides relative motion measurements which cannot correct previous drifts in error by an IMU. This makes the system highly dependent on the robustness of the camera updates. The strengths of MWIR listed above are why it should be explored as an alternative to EO spectrum cameras. This added robustness could prove essential for applications in military aircraft, especially those that operate at night and during other adverse conditions.

1.1 Problem Definition

The experiments presented utilized images from the Minor Area Motion Imagery (MAMI) data collect conducted by Air Force Research Laboratory (AFRL) over Wright-Patterson Air Force Base (WPAFB) in 2013. The aircraft used to collect imagery was equipped with both EO and Infrared (IR) cameras mounted onto side-looking gimbals. The system was instrumented with GPS and IMU sensors on each camera gimbal assembly to track their position and orientation over time. The MAMI data set contains both day and night flights.

Although GPS data was collected throughout all portions of the MAMI data collect, the experiments here are designed to characterize vision aided navigation solutions without

the aid of GPS. In particular, the scenario of interest is the case where an aircraft is navigating with a GPS signal which it suddenly loses. The goal of this research is to show that during the GPS outage, combining the inertial sensor with measurements from an image alignment algorithm called Structure from Motion (SfM) increases the quality of the navigation solution compared to letting the IMU run freely.

IMU sensors have small errors over short times that grow larger as random biases in the accelerometers and gyroscopes inside the sensor causing the solution to drift away from the truth at an exponentially increasing rate. GPS gives a position update to the navigation solution that may be up to a few meters off, but does not drift over time and is an absolute measurement. This allows a GPS and IMU coupled system to be accurate over very long times as the system can estimate the growing biases in the IMU.

Structure from Motion (SfM) is a computer vision tool that estimates the change in position and pointing angle of a camera between each image fed into it of a scene. This change in position, or velocity if divided by the time between frames, is a relative estimate that has noise depending on factors such as the number of features detected in the images, the distance to objects in the scene, and the resolution of the camera among others. The directions of these updates will drift away from the truth as the alignment between the SfM solution and a real world reference frame degrades over time. The drift in alignment between the SfM frame and the real world is different from the growing biases in the IMU so the SfM updates give observability of these IMU biases. Because neither sensor gives an absolute measurement, the system will ultimately drift far enough away that the solution exceeds a threshold of accuracy for operation.

If we compare notional navigation solutions over time through a two dimensional space we can see how various navigation approaches perform over time (Figure 1.2). All of the approaches begin with a small amount of error that grows differently over time. After a certain point the solutions without GPS corrections lose accuracy and diverge from the

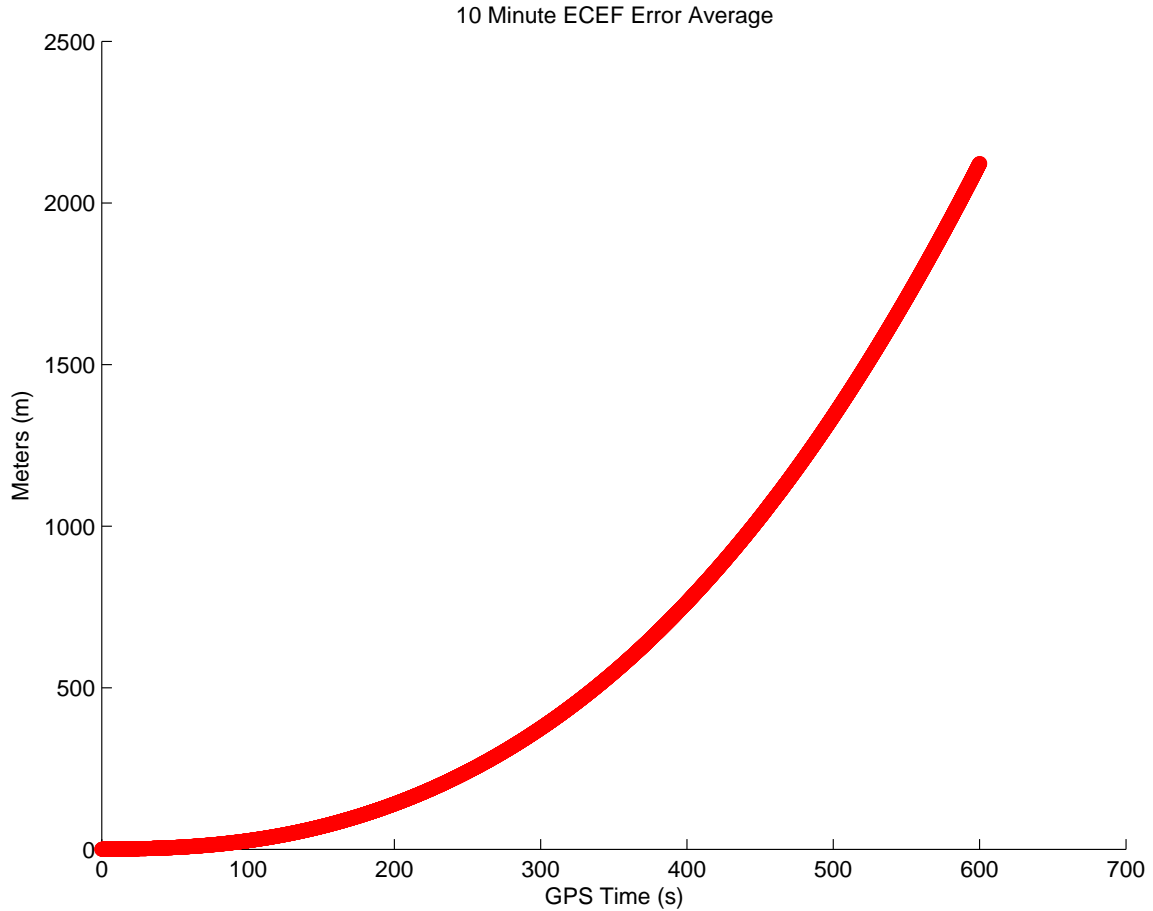


Figure 1.1: IMU Solution Example. An example of position error in a navigation solution using only a tactical grade IMU sensor. The error starts very low but grows exponentially over the 10 minute period. The plot represents the average error of 250 different IMU simulations.

truth. Proving that the SfM aided solution follows the truth more closely than the unaided solution is an important part of this research. The level of improvement is relative to the quality of sensors used for both the camera and inertial sensor, as well as the fitness of the environment for the camera solution.

1.2 Research Contributions

Although not measured directly, this thesis argues that SfM position estimates from MWIR imagery can be combined with an IMU to create an improved navigation solution. Unfortunately, the MAMI dataset did not have precision timing available for the MWIR

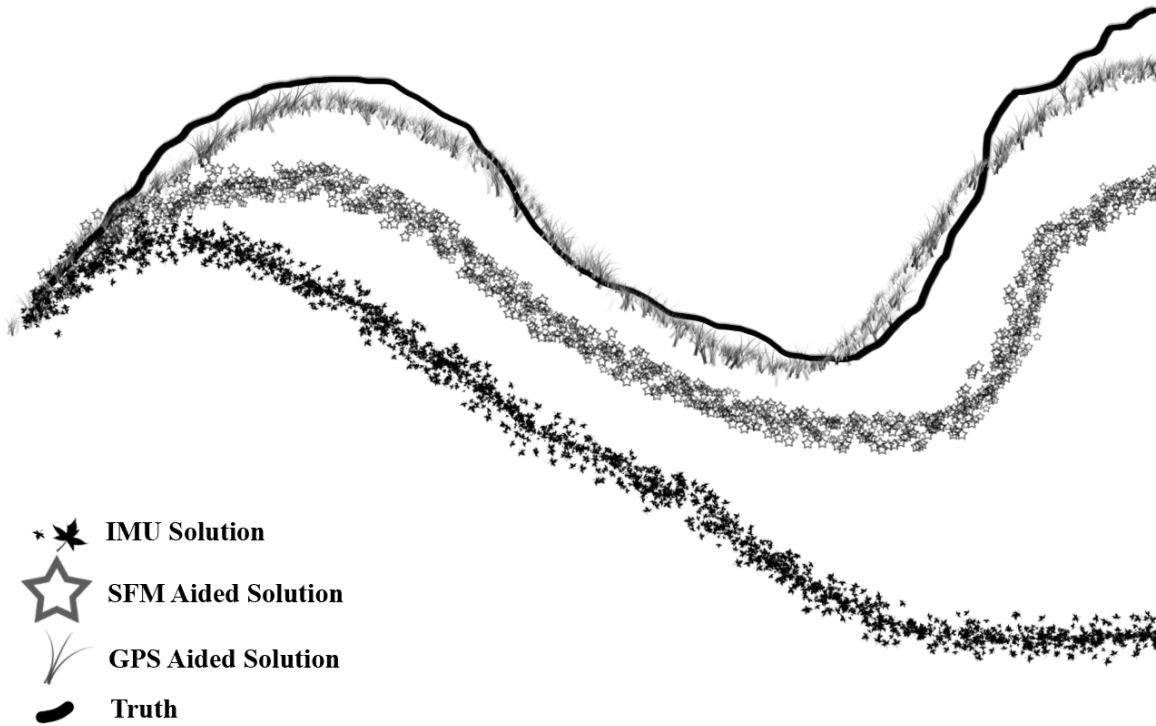


Figure 1.2: 2D Navigation Scenario. A qualitative view of the expected results of solution types used in this research for 2 dimensional navigation. The IMU solution, consisting of an unaided inertial sensor, drifts the fastest away from the truth. The GPS aided solution, a combination of GPS signals and an inertial sensor, tracks the truth most accurately. The Structure from Motion (SfM) aided solution, a combination of navigation information derived from SfM using EO images with inertial data, is somewhere in between the previous two solutions in terms of following the truth. Over a longer time, the SfM aided and IMU only solutions would continue to diverge while the GPS solution stays close to the truth.

imagery, thus it was not possible to directly combine the SfM MWIR position estimates with the IMU. In spite of this limitation, this thesis makes the following transitive argument: Given that SfM position estimates in the EO spectrum combined with IMU updates offer a significant navigation accuracy increase over a free running IMU solution, and that SfM position estimates from both EO and MWIR show similar trajectory reconstruction over the same scene, it can be assumed that MWIR-based SfM position updates will result in improved navigation solutions when combined with IMU updates.

The first part of this argument, i.e. combining velocity measurements from SfM with an IMU, is conducted with the MATLAB program called SPIDER[3] to incorporate

the sensor measurements into an estimate of the position, velocity, and orientation of the aircraft over time. Developed by the Autonomy and Navigation Technology (ANT) center at the Air Force Institute of Technology (AFIT), this program is a robust navigation tool that produces optimal estimates of navigation states using an easily customizable set of sensors. This research looked at the quality of navigation solution created by combining velocity measurements from SfM with inertial sensor updates in SPIDER to provide a more accurate solution over using only an unaided IMU.

The second part of the argument shows that the SfM position solutions, as well as the sparse point reconstructions of the observed scene, generated from EO and MWIR imagery are similar in quality. Interestingly, although the two sensors fundamentally differ in what they observe from the environment, the resulting sparse point clouds can be aligned for comparison because man-made structures and roads appear clearly, albeit characterized differently, in the reconstructions of both spectrums. The experimental results here seek to demonstrate that both phenomenologies are capable of similar quality position estimates to conclude that MWIR imagery can be combined with IMU updates to provide similar magnitudes of aiding seen with the EO navigation updates.

While this work argues for the integration of MWIR imagery into navigation solutions, it is not a panacea. It is important that practitioners understand the strengths and weaknesses of the EO and MWIR spectrums. The obvious advantage of MWIR imagery over EO cameras is its ability to penetrate through smoke and function at night. However, infrared imagery tends to have low contrast, making enhancement necessary to elicit a usable level of feature detection for navigation. Another limitation of MWIR imagery for navigation is that areas of vegetation tend to be featureless due to homogeneous temperatures. Thus, MWIR cameras are another potential tool to aid in navigation that offer large benefits in specific adverse situations.

1.3 Thesis Outline

The remainder of this thesis is structured as follows: Chapter 2 gives the background necessary to understand the key topics on which the experiments and approaches are based. In Chapter 3, the combination of EO SfM and IMU measurements is demonstrated in an improved navigation solution. Next, Chapter 4 discusses MWIR phenomenologies and compares SfM navigation solutions generated by MWIR and EO imagery. This work concludes with a discussion of the findings followed by suggestions for future research.

II. Background

This chapter introduces the key concepts on which the vision aided navigation algorithms used in this research are based. The discussion begins with the art of navigation including coordinate frames, inertial measurement units, and the global positioning system. Next, the combination of navigation sensors is explained. Lastly, key topics in the field of computer vision are covered.

2.1 Navigation

The motion of an aircraft can be tracked given sensor updates that measure either the *absolute values* of or *changes in* its navigation states given initial values. These states include the position, velocity, and orientation in each of the three dimensions of the body at a given time. These nine states capture the navigation information used in many modern systems today.

$$\mathbf{Nav} = \begin{bmatrix} \text{pos} \\ \text{vel} \\ \text{angle} \end{bmatrix} \quad (2.1)$$
$$\mathbf{pos} = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} \quad (2.2)$$
$$\mathbf{vel} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \quad (2.3)$$
$$\mathbf{angle} = \begin{bmatrix} \text{roll} \\ \text{pitch} \\ \text{yaw} \end{bmatrix} \quad (2.4)$$

Equations 2.1 to 2.4 above show the matrix representation of the nine navigation states. The determination of the x, y, and z directions are dependent upon the coordinate frame used to describe the navigation solution.

The conventional method of using an IMU coupled with a Global Positioning System (GPS) deals with the problem of measuring navigation states effectively over both short and long term applications. The problem addressed in this research is how to track these

navigation states in the absence, or under malicious interference, of GPS — which is a large concern for military operators.

2.1.1 Coordinate Frames.

The position and orientation of an object must be given in relation to an established frame of reference to utilize in real world applications. In navigation, there are established frames which define coordinates in terms of measurable points in the universe. All of these frames follow the right hand rule making 90 degree angles between each axis. The inertial frame (i) uses the center of the earth as the origin and maintains a constant direction for the x and y axes in terms of distant astral bodies [4]. The Earth Centered Earth Fixed (ECEF) coordinate frame (e) uses the same origin as the inertial frame but with a fixed x axis coming out of the equator at the Greenwich meridian and a z axis pointing out towards the north pole. Moving closer to a tracked object along the Earth's surface, the Navigation frame (n), while having many possible interpretations, can be considered as being along the surface of the Earth close to the object of interest with one axis pointing north, one east, and the last down to give the North East Down (NED) convention or one axis pointing east, one north, and one up to give the East North Up (ENU) convention. The body frame (b) is directly attached to the vehicle at some decided point and is initialized according to the NED or ENU conventions.

The interactions between these frames are important as different sensors give information about navigation in relation to the different frames that they reference. For example, GPS solutions provide absolute position estimates in the ECEF frame whereas the IMU gives relative motion and rotation in the body frame. In order to use multiple sensors to give navigation information about a vehicle, the respective position and orientation information coming from the sensors must all be rotated and translated into a similar reference frame.

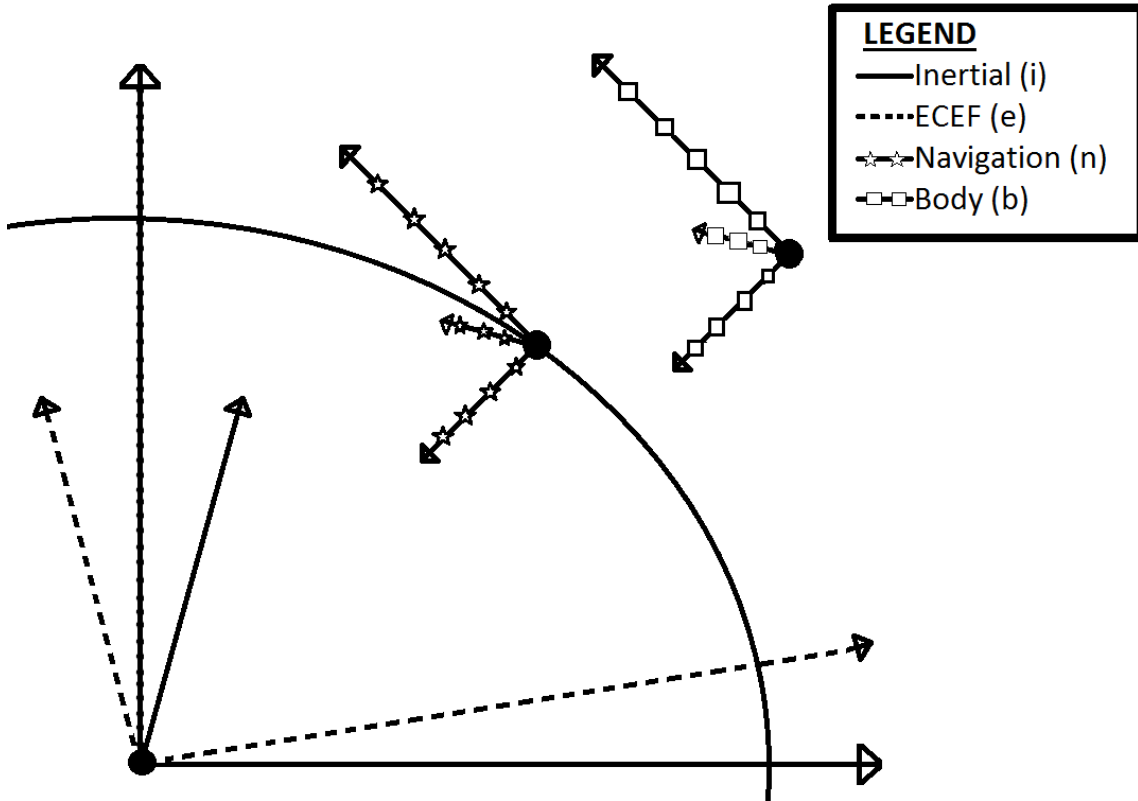


Figure 2.1: World Coordinate Frames. Four different coordinate frames and their relationships. The relationship shown is only valid for a single slice in time as rotation of the earth will change the way that all other frames reference the inertial which is fixed to points outside the Earth. The body frame is attached to an object flying above the surface of the Earth. The navigation frame tracks closely behind along Earth's surface. The Inertial and ECEF frames have their origin at the center of the Earth.

$$X^a = C_b^a X^b \quad (2.5)$$

Equation 2.5 shows the relationship between a measurement in the 'a' frame and that same measurement in the 'b' frame, assuming their origins match. The relationship between the frames is a rotation matrix, which for our example we will describe using the Direction Cosine Matrix (DCM) convention. The DCM matrix can be described in terms of individual rotations about each three axes.

$$\mathbf{C}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & \sin(\phi) \\ 0 & -\sin(\phi) & \cos(\phi) \end{bmatrix} \quad (2.6)$$

$$\mathbf{C}_2 = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \quad (2.7)$$

$$\mathbf{C}_1 = \begin{bmatrix} \cos(\psi) & \sin(\psi) & 0 \\ -\sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.8)$$

Equations 2.6 to 2.8 show the rotations along each axis in the original frame according to the DCM convention. C_3 is a rotation about the x axis, C_2 is about the y axis, and C_1 is about the z axis. The ψ , θ , and ϕ angles inside each of these matrices are called the Euler angles, which are a convenient way of describing rotations. In an aircraft these three angles are called roll, pitch, and yaw and are given in respect to the body frame.

$$C_b^a = C_3^b C_2^b C_1^b \quad (2.9)$$

Equation 2.9 describes the combination of the three DCM rotation matrices defined above to illustrate the rotation between the arbitrary ‘a’ and ‘b’ frames. The order of rotations is important as they are not commutative.

2.1.2 World Model.

In order to give accurate information about local navigation frames in relation to other systems, a standardized definition of the Earth’s surface is necessary. Because elevation above the surface of the Earth is easier to measure for local reference frames, it is important to be able to accurately model the shape and distance between the local surface of the Earth and the origin of the earth centered reference frames. Earth is not a perfect sphere and

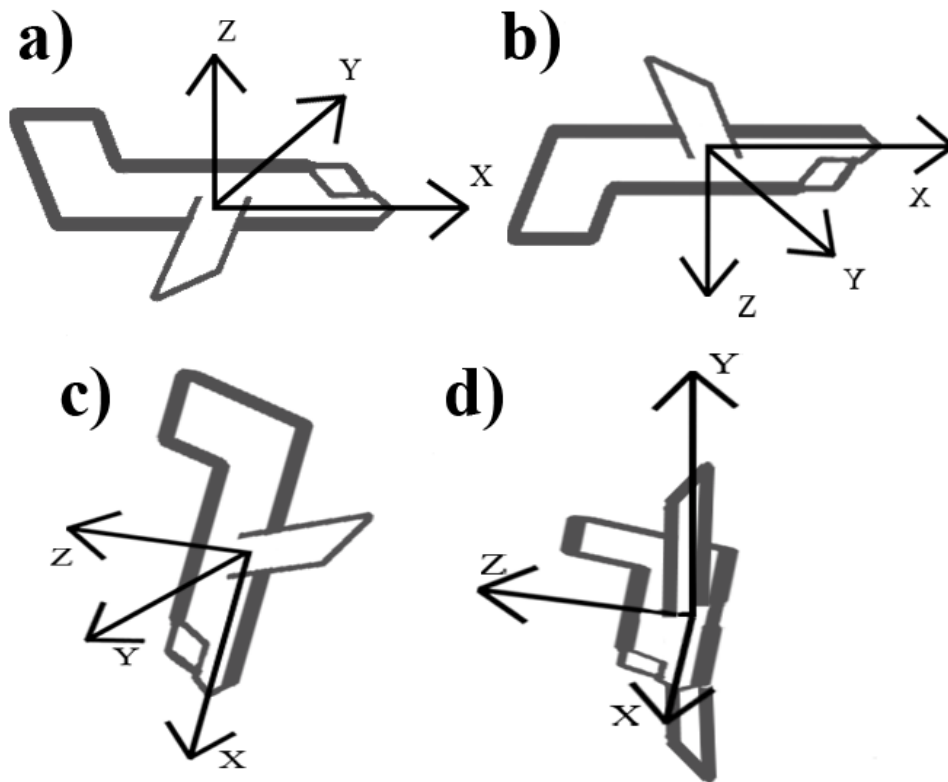


Figure 2.2: DCM Example. A three dimensional example of the different Euler angles in a DCM. Starting with the aircraft oriented according to a, b shows a rotation along the x axis, c shows a further rotation along the y axis, and then d shows the final rotation along the z axis to achieve a complete change in three-dimensional orientation. Each of the angles used to rotate these could be put into the matrices of equations 2.6 to 2.8 above and then combined using equation 2.9 to determine the relationship between the coordinate frames of images a and d.

is commonly modeled by the WGS-84 ellipsoid, which has precise parameters modeling the major and minor axes as well as the eccentricity, flattening, and turn rate of the Earth [5]. Position on earth is commonly described in terms of latitude and longitude, angular distance between the Equator and Prime Meridian respectively, and elevation in terms of distance from the modeled surface of the Earth directly between the body of interest and the center of the frame. More accurate models of the surface of the Earth such as the Digital Terrain Elevation Data (DTED) database are available for precise applications.

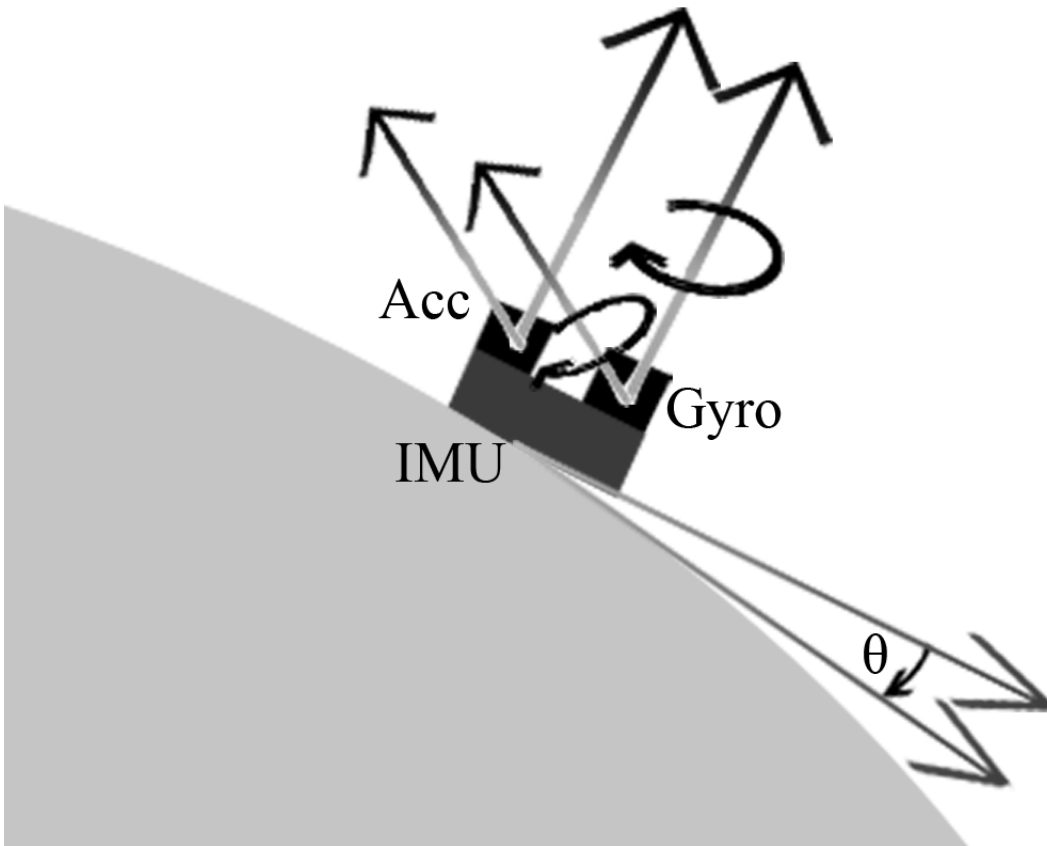


Figure 2.3: IMU Measurements. Different measurements read in by the accelerometers and gyroscopes onboard an IMU during navigation. The IMU is traveling along the surface of the earth with a difference in alignment with the gravity vector defined by θ .

2.1.3 Inertial Measurement Unit (IMU).

A standard sensor used in most aircraft for determining information about the navigation states is an IMU. An IMU determines the changes in both acceleration and orientation of the aircraft over time using accelerometers and gyroscopes. Given initial values of the nine navigation states of an aircraft for a certain trial, estimates of these states at any time afterwards can be determined through integration of the acceleration and rotation measurements coming from the IMU sensors.

The accelerometers and gyroscopes used inside IMUs accrue small errors over time due to various factors [4]. As these errors in the measurements of the change in acceleration and orientation are integrated together with the errors present in all previous measurements, what begins as a slow drift away from the truth becomes increasingly worse over time. Due to any small discrepancies in orientation information, gravity modeling effects can bleed over into the other horizontal directions, further contributing to the growing error. Using only an IMU, a navigation solution can be very accurate over short periods of time depending on the quality of the initialization as well as the quality of the sensors. However, even the most expensive and accurate sensors will accrue biases over time, and over a long enough period their integrated solutions will drift too far away from the truth to be usable.

The drifting error in an IMU is commonly modeled as a bias that grows over time as a random walk with variance dependent on the quality of the sensor being used. Mitigating such biases is the goal of supplementary navigation sensors. Navigation systems using an IMU track these biases as extra states to allow for estimation by other sensors.

The IMU captures high dynamic motion of the aircraft accurately, making calculations of the real world trajectory unnecessary in the filter [6]. The errors of the IMU measurements propagate at a much lower frequency and the deviation of the IMU from truth is known well enough to model for filtering [7]. Navigation estimation can be accomplished by estimating errors in the IMU as opposed to modeling dynamics of the navigating body. The final solution in terms of real world navigation states is constructed by adding the estimates of IMU error onto integrated navigation estimates over the trial.

The Pinson 15 error model [4] defines the propagation of error for the navigation states listed in equation 2.1 as well as the biases in the Gyro and Accelerometer in the IMU. There are three states for each axis in the biases of both the gyroscope and the accelerometer. This model captures inertial data useful for navigation while being small enough to be

implemented in real-time operations. Sensors that measure these added bias states in an IMU system offer great potential to increasing navigation accuracy.

The following equation shows the Pinson 15 error model in state space form for error states of the system developed in Veth's research [1]. The 15 error states, 3 along each axis, of position, velocity, orientation, accelerometer bias, and gyroscope bias are represented by $\delta\mathbf{x}$. I is an identity matrix of diagonal 1 values, which is multiplied by all 3x1 vector values described below to give 3x3 matrices. The C matrices are DCMs from various reference frames described in Section 2.1.1. f represents the specific force vector, G is the gradient of gravity, γ_{ie}^e is the angular rate matrix of the Earth, ω_{ie}^e is the Earth's sidereal angular rate vector, and τ is the bias time constant from the accelerometer and gyroscope. w represents the process noise determined by the quality of sensor being used.

$$\delta\dot{\mathbf{x}} = \begin{bmatrix} 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ C_n^e G C_n^e & -2C_n^e \gamma_{ie}^e C_n^e & f^n & C_b^n & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & -C_n^e \omega_{ie}^e & 0_{3 \times 3} & -C_b^n \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & -\frac{1}{\tau_a} I_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & -\frac{1}{\tau_b} I_{3 \times 3} \end{bmatrix} \delta\mathbf{x} + \begin{bmatrix} 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ C_b^n & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & -C_b^n & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} \end{bmatrix} w \quad (2.10)$$

2.1.4 Global Positioning System (GPS).

A commonly used method for correcting error in an IMU solution is to use GPS measurements that give absolute position information from which velocity and acceleration can be derived. GPS solutions are created from signals broadcast by satellites orbiting Earth. The determination of position is done through relative timing between broadcast and received signals. The solutions coming from a GPS can be accurate down to millimeter level, but robust pseudorange tracking techniques will typically see errors in the low meter level [5]. These errors come from a variety of sources (e.g., the troposphere, ionosphere, satellite geometry, and surrounding objects) whose effects can be modeled to increase

accuracy. Due to the receiver clock error being unknown, a minimum of four satellites are required to build a deterministic position solution in the three reference axes.

2.2 Integrated Navigation Solutions

A navigation solution is created by combining measurements from various sources to determine a best estimate of navigation states for a system. Each sensor involved in the estimation will give different types of measurements and have different types of noise and bias models. Ideally, the different types of information can be combined to build a solution that is more accurate than any one of the sensors alone. Combining different measurements requires accurate modeling of their noise and understanding how that noise changes over time and in different situations.

A key algorithm for estimating navigation states using measurements from various sensors with differing uncertainties is the Kalman filter [8]. The Kalman filter combines sensors with stochastically defined noise to create a Minimum Mean Squared Error (MMSE) optimal solution based on these measurements. The filter is recursive in that it provides an estimate of each state after each measurement is incorporated, meaning it can be used for real-time applications.

The filter tracks not only the estimate of the states, but also the quality (i.e. confidence in those states) as a covariance associated with each state. This covariance continually grows over time as the states are propagated forward until information from sensors is included, which increase the confidence of the model in its estimates. The filter works by predicting the future values of states based on current navigation information and then comparing these values to what is measured by the sensors. The sensor information is weighted based on confidence in the measurements compared to confidence in the prediction. The weighted results are combined with the predicted state to update the filter's estimates.

The following equations describe the propagation step of the filter from time $k - 1$ to time k :

$$\hat{x}_k^- = \phi_{k,k-1} \hat{x}_{k-1}^+ \quad (2.11)$$

$$P_k^- = \phi_{k,k-1} P_{k-1}^+ \phi_{k,k-1}^T + Q_k \quad (2.12)$$

The \hat{x} variable is the estimate of the state while the P variable represents the corresponding covariance. The state value before updating is shown with a $-$ in the top right corner, whereas a $+$ shows the value after updating. ϕ is the state transition matrix described by the model of the state, in the case of error state modeling the Pinson 15 could be used, to determine how it will change over time. Q is the noise strength for the state also defined by the selected model. The covariance will always grow over time through propagation due to the effects of the additive state noise. In order to add information to the filter, updates must be made to the state estimates:

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \quad (2.13)$$

$$P_k^+ = (I - K_k H) P_k^- \quad (2.14)$$

The H matrix is the measurement observation matrix (i.e. the states that the sensor in question observes). The z vector is the measurement itself. The K matrix, called the Kalman gain, is a weighting matrix that determines the amount to which the update influences the filter's estimate of the state. These fundamental equations describe the basic operation of a Kalman filter.

The traditional Kalman filter only works on systems that are linear with additive white noise [6]. The Extended Kalman Filter (EKF) is a variant of the Kalman filter designed to deal with nonlinearities. It can have both nonlinear propagation and measurement

equations. The EKF linearizes about the estimate of the state to deal with these nonlinear functions. Besides this linearization, the rest of its functionality is fairly similar to the Kalman filter. The EKF is widely used in industry for integrated inertial and GPS navigation systems [9] and will also be utilized for vision-based experiments conducted in this research.

2.3 Computer Vision

A sensor that helps correct the growing biases of an IMU and that is also a passive sensor which cannot reasonably be maliciously interfered with is a vision sensor. The idea behind vision-aided navigation is to use tracking and estimation of the position of landmarks on the ground through computer vision to estimate and correct for errors in an IMU. This type of navigation works passively off of the scene around the aircraft making it less susceptible to interference than GPS. The angles between the position of the aircraft and the tracked landmarks that come from the intrinsic calibration of a camera (Section 2.3.1) give us information on the motion of the aircraft.

2.3.1 Camera Model and Calibration.

A camera creates a two dimensional image representation of a three dimensional scene. The relationships between the scene and the created images allow reconstruction of the scene using only images. The pinhole camera model [10] simplifies these relationships while maintaining a level of accuracy useful for a wide variety of applications. The camera center is seen as a point in 3D space with a forward facing z axis coming out of it. A perpendicular plane intersecting the z axis is the image plane, representing what is captured by the camera. The distance along the z axis to this plane is the focal length. The intersection between these planes is the principal point. The x and y axes run parallel to the image plane but perpendicular to each other, the x axis running horizontally along the plane and the y axis vertically. On the other side of the image plane lies the real world from

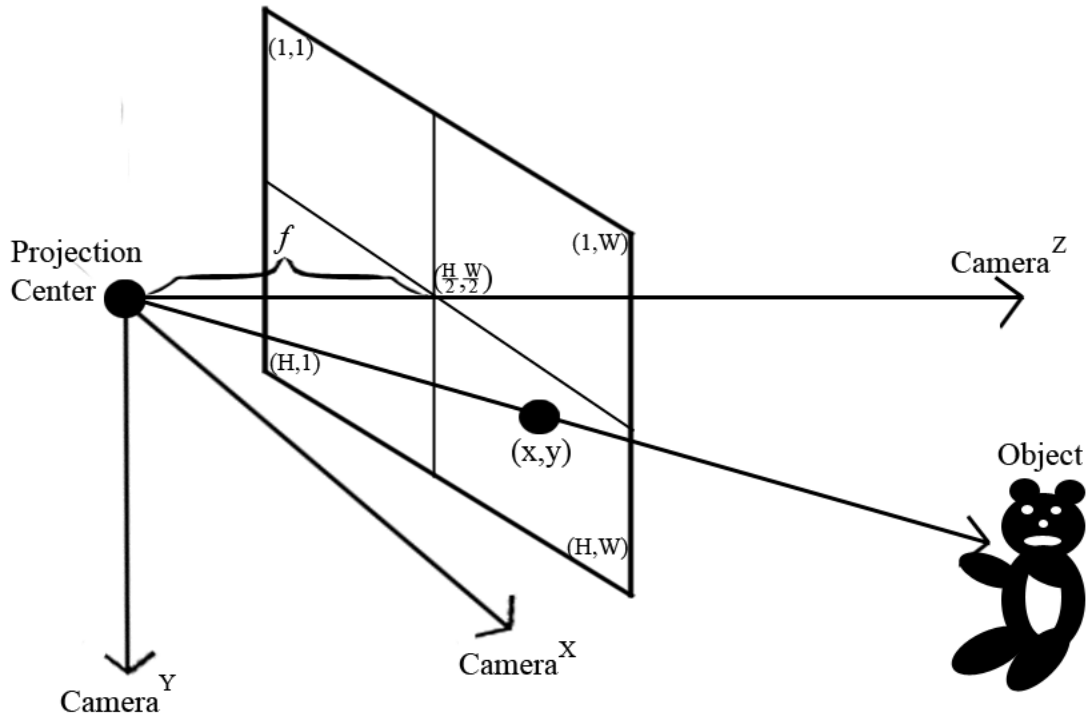


Figure 2.4: Pinhole Camera Model. A three dimensional example of the pinhole camera model. f represents the focal length, (x, y) the pixel coordinates of the image, and $Camera^{x,y,z}$ are the axes in the camera frame. This model does not show the effects of warping and distortion on images, but rather the direct relationships between the camera and the three dimensional scene if these effects were already accounted for.

which the camera senses. Given no distortions, the point at which a line drawn from any point outside the camera to the camera center intersects the image plane is its pixel point.

Using the detected features in an image of a scene to give information about the motion of an aircraft requires very precise knowledge of the relationship between pixel position on an image and the real world angle between the central axis of the camera and the detected feature. The characterization of this relationship is called intrinsic calibration. This can be accomplished using objects found in images with prior knowledge about the dimensions of the object. A common method of performing intrinsic calibration involves holding a checkerboarded pattern of known dimensions in front of the camera at different angles and



(a) Uncalibrated Image

(b) Calibrated Image

Figure 2.5: Image Calibration Example. An example of the distorting effects of a camera. The calibrated image was re-plotted to remove the warping characteristics of the camera used to take the photo. Note the bend in the metal stand along the left edge of the uncalibrated image and how it is fixed after calibration. Camera calibration estimates the warping characteristics of the camera. The calibrated image can now be used according to the pinhole model to obtain three dimensional information about objects in the scene.

positions [11]. Once the corners of the checkerboard are detected, the rest of the pattern can be found and the comparison between what the camera sees and the known physical dimensions of the object give the intrinsic calibration of the camera. This calibration can model the effects of warping in an image, wherein the sensor or lens of the camera distort and bend the representation of the three dimensional scene in image creation.

An intrinsic calibration estimates the focal length, principal point, and skew coefficient in a camera. These are modeled by the following equations:

$$x_p = f_c(1)(x_d(1) + \alpha_c x_d(2)) + c_c(1) \quad (2.15)$$

$$y_p = f_c(2)x_d(2) + c_c(2) \quad (2.16)$$

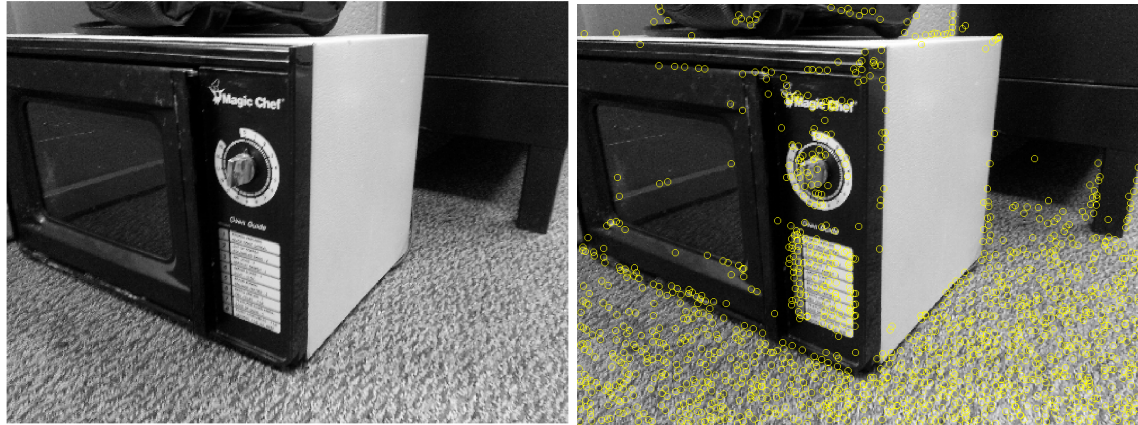
$$K = \begin{bmatrix} f_c(1) & \alpha_c f_c(1) & c_c(1) \\ 0 & f_c(2) & c_c(2) \\ 0 & 0 & 1 \end{bmatrix} \quad (2.17)$$

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = K \begin{bmatrix} x_d(1) \\ x_d(2) \\ 1 \end{bmatrix} \quad (2.18)$$

x_p and y_p in equations 2.15 and 2.16 are the pixel values of a point in a camera image. f_c represents the focal length of the camera and c_c is the principal point. α_c is the skew coefficient which relates the x and y axes of the image plane. All of these are captured in the matrix K shown in equation 2.18 which is a standard format of presenting a camera calibration. With the calibration matrix, relevant three dimensional information can be obtained from images of a scene.

2.3.2 Feature Detection and Extraction.

An image is a collection of pixels with varying intensities that represent the scene from which the image was taken. In order to use the image for a higher level purpose, in our case navigation, it is necessary to extract information from the image and collect/sort this in a usable way. The method of gaining information from images commonly used is feature-based, wherein the image is decomposed into features describing identifiable points in the image that could ideally be detected again from an image of the same scene taken at a different point, given some constraints. Feature detection is the process of looking at the image and determining which points are unique within the scene. Feature extraction is the process of defining the detected features using parameters and information that is ideally



(a) Base Image

(b) SIFT Overlay

Figure 2.6: SIFT Feature Example. An example of detected SIFT features in an image. Notice that areas of high texture and relative contrast tend to have more features detected. The side of the microwave has no distinct points so no features were detected along it. On the other hand, the power settings for the microwave and the carpet have many overlaid features due to their highly textured appearance.

robust enough that the same point would have a highly similar description if the image was taken from a different perspective.

The Scale Invariant Feature Transform (SIFT) method of both feature detection and feature extraction introduced by Lowe [12] has proven robust for matching points across multiple images taken of a single scene. The feature detection part of SIFT looks at points of interest that are identifiable over multiple Gaussian blurs of the image. The image is consecutively convolved with Gaussian filters and successive images are subtracted from each other to result in a Difference of Gaussian filter which accentuates edges and other detectable points from which SIFT selects. These copies of the image are continually down-sampled as the images become increasingly blurred, making those points identifiable across all images invariant to scale.

The SIFT method of feature extraction looks at the gradients of the pixels surrounding the points detected and creates normalized vector groups to generate descriptors for each

feature. By labeling feature points in terms of gradients in the local area these points are less affected by lighting conditions and invariant to rotation and scaling as well as partially invariant to affine transformations, making SIFT ideal for use in navigation. Once features have been identified and described, they can be used to identify corresponding points in different images.

The SIFT method of matching features looks at the euclidean distance between the descriptors of points of interest in two images and compares the strongest match to the second strongest for each point. If the strongest match is a certain magnitude stronger than any other matches for the point, the match is considered to be valid. The parameter that controls this percentage threshold to determine the difference needed between match strengths, *distRatio*, helps control both the quantity and quality of matches. A higher ratio relaxes the requirements for a successful match thus allowing more matches to be made at the cost of allowing more erroneous matches. Similarly, the inverse occurs when lowering the ratio in that less quantity of higher quality matches are selected. The optimal value for this parameter changes for each set of data it is used on.

2.4 Structure from Motion (SfM)

Outside the field of navigation, practitioners of computer vision endeavored to reconstruct three dimensional scenes from unordered sets of images. The process, called Structure from Motion (SfM), creates these solutions and has been solved using no a priori information about the trial or images being fed into the system [13] [14] [15]. SfM is a robust computer vision technology that combines images of a scene and produces an estimate of the positions and orientations of the camera at those times and three dimensional positions of features matched across the images. The order of images is unnecessary to build a solution, but can be incorporated to constrain matching. In order to recreate a scene using SfM, the program must derive information about the relative positions and orientations of the cameras producing the images as a necessary stepping stone to

determining three dimensional feature locations. This by-product is an estimate of the navigation states of the camera. Even without timing or inertial data, given only images SfM is able to obtain three dimensional position and orientation information along with other types of navigation and motion measurements from this reconstruction.

The Building Rome in a Day project [16] showed how images taken across multiple sensor platforms can be compared and features matched to build a three dimensional reconstruction of a scene. A massive collection of images of downtown Rome from the popular photo-sharing site Flickr were input into SfM without any prior knowledge about the cameras used to take the pictures. With enough images, the team was able to recreate a virtual Rome with position and pose estimates for each camera used.

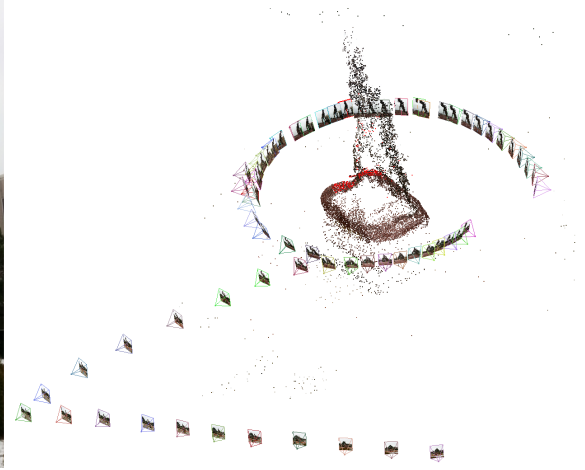
Interestingly, SfM can produce an estimate of intrinsic camera calibration if none is given initially, including a radial distortion estimate. SfM is a powerful algorithm that this work leverages as a source for motion estimates that can be used alone or in concert with other navigation information (e.g., IMU updates and GPS signals). The SfM process is a collection of various tools that all come together to give the three dimensional information that is central to this work. Feature detection and matching have already been discussed, but two other key tools in SfM will be explored to explain the source of estimates coming from SfM.

2.4.1 Bundle Adjustment.

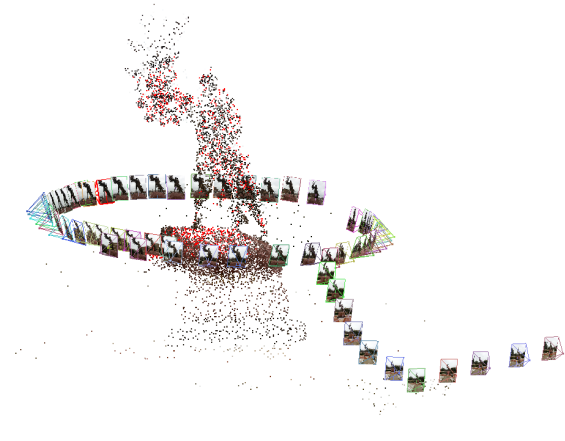
SfM is performed by determining the relative change in position and rotation that best fit detected correspondence between images. A sparse reconstruction is built of the scene based on matched features giving a three dimensional estimate of the features in addition to the positions and pointing angles of the camera for each image [14]. The experiments in this work leverage these estimates to provide translation information about the camera from frame to frame.



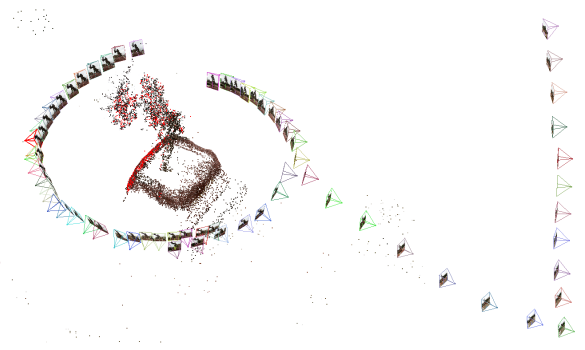
(a) Picture of Icarus Statue



(b) SfM Front View



(c) SfM Low View



(d) SfM Top View

Figure 2.7: SfM Position and Scene Reconstruction Example. An example of the position, pointing angle, and feature location estimates from SfM. The example is a reconstruction of the Icarus statue at the Air Force Institute of Technology (AFIT). The pyramid shaped objects represent the camera with position at the tops of the point and the pointing angle out the flat bottom. The point cloud is a reconstruction of all features matched across the images.

Bundle Adjustment (BA) is the SfM tool that estimates camera pose and position as well as calibration parameters [17]. This is accomplished by refining the sparse three dimensional reconstruction originally estimated in the program. First, a single image pair

is matched and correspondence between the two frames is determined through geometric constraints [14]. Next, another image is added and its intrinsic and extrinsic parameters are propagated forward from parameter estimates of previous frames. Tracks, or features matched across multiple images, present in the previous and current images are compared in terms of pixel location with estimated locations based on initial parameter estimates. The intrinsic and extrinsic parameters are refined through a minimization of reprojection error between predicted pixel locations using existing camera parameters and those observed in the current image. Once the parameters are refined for the current image, the same process is accomplished iteratively for all future frames until the set is complete.

$$\sum_{a=1}^y \sum_{b=1}^z = x_{a,b} * \|q_{a,b} - P(\theta_a, p_b)\| \quad (2.19)$$

Equation 2.19 describes the reprojection error seen between predicted and observed pixel locations of tracks in images added to BA. θ is the camera model defining the estimates of camera position, orientation, and focal length. p is the three-dimensional predicted position of each track, and $P(\theta, p)$ is the predicted pixel location of the track based on the camera model and three-dimensional position estimate of the features. q is the observed pixel of the track in the image. y is the number of camera frames in BA and z is the number of tracks in the image. The variable $x_{a,b}$ equals one if the track b is in frame a , otherwise it is equal to zero. Minimizing the right side of this equation is the goal of BA, which is accomplished through a refinement of the camera model and three-dimensional feature locations.

Reprojection error minimization is a non-linear least squares problem commonly solved using the Levenberg-Marquardt algorithm [18]. This algorithm iteratively adjusts parameters by a certain delta based on the gradient of the error equation at the current estimate. It finds local minimums which may not be global for a specified problem. If only one minimum is present, it is robust in that it can find the minimum with poor initialization.

The ability to fit the current image on to the previously constructed three-dimensional model improves the detail of the model by adding new points that are aligned with previously seen features. The natural effect that this has on refining the camera position and orientation estimate is beneficial to the navigation application demonstrated in this thesis. This allows imagery to be used as a navigation instrument by providing position estimates over time.

2.4.2 RANSAC.

Random Sample Consensus (RANSAC) [19] is a process by which data is made to fit a model based on general consensus by avoiding outliers that do not fit this model. This is not performed through using all of the initial data and determining which points are outliers, but rather by starting with a random small set of points and only adding on from there based on consistency with the initialization. Running this algorithm many times with different initializations narrows down the points to a model that best fits the data. RANSAC helps eliminate outliers in image matches that would detriment parameter estimates between images in BA.

2.5 Vision Aided Navigation

Once the images from a trial are processed and matches found, it is necessary to incorporate their measurements into the navigation system of the aircraft. The simplest method of doing this is by dead reckoning, wherein the translation and rotation estimated by feature matching between each successive set of images is used as a separate source of information about the movement of the aircraft. This method is not at all coupled with the inertial system in the sense that both the vision systems and the inertial systems are producing separate measurements without information from each other.

The method to update the inertial system using vision measurements coming from Veth's research [1] involves comparing the predicted pixel point of a tracked feature to its detected location in a subsequent image. The distance and angles between those two

points gives information as to how the aircraft has moved between those two instances of time, which captures motion in a different way than the inertial system. In his research these residuals of the predicted to measured pixels are incorporated into a Kalman filter in order to give an update to the navigation state estimates. This method couples the vision and inertial systems to improve feature matching and it incorporates the covariance of the position estimate to get information from the residuals of the feature location comparisons. This method requires higher fidelity Digital Terrain and Elevation Database (DTED) information about the local area being viewed by the camera in order to estimate the height of the ground — a requirement that is avoided by the approach taken in this study.

IMUs are usually corrected by GPS signals because of their ability to determine the biases in IMUs over both short and long intervals. The GPS solutions are absolute measurements that are given with respect to a world frame, and thus should not drift over time. A drawback, however, is that GPS is highly dependent upon tracking at least the 4 signals necessary to navigate without drifting errors. Without enough pseudoranges from satellites, due to signal jamming making these signals untrackable, the GPS will not be able to give position updates to the navigation system and it will begin to drift. This work focuses on incorporating vision systems (with an aim at MWIR in particular) to provide similar navigation benefits without the potential for jamming.

III. Structure from Motion and IMU Combined Solution

In order to demonstrate the usefulness of a new technology for vision-aided navigation, we need a tool with which we can directly compare technology against existing implementations. The tool that works for both the visible light and MWIR cases is the Structure from Motion (SfM) [20] algorithm. This tool is employed because it can produce navigation estimates from sequential images by aligning distinct features in an image that can be matched with corresponding features in a reference image. This chapter explores the potential for the navigation estimates from SfM to aid in navigation combined with an IMU.

3.1 Dataset

The Minor Area Motion Imagery (MAMI) trial, conducted by Air Force Research Laboratory (AFRL) in June of 2013, flew a NASA Twin Otter aircraft with various side looking cameras over Wright Patterson Air Force Base (WPAFB). The trial collected maximal framerate and resolution data from various camera phenomenologies (i.e., EO, Short Wave Infrared (SWIR), and MWIR) to allow comparison between the different imaging domains.

The sets of cameras were attached to gimbals on the aircraft that rotated both left/right and up/down in motion. Each camera gimbal was fitted with an Inertial Navigation System (INS), a combination of a GPS antenna and IMU, to obtain precise positioning and angle of the gimbals over time. The INS triggered the EO and SWIR cameras on each gimbal to give accurate time stamps for the images.

There were 2 different flights with usable EO and MWIR imagery over a long period of time in the MAMI dataset. They all flew in circular trajectories over specific ground points of WPAFB as the original intent of this data collect was to stitch massive image

collections together to create a large scene. The planes maintained an altitude between 1000-3000 meters across all of the trials.

3.2 Equipment

This research is based on camera 6 (Prosilica GE2040) from the MAMI dataset. The resolution of the camera is 2040 x 2048 pixels and it takes full resolution images at 15 Hertz. The camera's high resolution makes it more attractive for navigation aiding by detecting more unique content in a scene. For this research, downsampling the image rate to 3 Hertz was found to produce adequate results.

The GPS and IMU combined INSs on board the aircraft were Novatel SPAN Propak V3's. The tactical grade inertial sensor inside these INSs was a Honeywell HG1700 IMU. This IMU runs at 100 Hertz and its sensor quality gives in the classification of a tactical grade IMU. The combined solution from both of these sensors provided an estimated absolute position accuracy of 1.5 meters, which will serve as the truth reference for the following experiments because it should remain at least an order of magnitude more accurate than the SfM combined and free running IMU solutions over time.

3.3 Assumptions

In order to use the data provided from the MAMI data collect, the following assumptions were made to allow processing of the given measurements: The extrinsic calibration (i.e. the rotation and translation between the cameras and the INS) is assumed to be zero. According to diagrams and pictures from the trial, the two were mounted onto a single metallic skeleton within a meter of each other, thus for the purposes of this research, it is assumed that they are co-located in space with a fixed rotation that was determined experimentally. The second assumption is that both aircraft and gimbal motion would affect the INS and camera in the same way. Thus by tracking navigation with respect to the camera frame, the INS nav solution would be directly relatable to camera

position. The next assumption is that the post processed INS solution measured on the gimbal associated with a particular camera was assumed to be the truth against which all other solutions in this research were compared against. While it would be ideal to have an external measurement on the aircraft position (e.g. differential GPS; with a 1 cm position error), such an assumption is considered to be reasonable because the magnitude of error for the GPS solution is rated to be within 1.5 meters over time, which is at least an order of magnitude lower than the expected performance of the alternative solutions explored in this work. The last assumption is that the error of each velocity measurement created by SfM has known error description. Accurate modeling of error for the SfM solution was beyond the scope of work for this research, so the noise modeling used to combine measurements used error calculated between SfM measurements and the post processed INS solution.

3.4 Navigation Solution

The navigation solutions computed in this research combined motion data from inertial sensors with velocity information from SfM. The process of utilizing these sensors for navigation required various tools and prerequisite steps to set up the measurements.

3.4.1 IMU Simulation from INS solution.

The IMU that was part of the dataset given for the trial did not record the necessary alignment information that should have been recorded during its operation. Due to this constraint, it was deemed appropriate to simulate an IMU using the INS solution as the baseline. The INS solution was converted into changes in velocity and orientation to format it like an IMU. Growing noise that was randomly generated according to the noise characteristics of the specific IMU model for the trial was added onto these accurate measurements to simulate appropriate biases and measurement error. This research modeled the Honeywell HG1700 tactical grade IMU that was used to create the provided INS solution.

Table 3.1: HG1700 IMU Parameters

σ_{gyro}	4.8E-6
GyroTimeConstant	3600
AngularRandomWalk (ARW)	8.7E-5
σ_{accel}	9.8E-3
AccelTimeConstant	3600
VelocityRandomWalk (VRW)	9.5E-3

Table 3.1 shows the parameter values for the HG1700 necessary to simulate one using a truth trajectory. The values reflect data derived from experimental tests on HG1700 IMUs. Commercial grade IMUs would have greater σ and random walk values, while navigation grade ones would have the opposite. These parameters determine the ability of an integrated free running inertial navigation solution to stay close to the true trajectory over time.

$$\text{Measurement Noise} = X\text{RandomWalk} \sqrt{dt}\text{randn} \quad (3.1)$$

$$\text{Bias}(z) = \phi_X \text{Bias}(z-1) + \sqrt{\frac{2(\sigma_X * dt)^2}{X\text{TimeConstant}}}\text{randn} \quad (3.2)$$

Equation 3.1 shows the magnitude of the measurement noise for each sensor while equation 3.2 shows the growth of drifting biases in the inertial sensors from time $z-1$ to time z according to the parameters listed above in table 3.1 (replace X with the corresponding gyro or accelerometer parameter). ϕ represents the effects of the time constant for each sensor on the previous bias value. The randn variable represents a pseudorandom number selected by a Gaussian distribution centered on 0 with a variance of 1, showing how noise generated for each iteration of the IMU is incorporated into simulated sensor measurements.

$$\delta\theta_{\text{Gyro}} = \delta\theta_{\text{true}} + \text{Gyro Measurement Noise} + \text{Gyro Bias} \quad (3.3)$$

$$\delta v_{\text{Accel}} = \delta v_{\text{true}} + \text{Accel Measurement Noise} + \text{Accel Bias} \quad (3.4)$$

Equations 3.3 and 3.4 are the model for the simulated IMU measurements. These measurements are a function of the calculated changes in orientation and velocity, the measurement noise, and drifting biases. Using the noise generated by *randn* in equations 3.1 to 3.2, these summarize the simulated measurements used for navigation solutions in this research.

3.4.2 SPIDER.

This research used Sensor Processing and Inertial Dynamics Error Reduction (SPIDER) [3] to combine sensor measurements and create navigation solutions. SPIDER is a software tool used to give versatility and modular design of a Kalman filter to the user. It was designed by the Autonomy and Navigation Technology (ANT) center at the Air Force Institute of Technology (AFIT) to simplify the process of building navigation filters. It provides a modular set of predesigned models for various sensors that can be combined. In this way, the filter can run for various lengths of time under various sensor arrangements with only minor parameter adjustments. For this research the EKF [6] was used in SPIDER for navigation estimation as it can deal with non-linear propagation and update equations.

3.4.3 Navigation State Estimation.

Within SPIDER, the IMU was used as the reference trajectory of the navigation solution. The states propagated forward in time were error states of the IMU. The velocity measurements coming from SfM are incorporated into the system as measurements of the error state of velocity in the IMU.

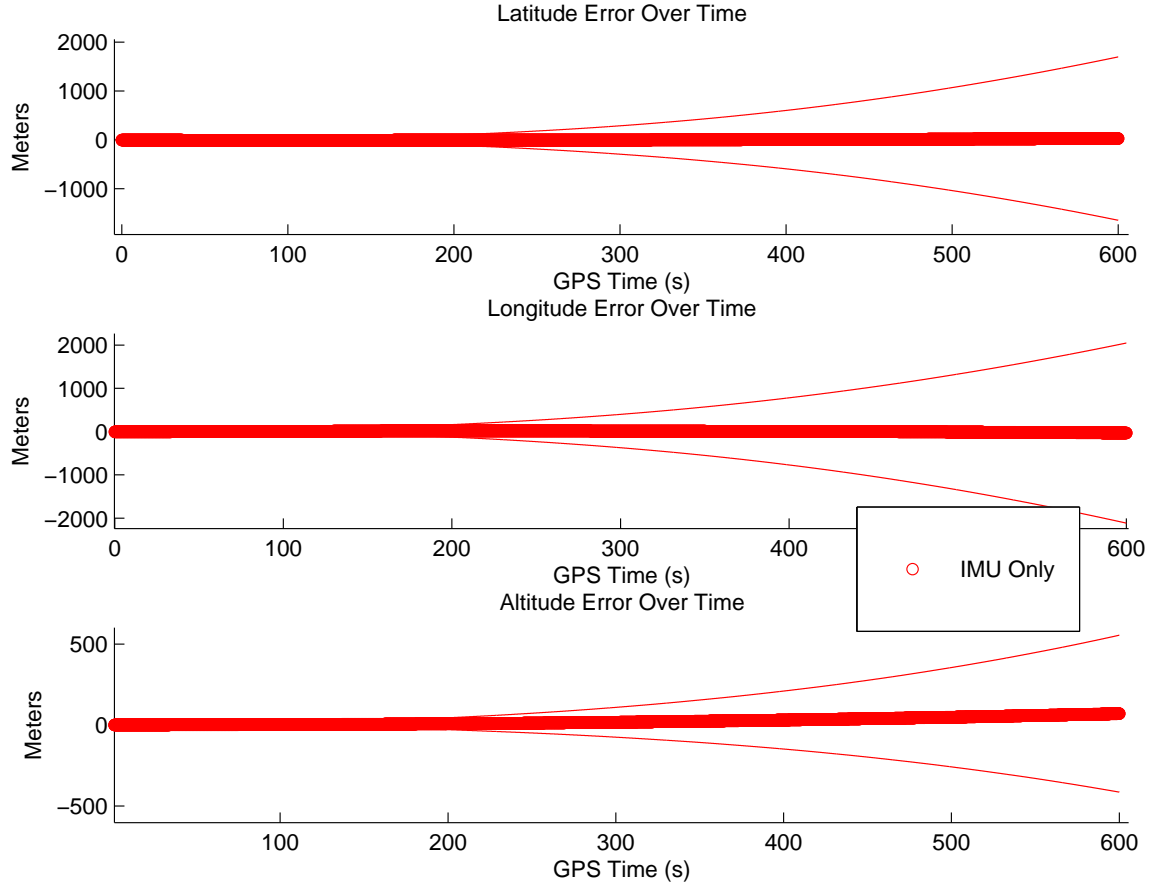


Figure 3.1: IMU Solution Over 10 Minutes. Position error of the simulated IMU solution over 10 minutes. This error is an average of 250 different noise realizations for a single trial using noise simulated using the parameters described in table 3.1. The error for the HG1700 simulation is the same magnitude as observed error from previous research with tactical grade IMUs [21].

$$v_{\text{IMU}}(t_k) = v_{\text{true}}(t_k) + \delta v(t_k) \quad (3.5)$$

$$v_{\text{SfM}}(t_k) = v_{\text{true}}(t_k) - \sigma_{\text{SfM}}(t_k) \quad (3.6)$$

$$z(t_k) = v_{\text{IMU}}(t_k) - v_{\text{SfM}}(t_k) = \delta v(t_k) + \sigma_{\text{SfM}}(t_k) \quad (3.7)$$

The velocities v above are not states tracked by the system but rather interpretations of velocities coming from measurements. $\delta v(t_k)$ represents the error state for velocity in the IMU. Equation 3.5 shows how the velocity measurement from the IMU relates to the real world velocity through the error state at time k . Equation 3.6 shows the same for the SfM velocity measurements through the measurement error present at each time, $\sigma_{SfM}(t_k)$. These velocity measurements are compared to obtain the measurement used in SPIDER, z , which represents the measurement noise from SfM added onto the filter error state.

$$\hat{\delta v}^+(t_k) = \hat{\delta v}^-(t_k) + K(t_k)(z(t_k) - \hat{\delta v}^-(t_k)) \quad (3.8)$$

Equation 3.8 shows the incorporation of the measurement determined above into the update of the velocity error state. The error state after update, denoted by the + sign is changed from its previous state, $\hat{\delta v}^-(t_k)$, by a weighted amount, determined by the Kalman gain K discussed in Section 2.2, of the residual of the measurement, or difference between the measurement and the current estimate of the state before update. The SfM measurements have observability on the growing error in the IMU velocity.

3.5 Visual Structure from Motion

Visual Structure from Motion (VisualSfM) is an application that performs incremental SfM using images fed into it of a scene [20]. The program creates a three dimensional sparse reconstruction of the features detected in the scene that were matched across multiple images [14]. Although it was not created for navigation purposes, the program inherently estimates the position and orientation of the camera for each image. Interestingly, because the algorithm tracks features through the camera's field of view, SfM is also able to estimate the internal (i.e. intrinsic) calibration of the camera for a given trial. Yet another advantage of SfM is that the program leverages graphical processing units to parallelize processing, thus making it ideal for long trials with many high resolution images [22] [23].

With any SfM solution, correspondence is needed between images to determine a fit for the current selection. Such failures can occur for a single moving body when the images are taken too far apart or the scene changes too drastically in between frames. When this happens, VisualSfM will begin creating a new model with images that do not fit to the old one. These new models do not have any connection of information with the old model, a condition that is problematic for the purposes of navigation. This effect was considered when downsampling the framerate of the images as it hindered use of the SfM models for navigation.

For this research, the motion estimates between images from VisualSfM were used to update the Kalman filter in SPIDER. The images were matched in sequence across a 15 second sliding window to best determine the position of the camera at each capture while still leaving the possibility open that this processing could be done in real time with a delay on the navigation estimate. Feature matching success naturally decreases as time between images increases as it is not invariant to affine transformations, making images further away from the origin of each match less relevant and helpful to finding a best fit solution. In fact, correspondence between 2 images of completely different scenes is undesired and this time window limits such effects.

Limiting the time allowed to match images across also prevented the solution from attempting to 'close the loop' if the aircraft were to return close to a previous position later in the trial. The purpose of this research was for the aircraft to operate in an unknown environment, so potential benefits offered by passing over the same ground space are intentionally ignored as they are not the focus of this study.

The output from running VisualSfM over the images for this trial is a file giving estimates of the orientation, position, and radial distortion of the camera for each image as well as a full set of position estimates for each feature tracked in the program. The point cloud of tracked features is a useful analysis tool that will be looked at later in the thesis.

The orientation information was not used to update the navigation solution as the gimbal motion made very drastic movements not captured in the INS.

3.6 SfM Measurements

In order to incorporate the SfM measurements into SPIDER, they must be interpreted in a way that can be combined with the inertial measurements. The first issue is that the SfM position estimates are in a different frame of reference from the navigation system. The second issue is that there is no information characterizing noise of the SfM position estimates. Both of these problems must be addressed before a combined navigation solution can be created.

3.6.1 SfM Alignment.

The position estimates created by VisualSfM are in reference to an arbitrary frame that exists in the reconstruction it creates. In order to use these position estimates for navigation, this arbitrary frame must be aligned to a real world reference frame.

$$p_{\text{SfM}} T_{\text{SfM}}^{\text{INS}} = p_{\text{INS}} \quad (3.9)$$

$$T_{\text{SfM}}^{\text{INS}} = p_{\text{SfM}}^{-1} p_{\text{INS}} \quad (3.10)$$

Equation 3.9 show the mapping between the homogenous vectors p_{SfM} and p_{INS} that represent the SfM trajectory estimate and GPS aided INS position solution for a fixed amount of time before the simulated GPS outage. The relationship between them, transformation matrix $T_{\text{SfM}}^{\text{INS}}$, represents the rotation, translation, and scaling to align the frames. The inverse function of vector p_{SfM} in equation 3.10 is a pseudo-inverse of the SfM position estimate. This equation shows the calculation used to obtain the transformation matrix between the two trajectories.

$$\mathbf{T}_{SfM}^{INS} = \begin{bmatrix} \text{Rotation}_{3 \times 3} & \text{Translation}_{3 \times 1} \end{bmatrix} \quad (3.11)$$

Equation 3.11 breaks down the transformation matrix calculated by matching the two trajectories. It consists of a scaled rotation matrix and a translation vector. The scaling of the transformation, which maps the difference between the SfM frame and meters in the real world, is determined by taking the norm of the rotation matrix. The normalized rotation matrix is a DCM moving the SfM trajectory into the real world reference frame. In order to deal with the differences in scale between the values of the two trajectories (the INS solution is in ECEF coordinates, while the SfM solution is in an arbitrary zero-centered reference frame) the position of the first INS estimate is subtracted from all INS estimates and then added into the calculated translation afterwards.

The length of time for SfM alignment was chosen to be 10 minutes for the trials in this research. Experimental testing showed that too short of an alignment gave a poor rotation and translation matrix causing the SfM solution to diverge from the GPS corrected solution very quickly. Too long of an alignment caused similar problems as the SfM solution also drifts away from truth over time unlike the INS solution, causing an increasingly worse fit between the two over longer alignments.

3.6.2 SfM Trajectory.

The SfM solution appears to capture slightly different motion in the altitude domain compared to the INS solution (Figure 3.6). This could be gimbal motion not tracked in the INS, alignment errors, or some other unknown factor causing a discrepancy in position tracking. As an external measurement of the trajectory is non-existent for this trial, it is difficult to go back and infer what caused the differences. Either way, this difference is worth noting in that it will have some effect on the contribution of the SfM velocity measurements.

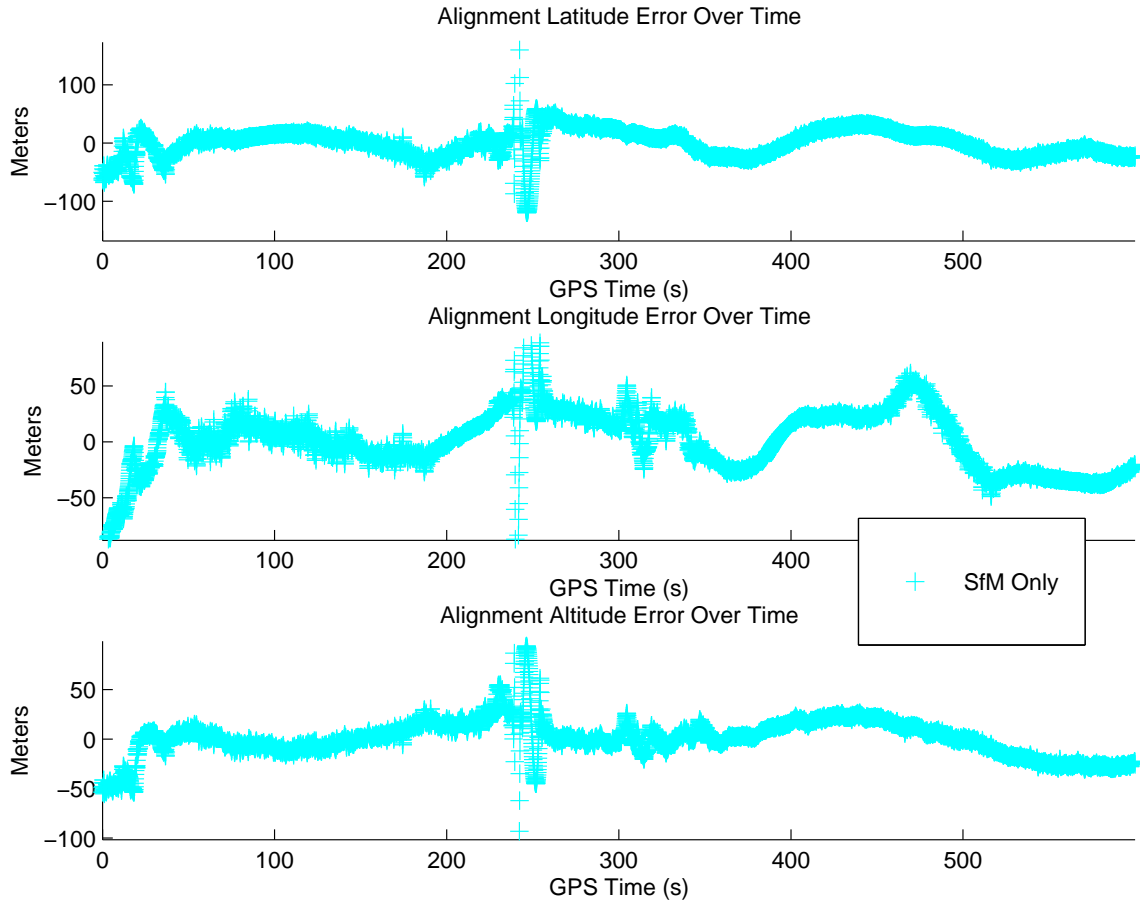


Figure 3.2: SfM Alignment Errors

Figure 3.3: Alignment Position Error. A picture of the comparisons between position estimation for the aligned SfM trajectory and the GPS aided INS solutions before the GPS outage. The position error tracks within around 50 meters or so for each axis over the entire alignment. The plot suggests some level of correlation between axes errors over time.

3.6.3 SfM Velocity Measurements.

SfM estimates the change in position between image captures for the camera in question. As precise timing of the image captures was known in this research, these changes in position were converted to a measurement of velocity once the SfM frame was aligned with the GPS solution.

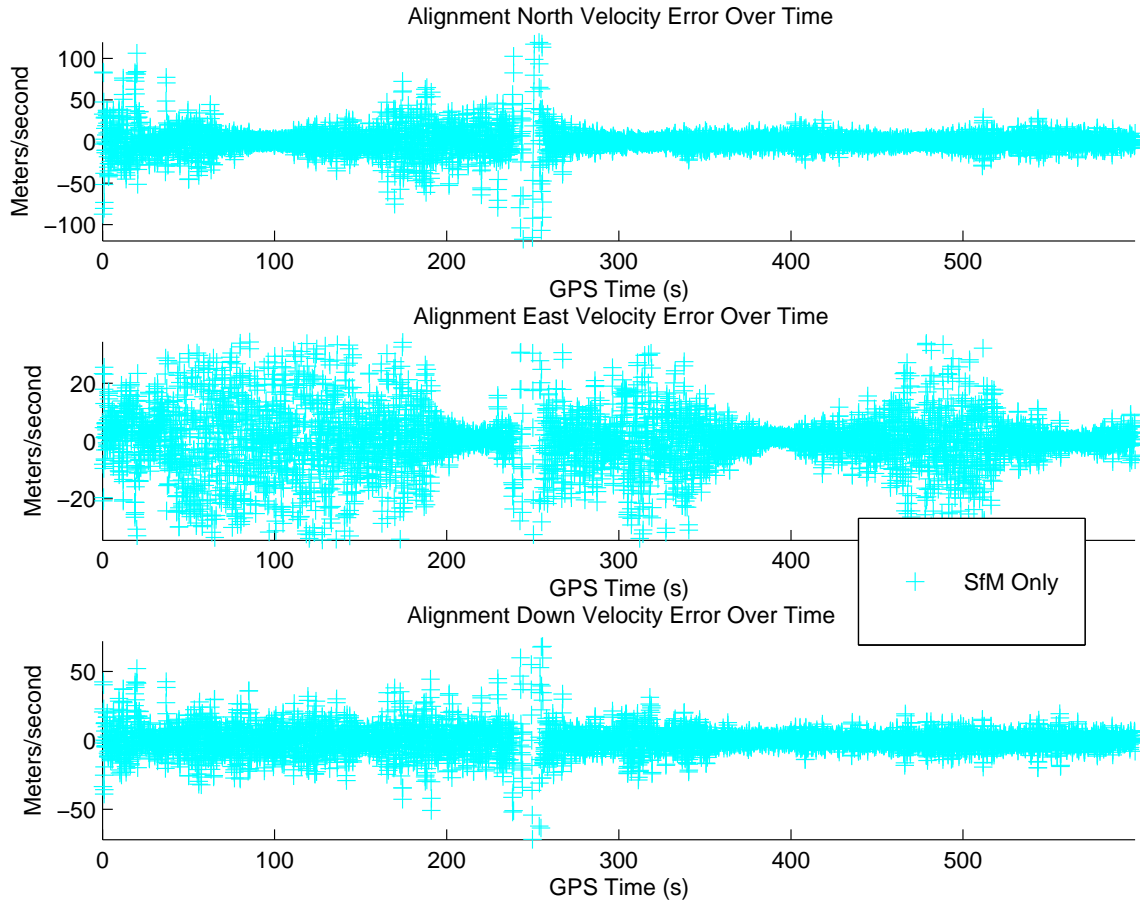


Figure 3.4: SfM Alignment Errors

Figure 3.5: Alignment Velocity Error. Error in the aligned SfM velocity before the GPS outage. The error appears to be zero mean with a time-dependent variance.

The velocity measurements from VisualSfM only give information about the 3 axis motion of the aircraft and not its pointing angles. This information aids the combined navigation solution by giving a second level of measurements that drift differently from the IMU biases.

3.6.4 Noise Modeling.

The velocity measurements from SfM capture absolute motion and will not grow or shrink in general magnitude over time unless the plane’s motion also does so. To capture

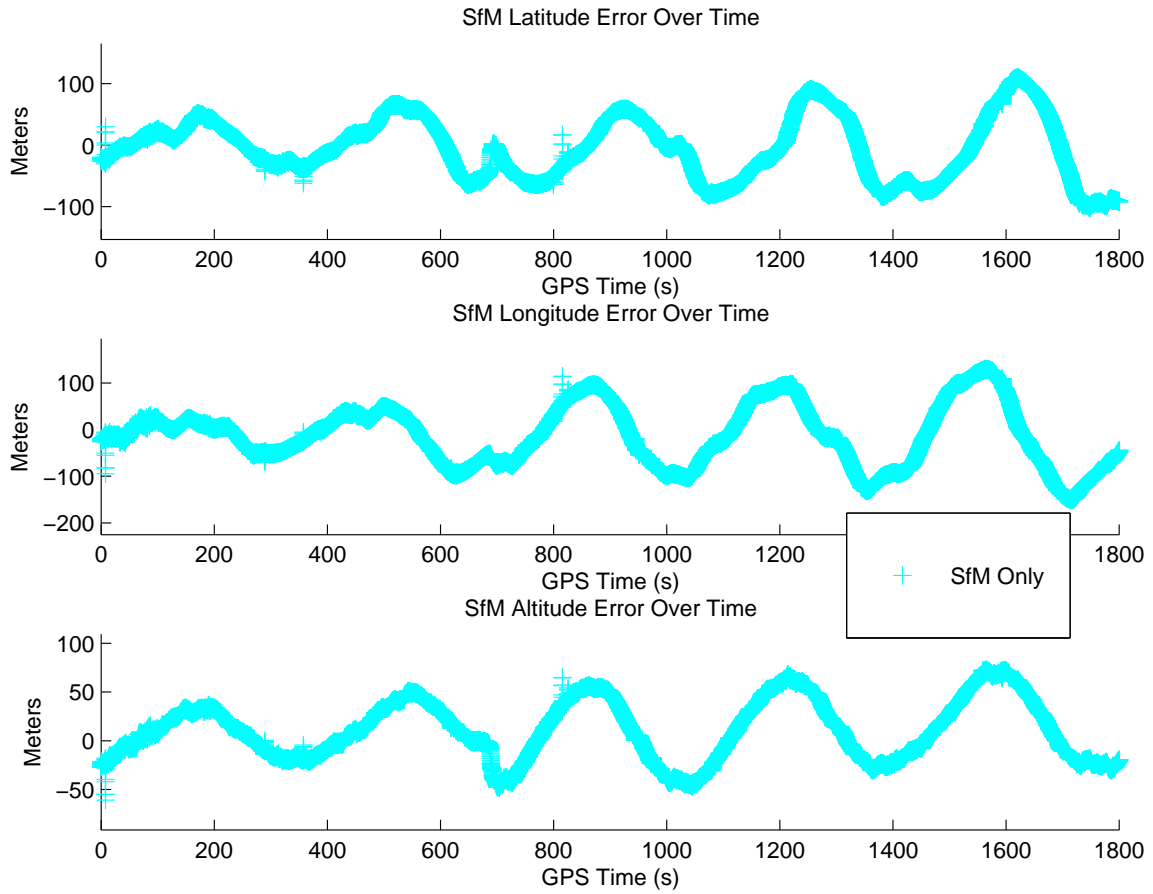


Figure 3.6: SfM Only Position Error

Figure 3.7: Latitude, Longitude, and Altitude Error of SfM Solution. Comparison of the position solution to the INS in terms of latitude, longitude, and altitude error over the entire trial. The position errors maintains similar magnitudes seen during alignment in figure 3.2

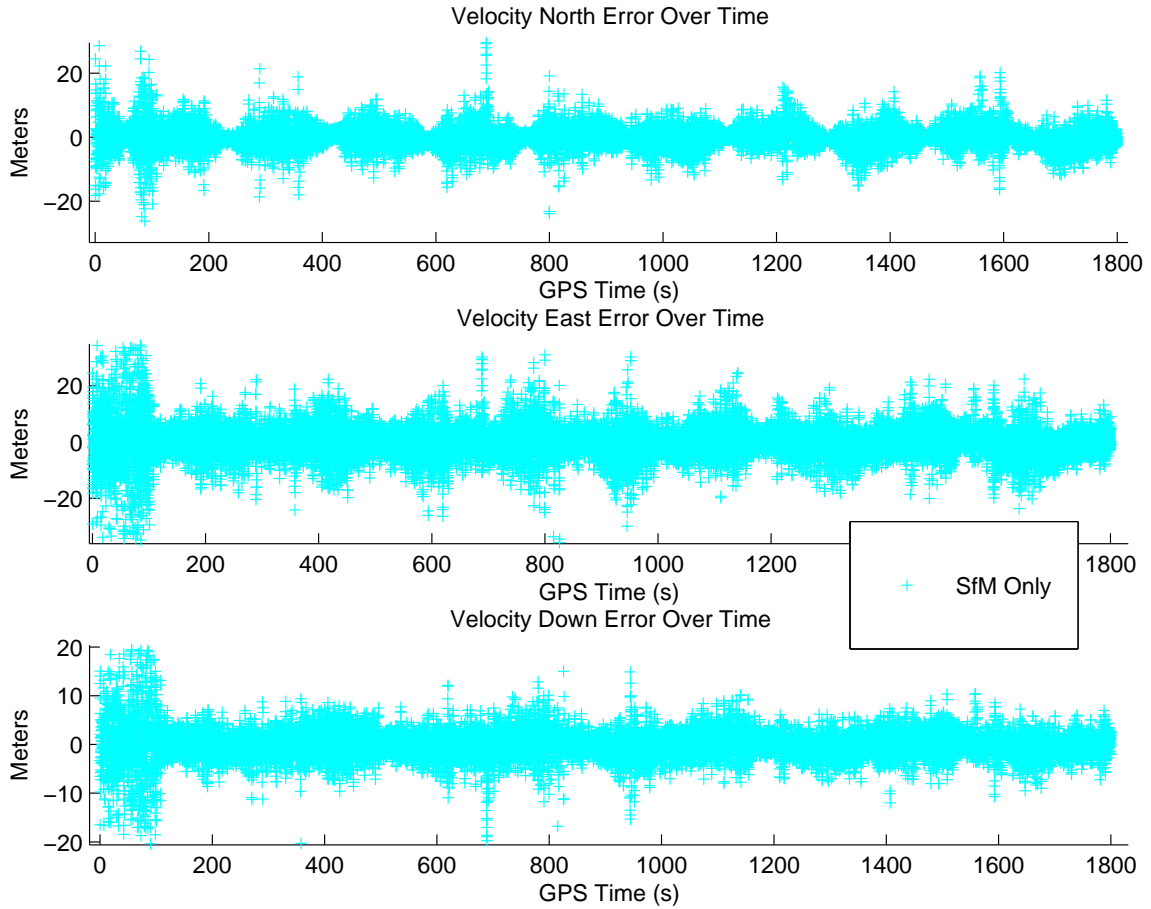


Figure 3.8: SfM Velocity Errors

Figure 3.9: Alignment Velocity Error. Error in the aligned SfM velocity after the GPS outage. The errors appear to mimic the velocity error during alignment shown in figure 3.8. The behavior of noise centered around a zero mean seems to extend beyond the alignment time frame for the SfM trajectory estimate.

this in the model fed into SPIDER, the noise was modeled as white Gaussian noise on the sensor for the velocity measurement along each axis. Due to the INS solution being used as the relative truth for this trial, the magnitude for the noise for the SfM velocity measurements was set as the standard deviation of the differences between the SfM solution and INS velocities before alignment.

The source of growing error over time that SfM is prone to is the drifting of its alignment from a real world reference frame due to the integration of velocity errors. The velocities should have a relatively accurate magnitude over time but the directions they point will slowly drift away from the real world reference frame they were aligned to. This degradation of alignment was modeled in SPIDER as a continual increase in the cross covariance of the measurements over time determined empirically.

The SfM alignment errors are intrinsically lower than the growing biases of the IMU. The SfM drifts much slower and will maintain the same accuracy of relative changes in position over a longer time. The fundamental difference in noise between these two measurement sources allows their combination to create a better estimate of the navigation states than using either two alone.

3.7 SfM Aiding Results

The velocity measurements from SfM and the simulated IMU measurements combined in SPIDER output a solution that stayed closer to the INS solution over a much longer time compared to the free running IMU. This result implies that the velocity measurements provided extra navigation information to the system that helped correct for the growing biases in the IMU.

The trials looked at from the Minor Area Motion Imagery (MAMI) dataset were the DEBU 2 and MSEE 1 flights. These were all done during the day and flew in circular paths over the ground. As was described in Section 3.5, the matches were specified to be over a short enough time to prevent the solution from self-correcting at a later time based on much earlier measurements of the same ground location. The circular motion did allow features to stay in view for the camera over a longer time which aids matching, thus SfM solutions of straight and level flight may not share the same amount of matches between images as this trial experienced. The time limit for matching in VisualSfM should help limit this discrepancy.

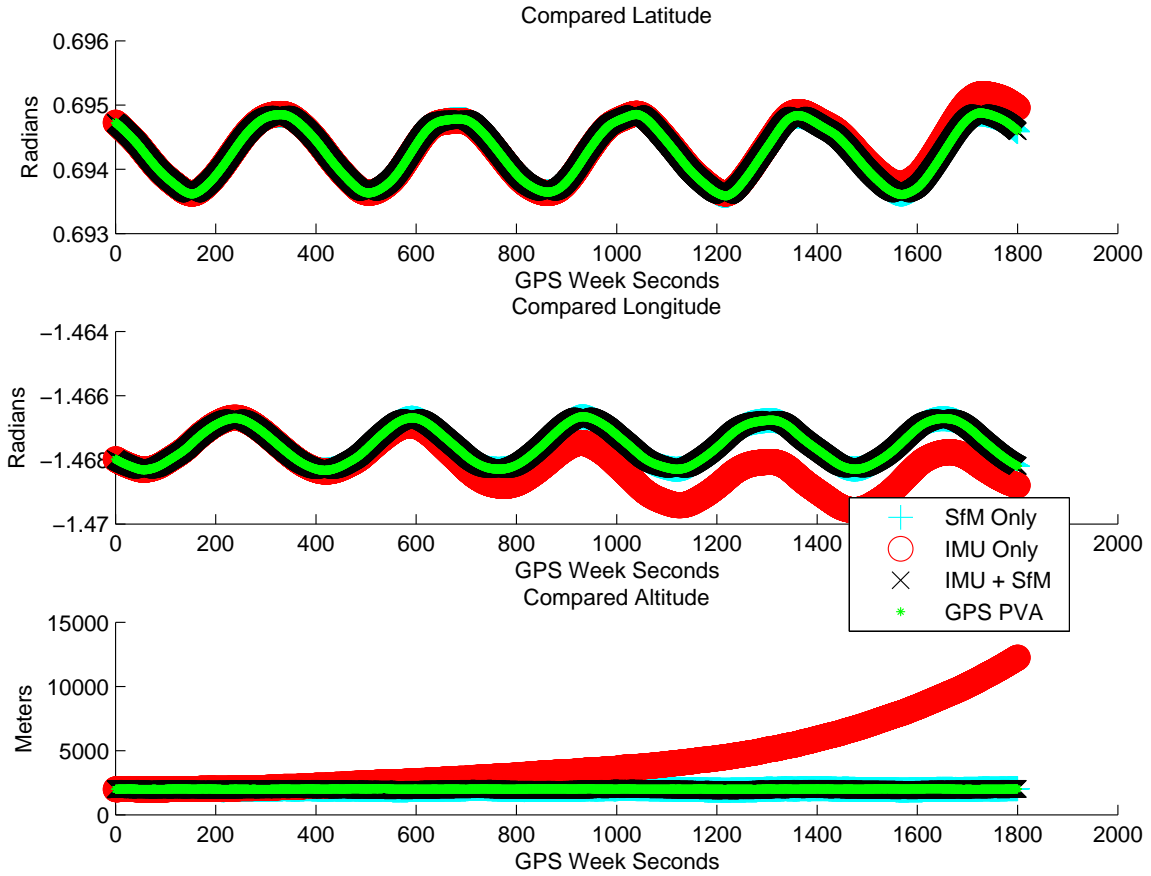


Figure 3.10: Latitude, Longitude, and Altitude Of All Solutions. This figure shows the values of latitude, longitude, and altitude, for all solutions considered in this research for a single simulation of the IMU. The free running IMU solution in red drifts away from the GPS truth reference shown in green, a common theme that will appear across all figures in this analysis. In comparison, the SfM only and combined SfM and IMU solutions stay much closer to the GPS corrected solution.

Spikes of rapidly changing states appear in the combined solution plots around 850 seconds into the MSEE 1 data due to a timing discrepancy in the images used to perform SfM, but they do not affect the quality of estimates after those points.

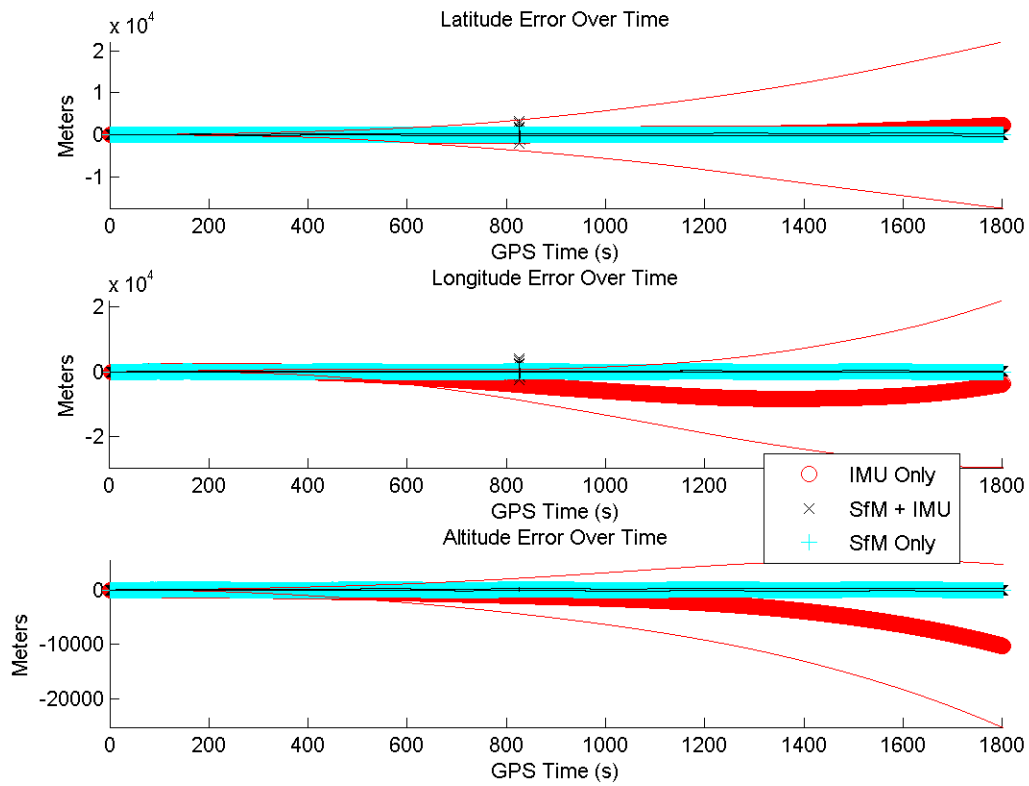


Figure 3.11: Position Error with Confidence. This figure shows the errors in position with covariance of the solutions plotted. The growth in the free running IMU dwarfs the confidence in the combined solution with SfM velocity updates.

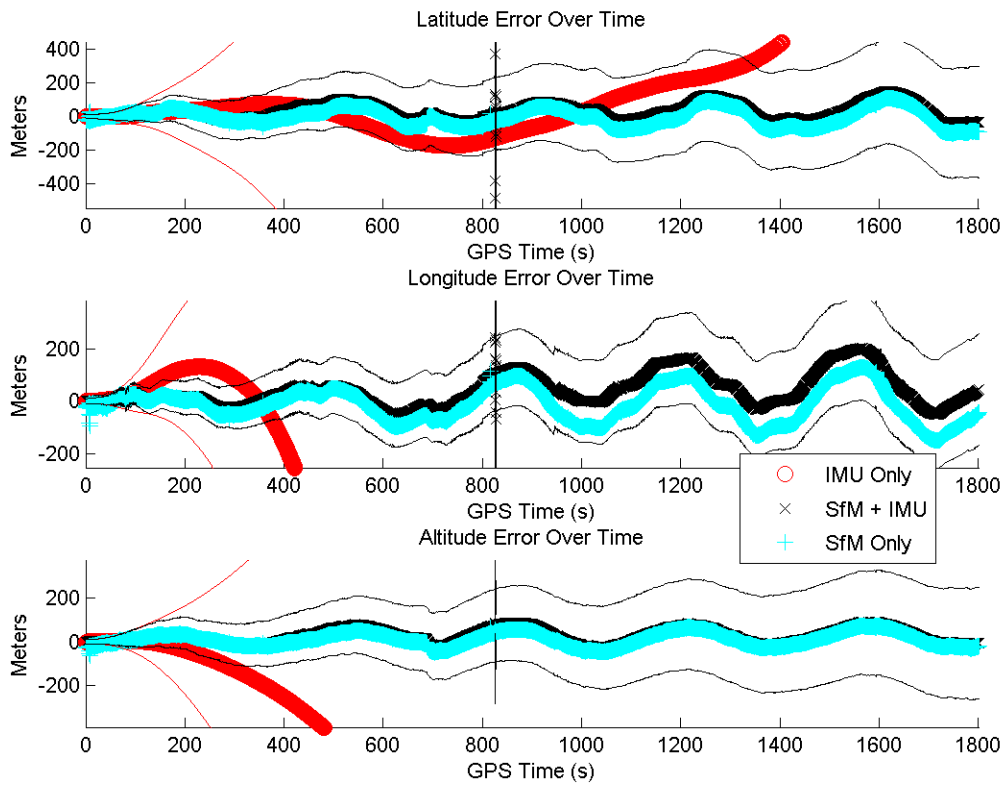
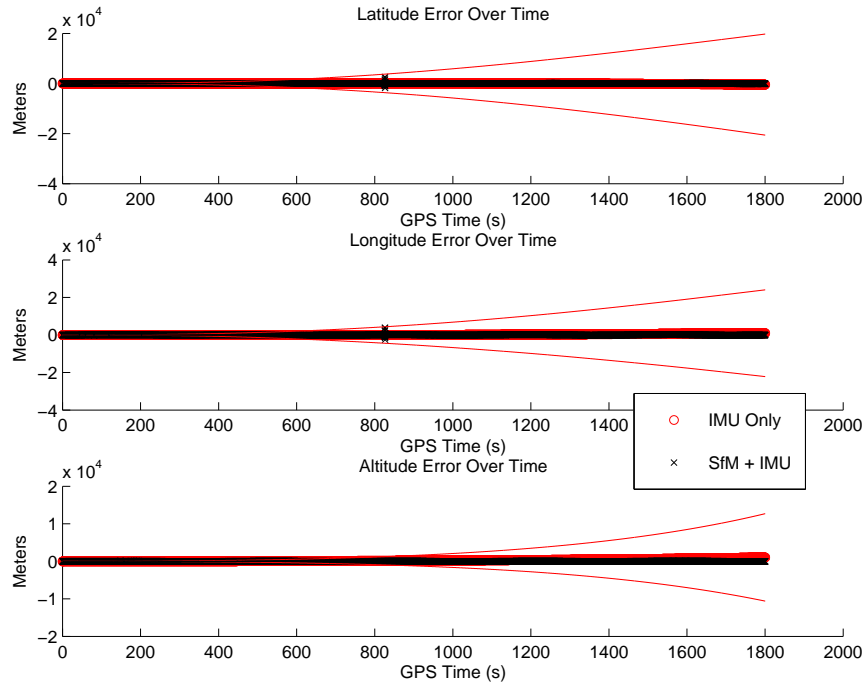
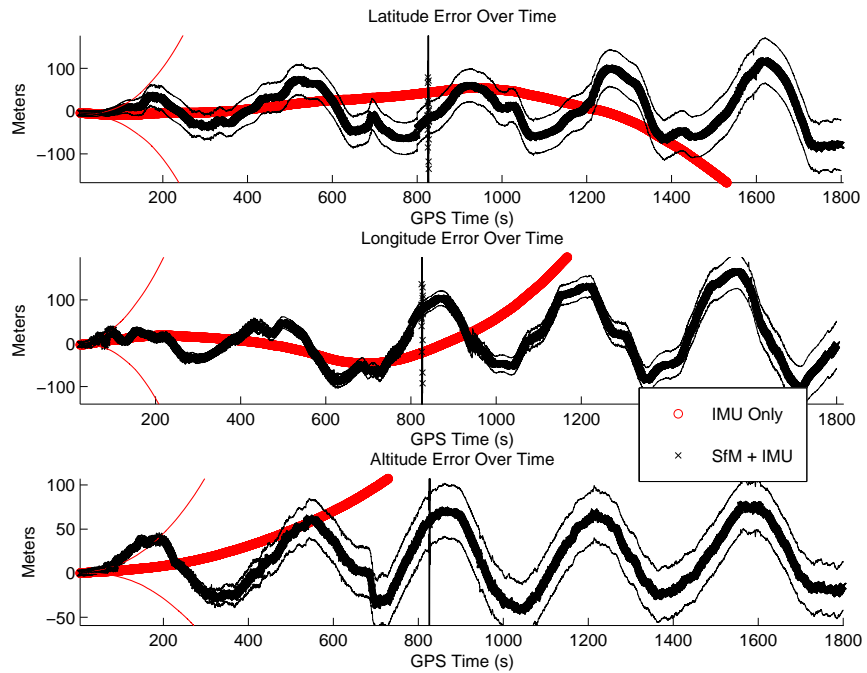


Figure 3.12: Position Error with Confidence Zoomed In. The zoomed in view of position error illustrated the growth of error in the SfM only and SfM aided solutions. The aided solution tracks error in the SfM only estimates, but maintains error bars that encompass the GPS truth reference. We see the error in SfM solutions grow over time corresponding to decreasing confidence in the IMU as well as integration of the SfM velocity errors.



(a) Position Error



(b) Position Error Zoomed In

Figure 3.13: Position Error with Variance over 250 Simulations. The thin lines show the variance in position error over 250 simulations of the IMU. The effect of the different IMU simulations change the SfM aided solution much less than the IMU only solutions. The combined solutions maintain an error of around 100 meters in each axis .

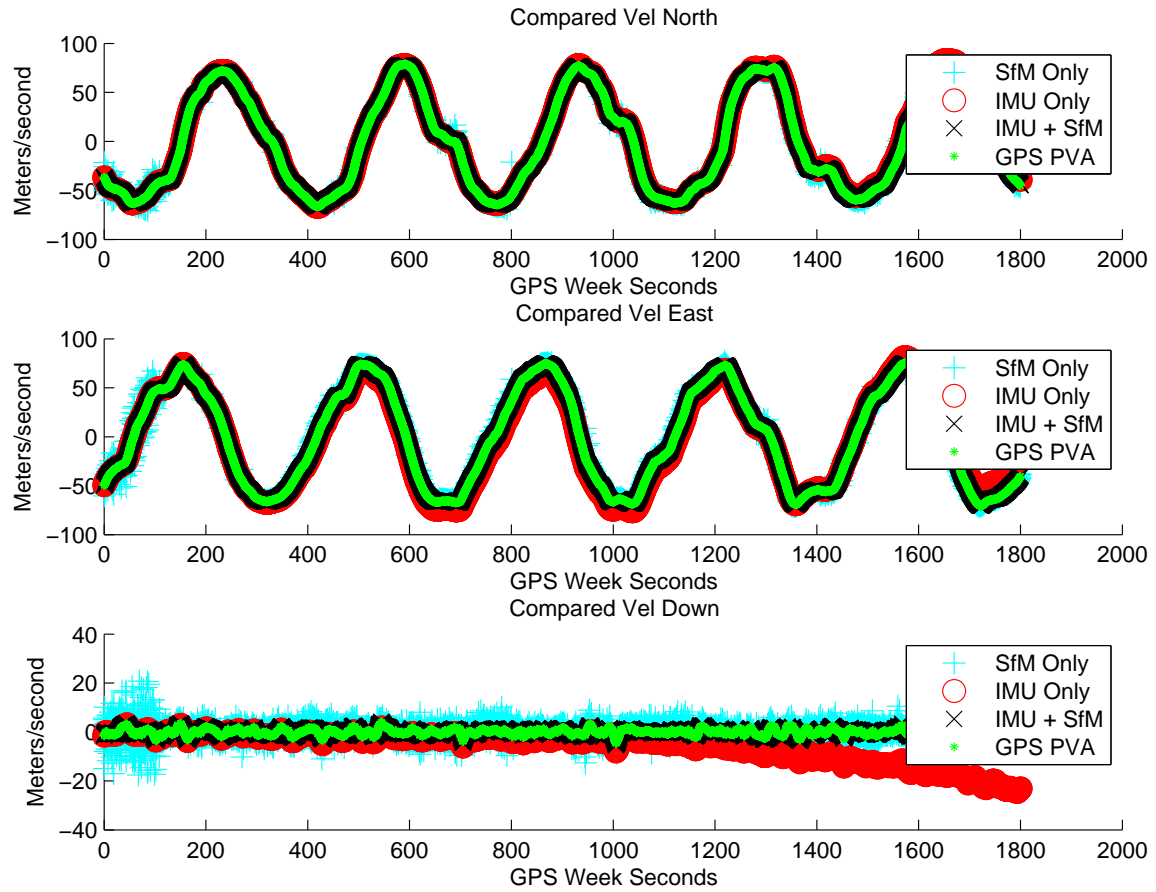


Figure 3.14: Velocity Of All Solutions. The IMU only velocity drifts away over time compared to the other solutions. The SfM solution has a large variance around the truth throughout the trial.

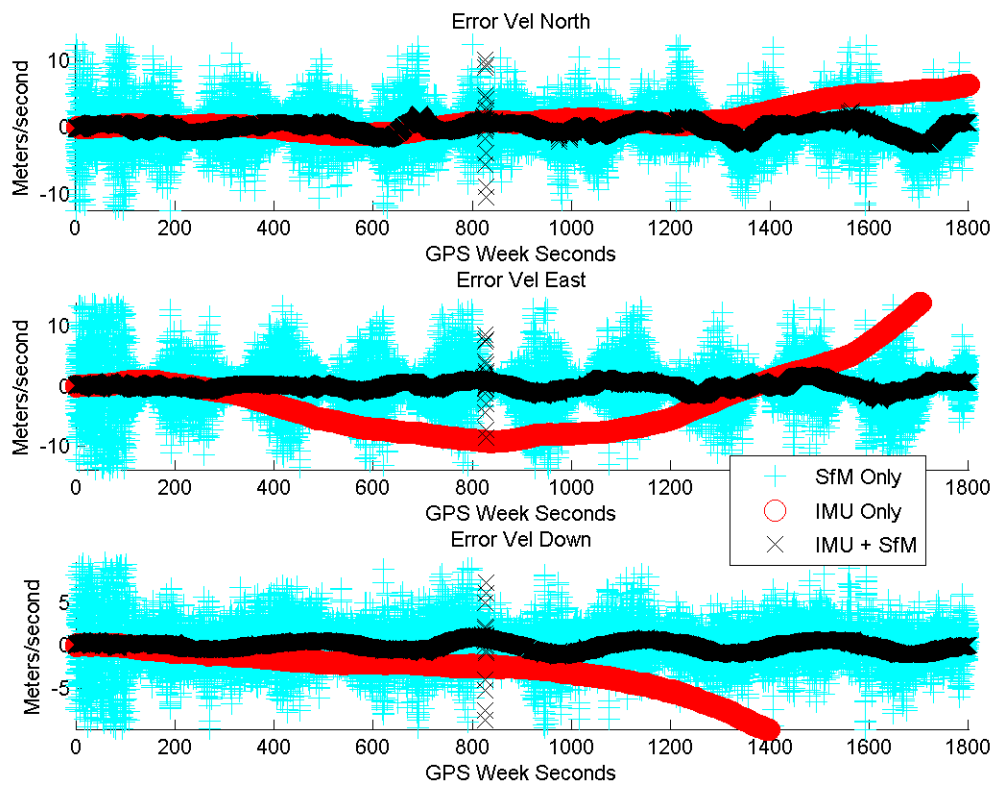


Figure 3.15: Velocity Error Of All Solutions. The combined SfM solution has significantly less variance than the SfM only solution. The combined solution has observability on the growing biases in the IMU and does not track the inertial sensor drift away from truth.

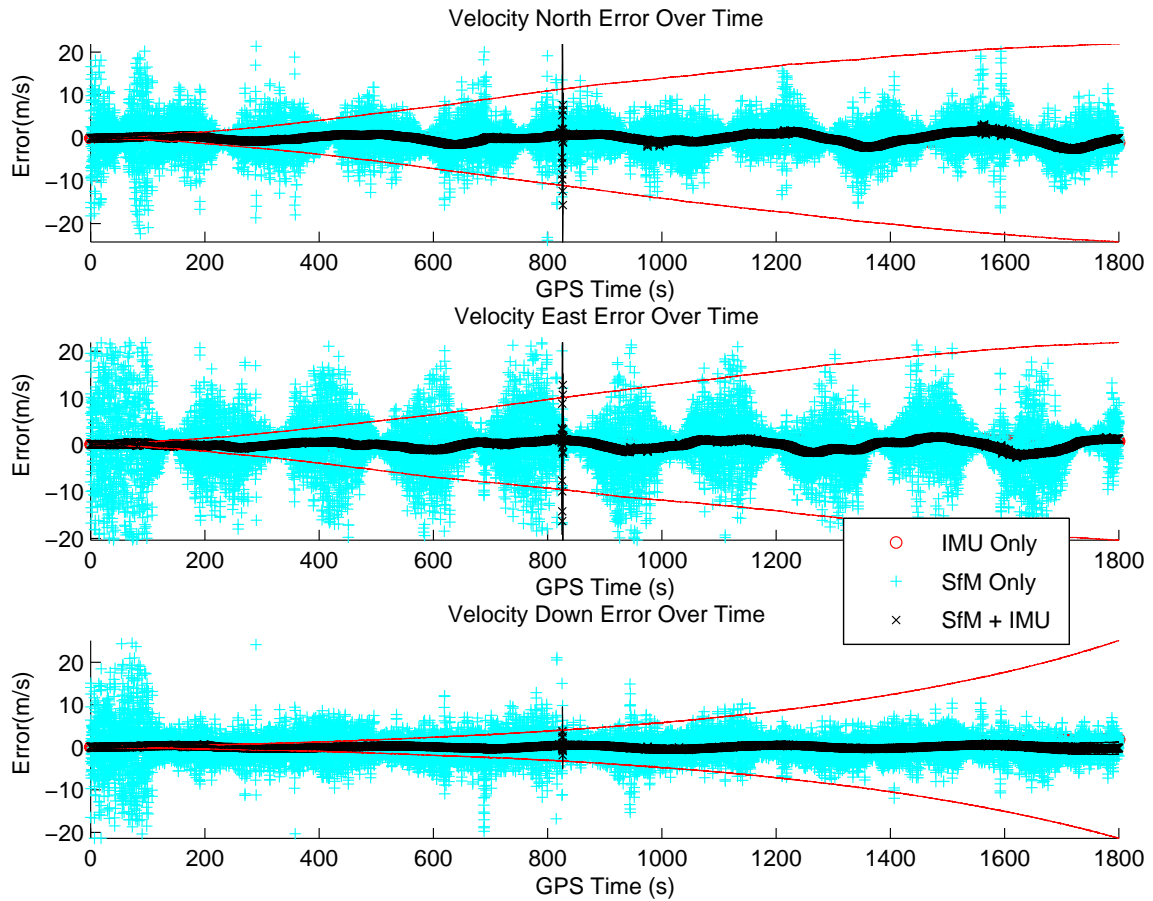


Figure 3.16: Velocity Error Over 250 Simulations with Variance. The variance in the SfM combined solution is much smaller than the IMU only solution as the SfM velocity measurements are the same for each IMU simulation.

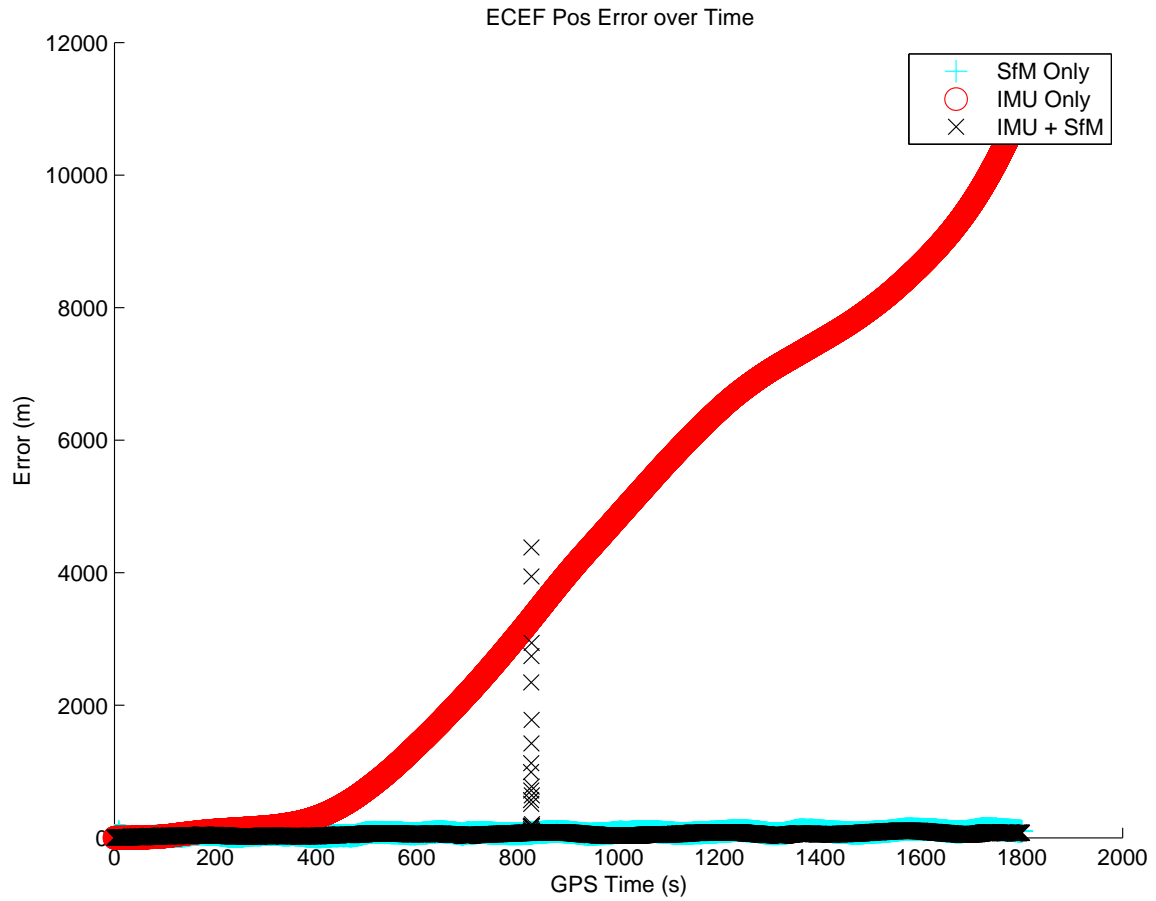


Figure 3.17: Average Position Error Over 250 Simulations. The total error in the combined solution is much less than the free running IMU solution. The average of the IMU only position error after 30 minutes is over 12000 meters, dwarfing the error seen by the solution using SfM aiding.

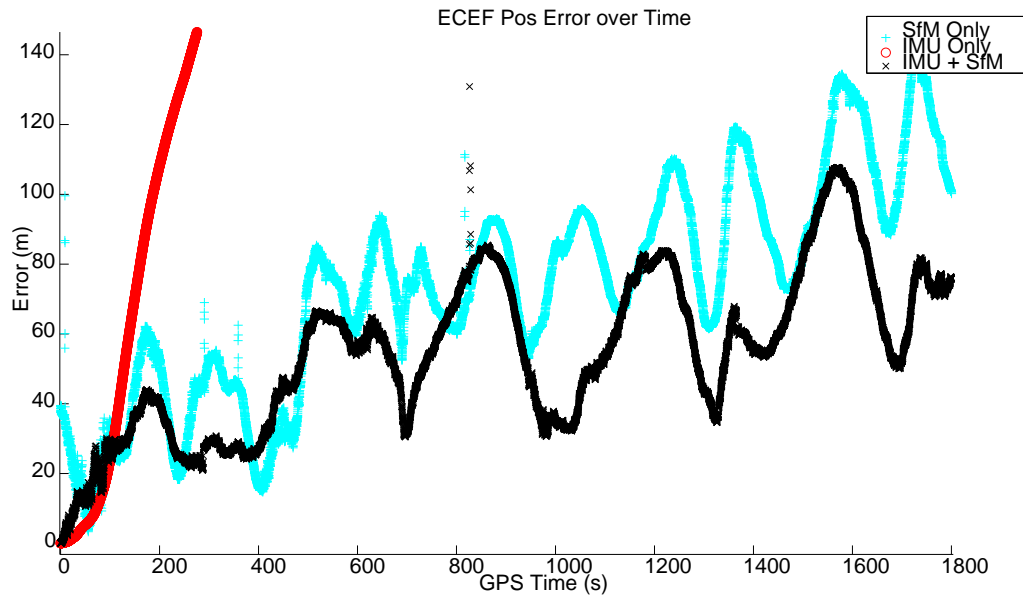


Figure 3.18: Average Position Error Over 250 Simulations Zoomed In. The total error in the combined solution appears to drift less than the SfM only solution over the flight. The average error of the combined solution only barely exceeds 100 meters over the 30 minute trial. Position errors in the combined solution seem to correlate with SfM only spikes in error due to the drifting alignment of the SfM solution.

3.8 Weaknesses of Visible Light Cameras

Visible light cameras are ideal for vision aided navigation under many circumstances. They are extremely cheap compared to similar quality cameras of other parts of the electromagnetic spectrum, they have very high resolution, and they can see very clear contrast between objects in a well-lit scene. For more extreme navigation purposes though, sensors need to be resistant to almost all plausible situations an aircraft might encounter.

Visible light cameras do not function well through artificial vision-occluding particles as these block most of the spectrum. The cameras essentially don't work at all at night time due to the low contrast between objects. These two weaknesses are of large importance to military operators as operation in these two cases is very regular, and robustness of solution is highly valued. A new tool is necessary to give vision aided navigation robust flexibility to these situations.

IV. Medium Wave Infrared (MWIR) Cameras for Navigation

Previously, this thesis showed how position estimates from an SfM solution can aid a free running inertial sensor. Interestingly, SfM is able to build such position estimates from MWIR imagery as well. Unfortunately, the MAMI dataset lacked the precision timing needed to integrate the MWIR-based position updates with the free running IMU. However, the quality of the SfM only position estimates from the visual spectrum imagery can be compared directly against those generated from the MWIR imagery, thus suggesting whether or not the contributions of an SfM solution based on MWIR imagery might be expected to provide similar benefits if it were properly timed.

In some of the areas where visible light cameras struggle to detect any features in a scene, Infrared (IR) cameras maintain a similar level of functionality to normal operation. In particular, the 3-5 μm infrared band (termed MWIR) stands out as a candidate for navigation due to its functionality at night and relatively high resolution for the infrared domain. The unique strengths of MWIR cameras make them ideal for use in critical systems that require operation at night, most notably in military and search and rescue missions. Their characteristics will be explored in comparison to EO camera solutions to prove their usefulness for navigation.

4.1 Equipment

The MWIR camera used in this research captures images in the MWIR domain from roughly 3 – 5 μm at 1024 x 1024 pixels and operates at 30 Hz. It was attached to a gimbal and pointed in the same direction as the EO camera explored previously. The exact fields of view for the EO and MWIR cameras were not recorded, but the amount of scene content viewable in each MWIR image is less than that in corresponding EO images. MWIR are

already present in many military systems for use in sensing operations, making the potential to aid in navigation for it of particular interest to military users.

4.2 Dataset

The MWIR camera was set to take images at full resolution and frame-rate in the MAMI dataset. Unlike the EO camera, the server connected to the MWIR camera was not configured to tag images with with precision timing. The sensor data was saved in a raw format that required post-processing with functions provided by AFRL to obtain usable images for navigation. MWIR was run on both of the same trials as the EO cameras, but with a single additional nighttime data collection.

4.3 Infrared Imaging

The light that the human eye can see lies in the 400 *nm* to 700 *nm* wavelength spectrum, which is only a very small slice of the electromagnetic energy in our environment. Extending beyond the visible spectrum is the IR spectrum in the energy bands between 700 *nm* and 14 μm [24]. EO cameras work in the visible spectrum, while infrared cameras sense different slices of the IR spectrum depending on intended usage.

4.3.1 Construction of Infrared Images.

According to Garnier et al. [26] the image of thermal radiation created by an infrared sensor is a function of spectral radiance, spectral irradiance, and spectral flux of the objects in the scene as well as the detector spectral responsiveness. These are considered the material properties of an object. These factors determine the amount and frequency of energy leaving visible objects, and then degradation of amplitude and phase by the atmosphere between the objects and the camera. This energy is collected as photons interacting with the camera's optics along with atmospheric background noise. A detector in the camera reads and saves these pixels as an image.

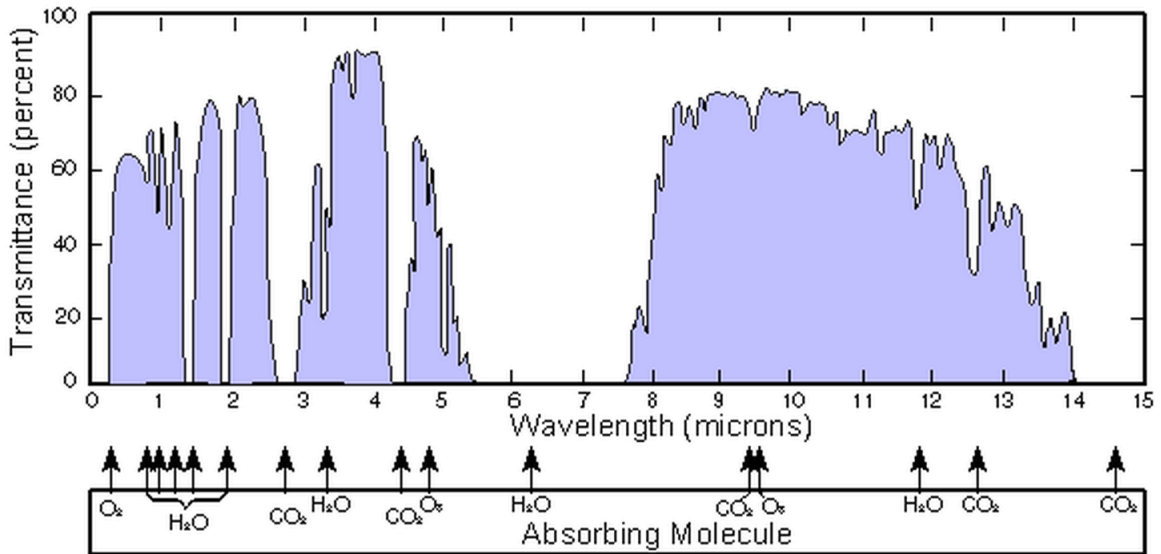


Figure 4.1: Atmospheric Transmittance[25]. The transmission of infrared energy in the atmosphere. Infrared energy can be subdivided into three different bands that do not experience severe atmospheric absorption. The 700 nm to 3 μm band is considered a blend of the Near Infrared (NIR) and Short Wave Infrared (SWIR) region. From 3 μm to 5 μm is the MWIR band. Beyond that is the Long Wave Infrared (LWIR) band occupying the 8 μm to 14 μm wavelengths. The wavelengths that were omitted from these selection of bands are mostly absorbed by the atmosphere. These distinctions of bands follow the guidance of passive sensor design and not the more rigorous definitions used by those in the scientific community characterizing the entire electromagnetic spectrum.

The sum of these factors that describe emission based on material properties is called emissivity, which is directly inverse to reflectivity (the amount of energy that a material will reflect in a certain band). The paper by Nandhakumar and Aggarwal [27] compares emissivity in the infrared spectrum of objects that are found in natural scenes. Their research uses a Long Wave Infrared (LWIR) camera to measure the amount of emitted versus reflected solar radiation for natural scene objects. The objects all show a high bias in emissivity due to the longer wavelength IR spectrum being more of a function of emissive energy rather than reflected energy, but there are differences in how much energy the object will give off based on its heat that become more apparent at shorter wavelengths. The objects that they looked at were among broad classes specified as buildings, vehicles,

vegetation, and pavement. These objects can be found in my different natural scenes seen by navigation systems.

Signals visible in the IR spectrum vary from reflected energy at shorter wavelengths, which is similar to the visible spectrum, to completely emissive heat radiation at longer wavelengths. Energy in the EO wavelengths is almost completely reflective in a natural Earth environment. This visible light is created by a separate source and then reflected off of objects in the scene. Cameras that operate in the EO spectrum can only capture the energy in this band. At the higher ends of the infrared spectrum, most of the energy that comes from the environment is emissive heat radiation. This type of radiation is emitted by objects themselves as a function of the object properties and temperature.

4.3.2 Unique Infrared Characteristics.

The difference in temperature between objects and their surroundings creates distinct contrast in a thermal image. Thermal conduction is the phenomenon wherein the temperature difference between two objects in direct contact fades over time as heat is transferred from the hotter object to the object with less heat [28]. This causes infrared images to have less distinct edges present as conduction creates a blurring effect along edges of an object that are in contact with another conductive material.

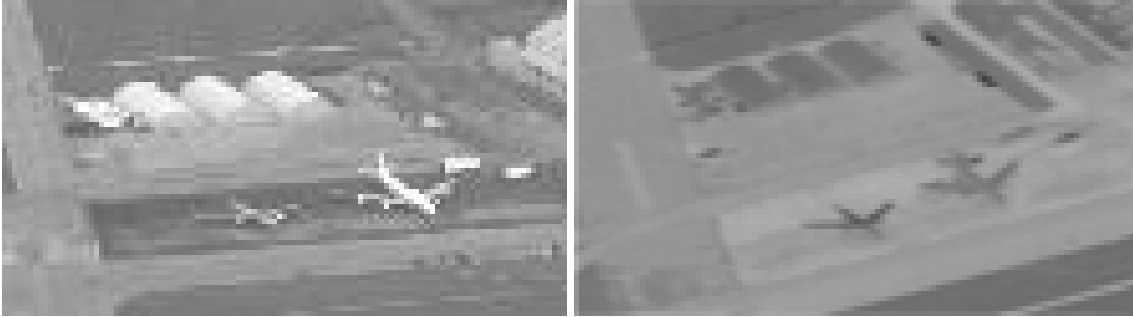
Transmission of infrared radiation is much more easily blocked by any type of solid medium compared to visible light. Even very thin plastic objects that are transparent in the electro-optical spectrum can completely block the transmission of infrared energy. These effects of material properties are much more obvious in IR images due to their heavy impact on the appearance of objects in a scene.

Vision systems looking at longer wave infrared radiation perform much better than visible light cameras under situations where the sun's light is absent because a temperature difference still exists between objects in the scene. At night, visual, Near Infrared (NIR), and Short Wave Infrared (SWIR) cameras are less effective because there is no strong

source of reflective energy to illuminate the scene. The emissive radiation coming from distinct objects in a scene in the MWIR and LWIR spectrums is contrasted to that coming from the Earth which both heats and cools at a different rate. These types of IR sensors can still passively observe the environment at night with comparatively little degradation in solution quality.

Thermal crossover is a phenomenon wherein a target of interest has the same signature as its background in an infrared image. For ground structures on Earth, this happens twice a day when objects that absorb the sun's radiation and the ground become the same temperature and radiate at a similar intensity. At this point it is very difficult for thermal imaging sensors to pick up individual objects in a scene as everything is at the same level of thermal radiation. This happens both in the morning as the sun heats the ground up and at night as the ground cools. On opposing sides of these crossovers, images of a scene will look inverted in comparison to each other (Figure 4.2a and 4.2b) because the emissions from the above ground objects will either have more or less heat relative to the backdrop of the Earth's surface.

Infrared cameras have also been known to increase visible range under certain vision impairing conditions compared to the visible light spectrum. Driggers et al. [24] shows that the amount which aerosols obstruct electromagnetic energy decreases with longer wave infrared sensors. In addition, Beier and Gemperlein [29] looked at the spectral transmission over the Electro-Optical (EO) and Infrared (IR) bands given varying visual range conditions on the ground. At a mild atmospheric interference, defined in the paper by a 1220 meter visual range on the ground, SWIR, MWIR, and LWIR all have significantly higher atmospheric transmission than in the visual range. The atmospheric interference was created with an artificial aerosol. Heavier atmospheric conditions created by radiative fog showed less of an improvement for the infrared bands, especially the SWIR and MWIR bands which were fairly similar to the visible band. We can still see that longer wavelength



(a) MWIR Day Image

(b) MWIR Night Image

Figure 4.2: Different Sides of Thermal Crossover. Two images taken from the MAMI dataset, one during the day and one at night. Both pictures show the same two aircraft and set of three hangars behind them on WPAFB. In the day image these objects appear much lighter than their surroundings due to the metal absorbing heat and having a unique infrared signature. At night, the metal loses more heat than the ground and appears darker in comparison. This relative appearance change happens during thermal crossover.

infrared sensors have much more penetration through unnatural vision occluders such as smoke. This property gives them robustness when used in sensing operations.

All of these effects cause an infrared image to have distinct differences in comparison to an image of the same object taken in the visible spectrum. These differences manifest themselves both as strengths and weaknesses in terms of navigation. Understanding both allows for the most effective implementation of such sensors.

4.3.3 Medium Wave Infrared (MWIR).

This research concerns the use of MWIR cameras and how they specifically compare to EO views of a scene. The portion of the infrared spectrum that occupies wavelengths between $3\text{-}5\ \mu\text{m}$ is considered MWIR. This portion of the spectrum can see both reflected and thermally self-emitted energy [30]. The $3\ \mu\text{m}$ to $4\ \mu\text{m}$ band has a very high transmittance in the atmosphere as well as another small band above the gap absorbed by CO_2 in the atmosphere which MWIR cameras can also look at (Figure 4.1). The MWIR bands of energy are especially sensitive to vehicle exhaust which is why cameras that sense in this band are more commonly found within military applications [31].

MWIR cameras often have higher resolution than their LWIR counterparts due to a smaller wavelength of light being captured, thus making them better suited for situations requiring higher image fidelity — as in vision-aided navigation. At the same time, MWIR cameras are more robust at night in comparison to SWIR sensors as they capture more emissive radiation from the scene. Such advantages, combined with the stated benefits over visible light cameras (particularly the ability to function at night), make MWIR cameras an interesting tool to investigate for its usefulness in vision-aided navigation.

4.4 Contrast Enhancement

In both cases where MWIR cameras are advantaged over visible light cameras (i.e. in vision occlusion and at night) the contrast of the images output is greatly decreased. In order for feature detection methods to work on these images, the contrast must be enhanced to highlight features. This research utilized a particular contrast enhancement method developed for analyzing very large images [32]. The method equalizes and spreads the histogram of intensity values in an image as to draw out contrasts hidden when the image is focused around a certain portion of gray-level. The strengths of this method are that the amount of spreading is established by a single parameter, allowing for easy implementation of a best value determination, and that it doesn't wash out existing features from before the spreading.

$$d = \frac{0.5\pi\log(20p_s + 1)}{\log(21)} \quad (4.1)$$

$$c = \frac{1}{2}\tan(d) \quad (4.2)$$

$$px_{\text{new}} = 0.5 + c\tan(d(2px_{\text{old}} - 1)) \quad (4.3)$$

Equations 4.1 to 4.3 show the calculations done on a pixel by pixel basis of the image to determine their new values. p_s is the parameter that controls the amount of spreading in each image. The first two equations create the variable d and c that are derived from p_s to simplify the final calculation. The final equation calculates px_{new} , the new value of the pixel being looked at, using c , d , and px_{old} , the pixel value before spreading. The pixel values are the gray-scale equivalents of the original picture.

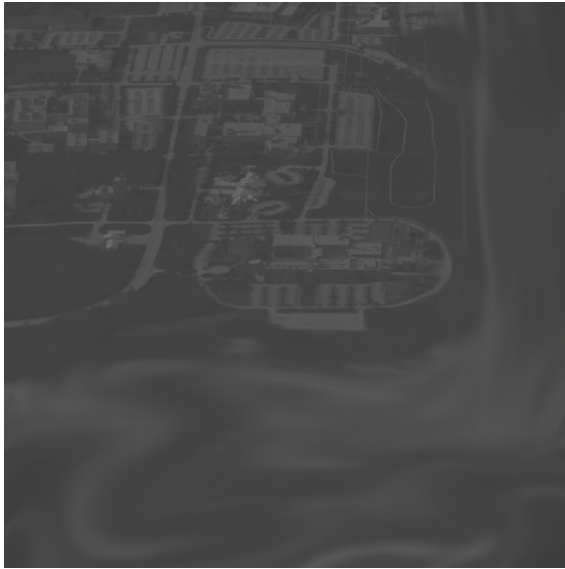
For this research, the ease of changing the spreading parameter per frame allowed for automated tuning based on the number of features detected in each image. As discussed later, the amount of features detected in MWIR images is dependent upon objects in the scene and dynamic parameter tuning during a trial is sometimes necessary to detect enough features for navigation.

4.5 MWIR Image Comparisons

MWIR images are based upon object characteristics, temperature differences, and present light in a scene. The same scene in visible light images will have distinct differences that can affect the quality of image matching over these areas. This manifests itself in Structure from Motion (SfM) as features both detected and matched across images, integral parts of determining the best fit solution between images. This section explores the differences in feature detection of specific portions of natural scenes to paint a picture of where features can be found and where they are lacking in both EO and MWIR images.

4.5.1 Vegetation.

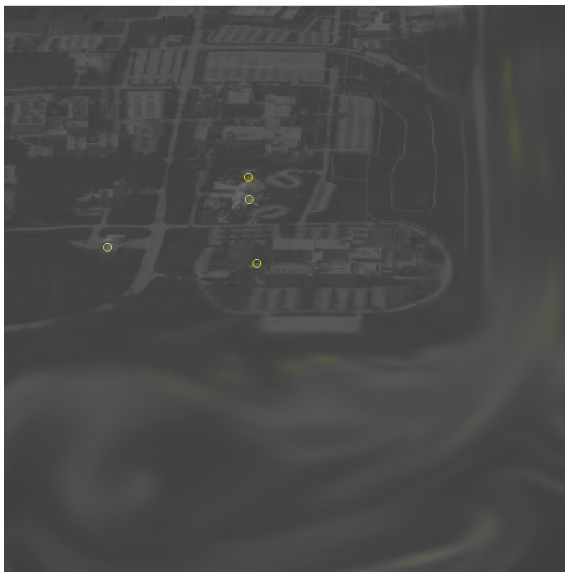
Vegetation tends to have very low feature density in MWIR images. Vegetation maintains a similar heat level across all of the plants included and they have very similar infrared signatures. These effects cause the images to have a blurring effect over dense vegetation. EO images do not tend to have the same problems in this environment. MWIR



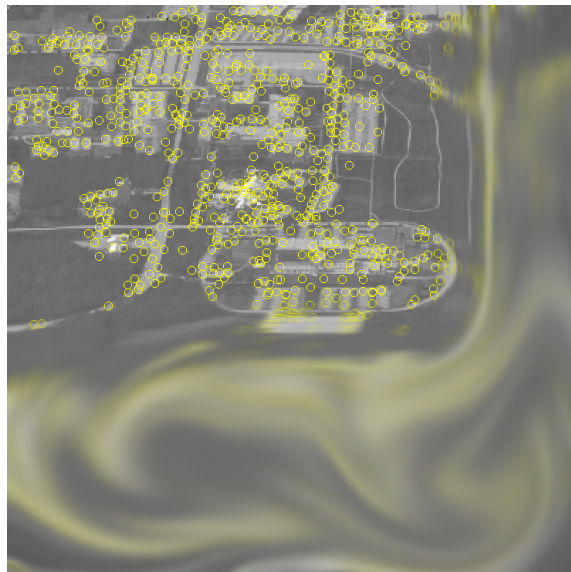
(a) MWIR Image Raw



(b) MWIR Image Enhanced



(c) MWIR Raw SIFT (37 Features)



(d) MWIR Enhanced SIFT (1695 Features)

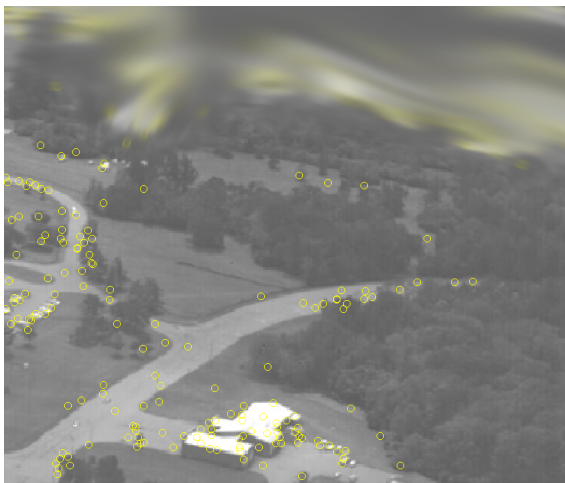
Figure 4.3: Histogram Spreading Example. The two images represent the same scene before and after histogram spreading implemented in this research. The image comes from the MWIR camera during the MSEE 1 Trial in the MAMI dataset. The number of features detected by SIFT increased from 37 to 1695 after contrast enhancement. This improvement is very useful to navigation as there are more potential matches between images for a given scene.



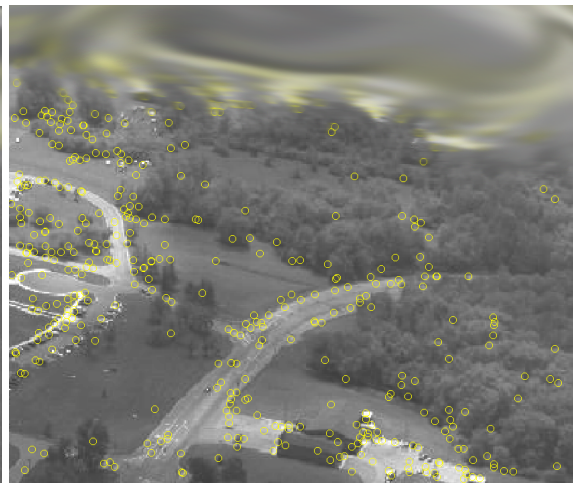
(a) MWIR Vegetation



(b) EO Vegetation



(c) MWIR SIFT (476 Features)



(d) EO Vegetation (772 Features)

Figure 4.4: Comparison of Vegetation Appearance. Two images from the same trial in the MAMI dataset. The images show a road coming from the right and curving downwards that divides two large section of wooded area. The trees appear less distinct in the MWIR image compared to the EO image. The SIFT images show features detected in the woods of the EO image while only capturing the edges of the forest in the MWIR image.

cameras might struggle to navigate over scenes highly populated with swaths of trees or plants.

4.5.2 Artificial Structures.

Man-made buildings in the MWIR domain also look different from their appearance in similar visible light images. Edges and corners are strongly contrasted against the ground based on temperature and material differences. Roofs with different materials appear lighter or darker than each other even at the same temperature. Across the roof itself, unless it is broken up by different objects there is much less feature density compared to those in visible light images. A large amount of smaller buildings and many roads or parking lots in a scene would provide a much greater amount of trackable features for MWIR image navigation than only a few large structures.

4.6 SfM Comparisons

With the unique characteristics of MWIR images in comparison to those from the visible spectrum in mind, the whole navigation solutions can be compared to determine the level of similarity between the two sources. Again, as precise timing information was not available for MWIR images, MWIR SfM was not able to be directly compared to real world positioning information. As SfM was used on both domains, comparisons directly between their trajectory solutions provide a good way to observe similarity.

4.6.1 Point Cloud Alignment.

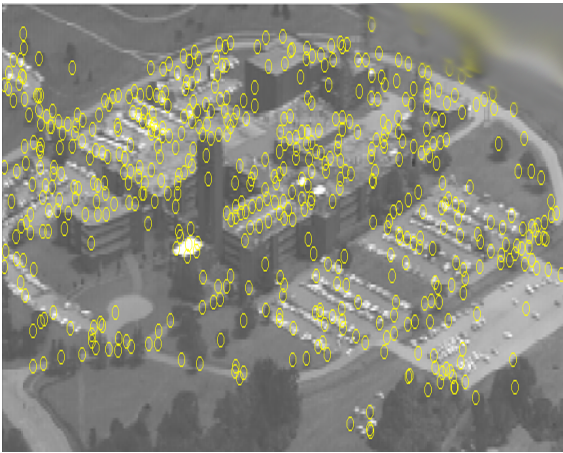
As SfM estimates positions and angles in arbitrary reference frames for each set of data it is given, the MWIR and EO solutions were not inherently aligned. As timing data was not present, the trajectories could not be compared to match their frames. A comparable aspect between the two domains does exist, and is also another useful point of analysis, which is the three dimensional reconstruction of the scene from matched images features. Matching these point clouds is a method of aligning the two reference frames and also illustrating the differences in feature detection on certain parts of the scene.



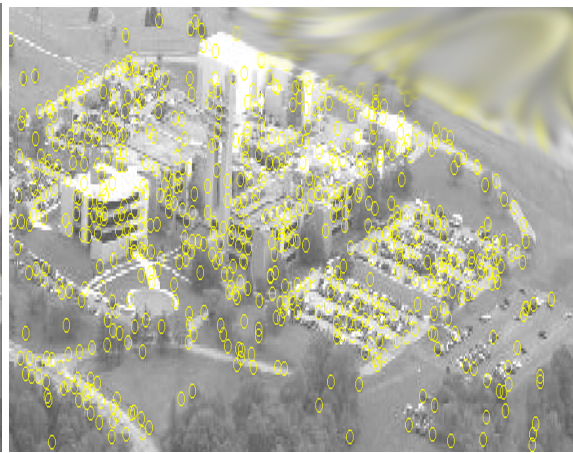
(a) MWIR Buildings



(b) EO Buildings



(c) MWIR SIFT (662 Features)



(d) EO SIFT (928 Features)

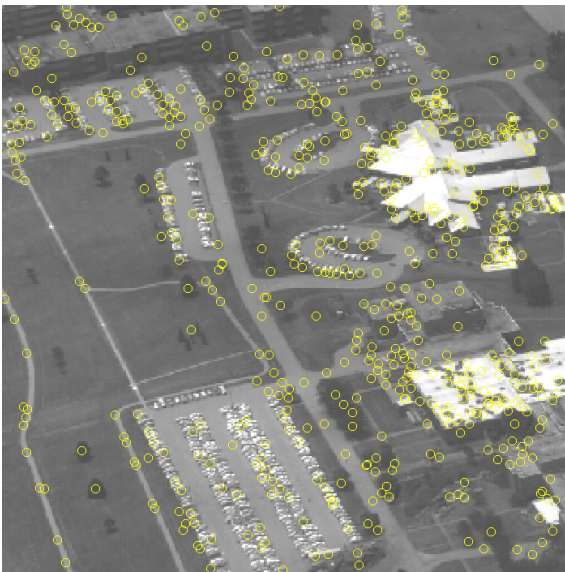
Figure 4.5: Comparison of Artificial Structures. The images show a cluster of buildings surround by parking lots in both the MWIR and EO domains taken at the same time. In the EO image, the buildings appear much brighter than their surroundings whereas they are darker in the MWIR image. In the MWIR image, the structural windows are not as strongly contrasted to the rest of the building as they are in the EO image. The cars in the MWIR image are all fairly uniform in appearance as hot bodies whereas almost each one appears different in the EO image. The buildings appear to have a higher density of detected features in the EO domain.



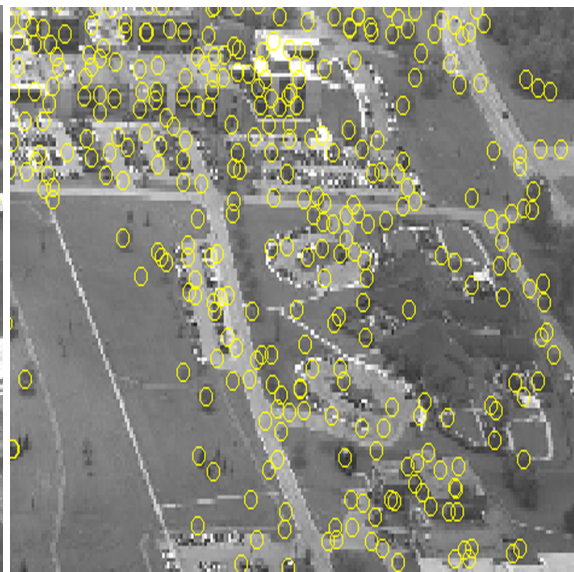
(a) MWIR Roads



(b) EO Roads



(c) MWIR SIFT (1165 Features)



(d) EO SIFT (1648 Features)

Figure 4.6: Comparison of Asphalt Appearance. The images show interconnected roads and parking lots in WPAFB. The roads appear to be slightly less feature dense in the MWIR domain than the EO domain. Both domains pick up a lot of features on the sidewalk running up/down along the left side of the image. The parking lots also appear to have similar feature density due to contrast against parked cars.

Cloud Compare was used to match the point clouds in this research. Cloud Compare is a point cloud processing program with many analysis and matching tools built into it. There are many plugins developed separately that work with the program to give more filter and analysis tools to its arsenal. It was designed to work with very large point collections, making it ideal for the scenes in this dataset. It allows pixel shading and gradients to highlight and display cloud shapes in a human friendly picture.

Alignment between the point clouds was completed by hand due to the differences in detected features between EO and MWIR SfM solutions. The matching tools in Cloud Compare rely on similar structures between the point themselves, which the differences in MWIR and EO characteristics differ too significantly. The point clouds were matched by finding corresponding points in each point cloud close to the four corners of the scene. Compared points consisted of distinct ground features that stood out in both point clouds (e.g. intersections of roads, corners of buildings, and parking lot shapes). While they did not have the same feature representation in both domains, the corners and edges of these structures were evident to a human observer. In this way, the alignment of the two point clouds provided a scaling, rotation, and translation matrix matching them together.

4.6.2 Further MWIR Characteristic Analysis.

In support of feature detection shown in Section 4.5, the point clouds from the MWIR and EO trials were inspected at the same points as the compared images of vegetation and artificial structures to further illustrate their conclusions. The point clouds were looked at from the same zoom and angle as the images compared for those sections to further allow comparisons to be built.

4.6.3 SfM Solutions.

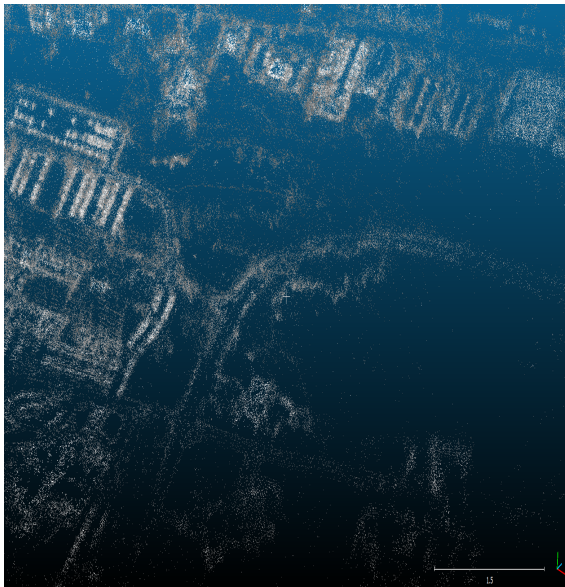
The position solutions from SfM were compared by aligning the trajectories with the transformation derived from point cloud alignment. The values coming from VisualSfM



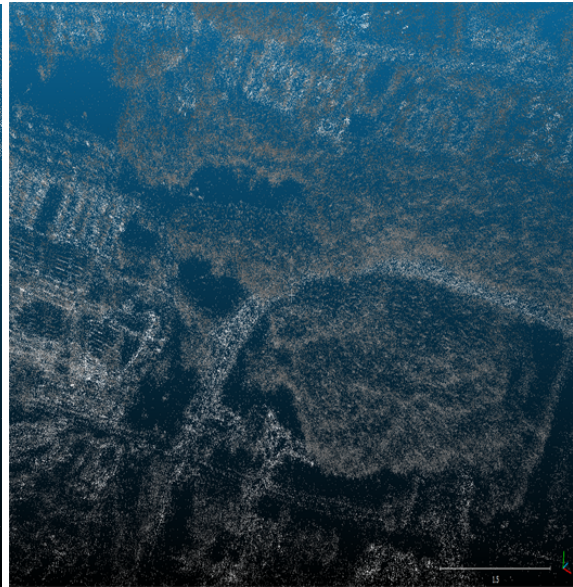
(a) MWIR Vegetation



(b) EO Vegetation



(c) MWIR Point Cloud



(d) EO Point Cloud

Figure 4.7: Comparison of Vegetation Point Reconstruction. The two top images show a scene from the MAMI dataset in both the MWIR and EO domains. The bottom images show similar views of these same locations on the ground in the point cloud reconstruction created from images in both domains. The blank point cloud over the tree clusters in the MWIR domain confirms earlier conclusions about low feature density in vegetation for the domain. This effect is not mirrored in the EO domain as the area appears feature rich.



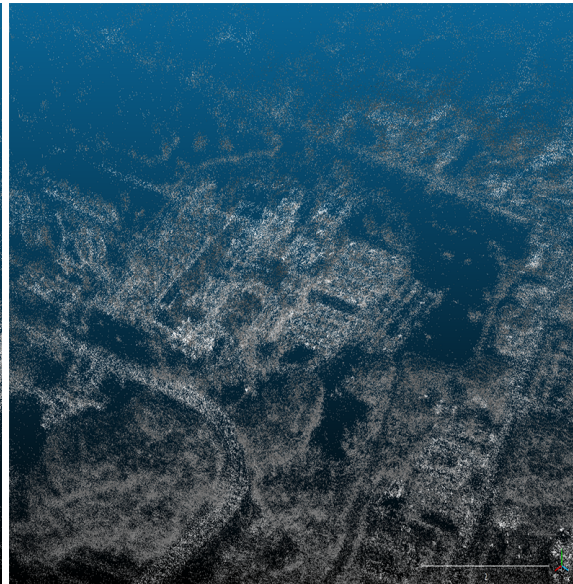
(a) MWIR Buildings



(b) EO Buildings



(c) MWIR Point Cloud



(d) EO Point Cloud

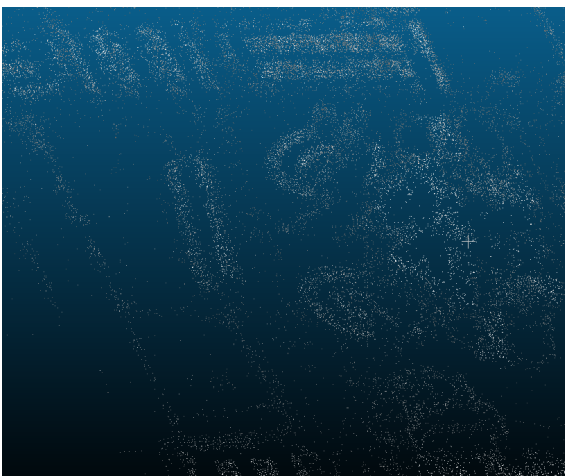
Figure 4.8: Comparison of Building Point Reconstruction. The top images are the MWIR and EO domains pictures of a building cluster while the bottom images are their point cloud reconstructions from entire trials. The roofs of buildings are void of features in the MWIR domain while the same is not seen in the EO feature cloud. The MWIR domain makes very distinct rows where cars are parked whereas the EO domain does not have such distinct lines in the parking lots.



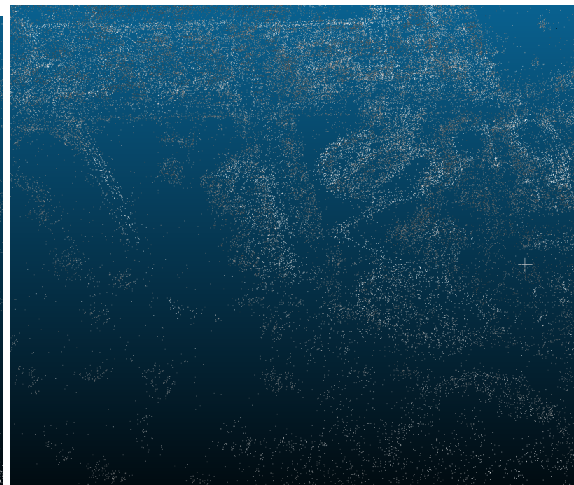
(a) MWIR Roads



(b) EO Roads



(c) MWIR Point Cloud



(d) EO Point Cloud

Figure 4.9: Comparison of Road Point Reconstruction. Additional images to support the comparison of artificial objects in a scene. The MWIR domain creates distinct outlines along the edges of the parking lots and roads, while these areas are not as unique in the feature rich EO point cloud. Again we see very distinct traces of the rows of cars in the parking lot as they highly contrast their surrounding environment.

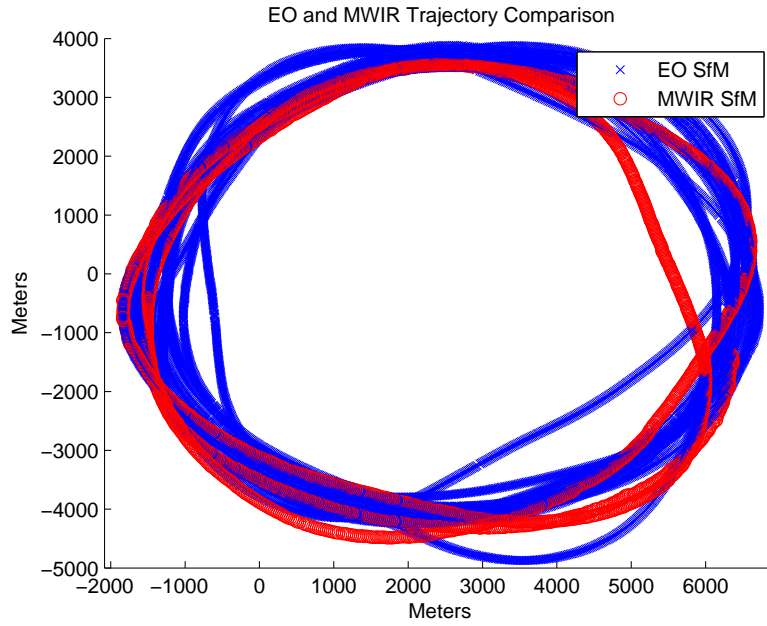
for trajectories are in the same frames as the point clouds, allowing this alignment to carry over. The unknown timing in the MWIR images prevents matching the same sections of flight together, but the plane flew a repeated pattern many times allowing the comparison of the shapes of trajectories to determine closeness of fit. The scaling of the EO SfM frame determined in alignment to the ECEF frame for the combined navigation solution allowed these comparisons to be done in terms of meters. A matching trajectory in both domains helps prove the validity of MWIR SfM solutions for navigation aiding.

4.6.4 MWIR Night Imagery.

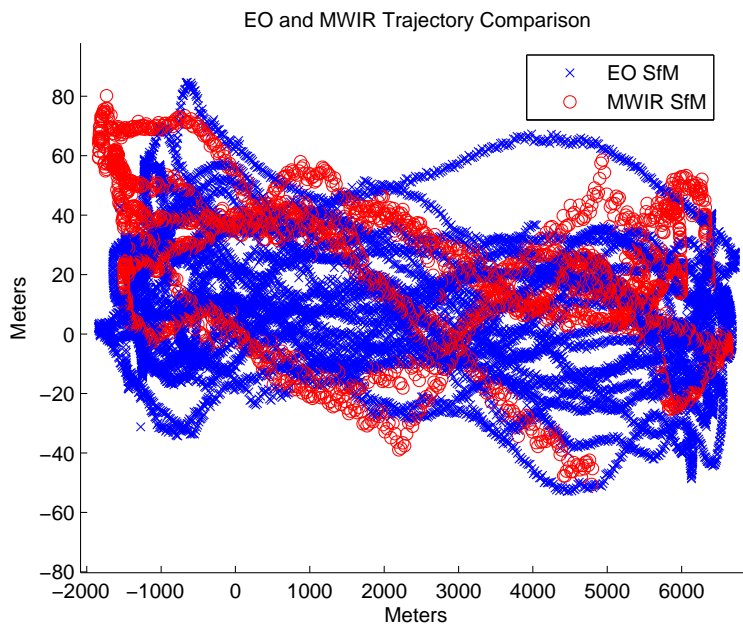
EO images taken at night can only sense independent sources of light in a scene, which for navigation purposes can not be expected and are not generally populous enough to navigate with. For this research, nighttime EO images were not explored as comparisons between them and MWIR were deemed trivial based on the lack of EO information. Instead, MWIR nighttime images were compared with MWIR daytime images of the same scene to illustrate the differences and similarities between them.

An interesting phenomena observed in MWIR images at night, and MWIR images in general from this trial, was the lasting impression on asphalt from cars left in parking lots. Night images show blurred shadows where rows of cars would be in day images. The shade provided by cars to the asphalt in parking lots creates a temperature differential that carries over in the infrared domain even hours after the cars leave the parking lot. The characteristics of this temperature difference change fairly quickly, making the descriptions of the features only valid for a relatively short time. It is still added feature density that would increase the number of feature matches between scenes for navigation and give a better quality solution.

SfM was able to estimate the trajectory for the MWIR camera along the night trial used in this research. There were no equivalent trajectories to compare this to given the

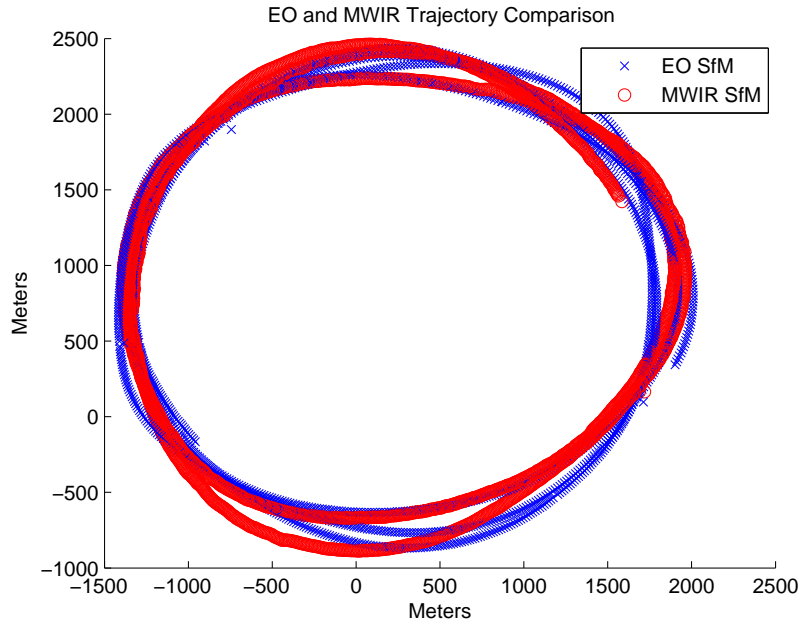


(a) SfM Trajectory Top View

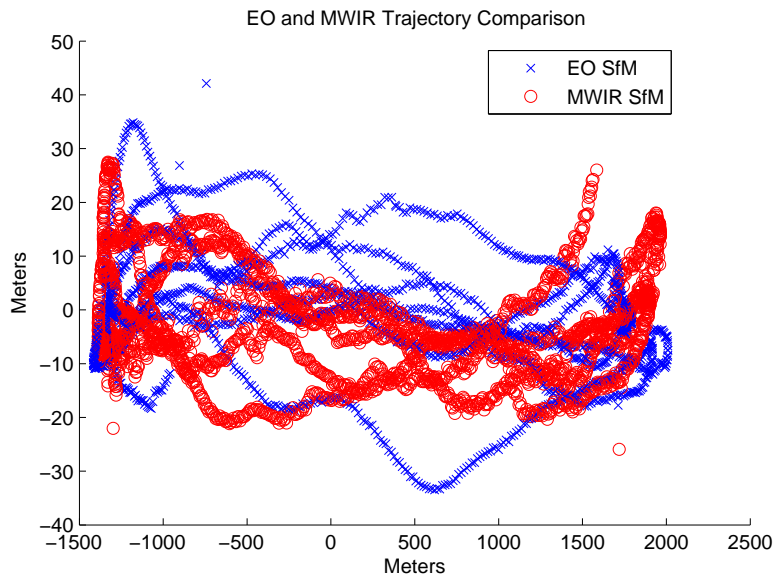


(b) SfM Trajectory Side View

Figure 4.10: Comparison of Aligned SfM Solutions MSEE 1. These images show a MWIR SfM solution aligned to an EO solution via point cloud comparison. The two tracks do not match up in time due to the lack of precision MWIR timing, but they are from the same trial and cover the same area. The top view shows that the trajectories track the same circular motion and maintain do not drift away over the 18 minutes of data used. The side view shows a similar result wherein the altitudes track within 40 meters or so of each other.



(a) SfM Trajectory Top View



(b) SfM Trajectory Side View

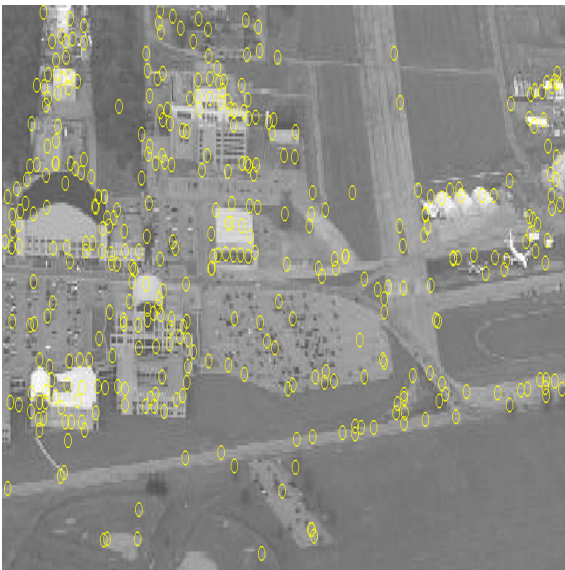
Figure 4.11: Comparison of Aligned SfM Solutions DEBU 2. Similar results are shown for the DEBU 2 trial. This alignment used 12 minutes of MWIR data. The level of tracking is similar to the MSEE 1 trial.



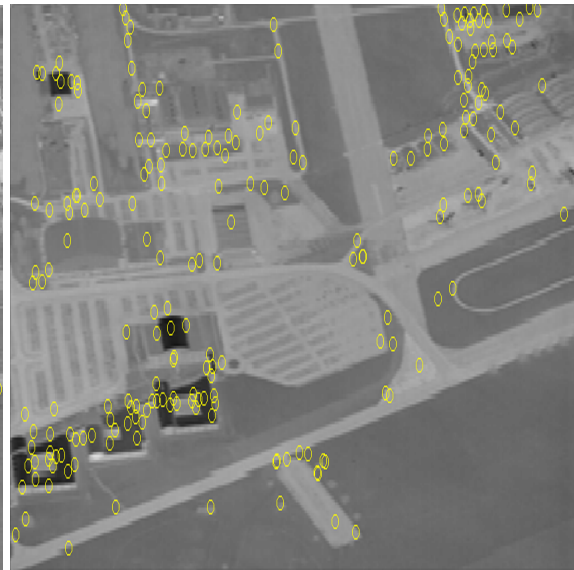
(a) Daytime Image



(b) Nighttime Image



(c) Daytime SIFT (489 Features)



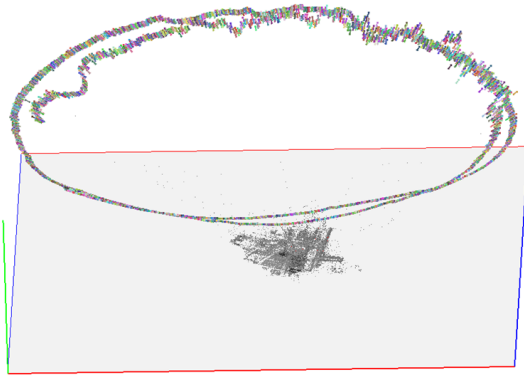
(d) Nighttime SIFT (266 Features)

Figure 4.12: Comparison of Day and Night MWIR Images. The two images are of the same location on WPAFB but from both day and night trials. The buildings with black roofs in the bottom left of the nighttime image have white roofs in the daytime image. The same goes for the planes in the upper right corner. This is due to the images being on opposite sides of thermal crossover. The night image has less feature density than the day image.

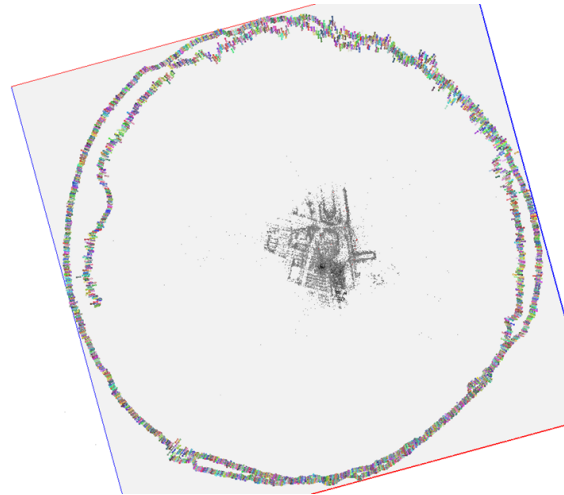
lack of EO SfM solutions at night and timing on the MWIR images for comparison to the GPS trajectory, but the circular motion captured by SfM estimates resembles that seen by the GPS solution during that flight. The reconstruction of the ground in the SfM point cloud resembles the images used to compare day and night feature detection.

4.7 MWIR Navigation Viability

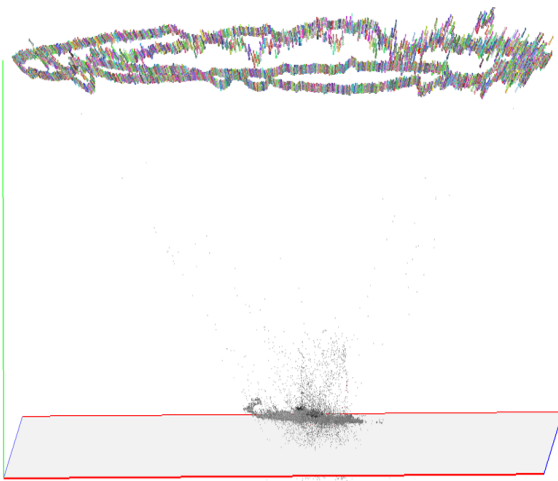
The second part of the transitive argument, that MWIR and EO SfM solutions are similar in quality, is illustrated by the comparisons in this chapter. The differences between sensing ground features were first highlighted to acknowledge that the solutions are not the same. One of the unique strengths of MWIR sensing, nighttime functionality, was explored to support the reason for looking at these sensors as additional navigation tools. The other strength, smoke penetration, was not explored as the MAMI trial did not experience this situation in flight. The side-by-side views of overall point cloud construction and derived trajectories were then shown to prove that despite the differences in creation, the end result is of similar quality. MWIR SfM is a different tool that can be used for navigation aiding in the same way that EO solutions can.



(a) Angled View



(b) Top View



(c) Side View



(d) Point Cloud View

Figure 4.13: MWIR Night Trajectory. This reconstruction of the scene and estimate of airplane trajectory was performed on MWIR images taken of a scene at night. The ability for MWIR to perform these estimates is a unique strength over the EO domain.

V. Conclusions and Future Work

This thesis explored the potential for Medium Wave Infrared (MWIR) cameras to be used in vision-aided navigation. Proving this utility expands the tools available to operators requiring precise navigation in potentially hostile environments.

5.1 Conclusions

In order to prove the validity of MWIR image navigation aiding, a transitive argument highlighting two points was presented: First, using measurements from the Structure from Motion (SfM) algorithm run on a set of images in conjunction with inertial data significantly improves the solution quality over only using an inertial sensor. Second, MWIR SfM position estimates are similar in quality to EO estimates from the same experiment. Both of these arguments were studied in this research through real-world data.

The first argument addressed the inclusion of SfM position data to an integrated navigation solution with inertial sensors. The first step in accomplishing this combined solution was to simulate an IMU from the INS data used as truth for the trial. This was accomplished by adding in artificial noise to integrated measurements from the truth solution according to the statistics of noise on a HG1700 IMU. Next, an SfM solution was created via VisualSfM to estimate relative position changes between images. The position estimates were aligned to the INS for the time leading up to the beginning of the trial in question. To turn the SfM position estimates into measurements, the positional changes between each image were divided by the time between them to give velocity measurements. These velocity measurements were combined with the IMU measurements in SPIDER to create a combined navigation solution. The error in this combined solution was compared with the unaided IMU solution error.

The resulting combined solution showed an improvement over the $2.7 * 10^4$ meters of error in the free running IMU solution by lowering it to approximately 120 meters of error at the end of a 30 minute trial. The exponentially increasing error due to the growing biases in the IMU were significantly less pronounced in the combined solution plots. Combining the SfM velocity measurements with the IMU extended the life of the usable navigation solution. This sets the framework for which MWIR measurements could be included in a combined solution.

The second argument addressed the comparison between EO and MWIR SfM solutions. Given that EO and SfM images differ in feature content, feature detection over specific areas of the scene was compared between imaging domains. Due to low contrast in the MWIR images, contrast enhancement was used to accentuate features in the scene. Detection showed lower feature density in the MWIR images over heavily vegetated scenes or those with little change in material consistency. MWIR images did detect many features along the borders between two different materials in a scene, such as the edges of roads or the corners of buildings. The strength of MWIR cameras to detect the scene at night was compared to similar images of the day trials. The night trials showed only a slightly lower feature density than the day trials, which is a large improvement over the almost negligible detection by EO cameras in the dark.

In order to compare the estimated SfM trajectories, the point clouds created by VisualSfM for the EO and MWIR images were scaled, rotated, and translated by hand into the same reference frame to allow trajectory comparison. Both SfM solutions maintained similar trajectories over time evidenced by their matching patterns. This qualitative analysis was valid for these trials as the aircraft taking images moved in a constant circular pattern over specific patches of ground in the scene. The trajectories followed a similar track despite not matching up in time due to the lack of timing in the MWIR imagery. This

similarity strengthens the proposal that MWIR imagery would give similar quality updates to a combined navigation solution if included with an IMU.

5.2 Future Work

The most direct segue from this research for future work would be a similar experiment performed with precision timing on the MWIR imagery. The framework implemented for this thesis combining SfM with the inertial sensors can be exploited to operate with MWIR imagery with little to no modification. The difference in quality over time between MWIR and EO image aided navigation would prove an interesting comparison for these combined solutions. Furthermore, MWIR imagery over other types of environments, especially heavily wooded areas, deserts, urban landscapes, and oceans, would further the study of feature detection in the scene on navigation quality done in this research.

Another potential approach might be to leverage inertial data to constrain SfM matching between frames such that the solution and the point cloud representation of the scene are simultaneously improved. This coupling could also be applied to MWIR imagery to study the effects on the point cloud and navigation solution.

Explorations of SWIR and LWIR sensors for navigation may give a more complete picture of the effects of different phenomenologies across the electro-magnetic spectrum on the quality of vision-aided navigation solutions.

Another important aspect to vision-based navigation that warrants further explanation is a study of how such approaches apply at various altitudes. In particular, high-altitude aircraft may suffer from a lack of useful detail, while lower flying aircraft may not have the opportunity to track features reliably from frame to frame. The consequences of these effects on navigation quality can be compared to be useful in mission planning.

Finally, significant though should be given to enhancing vision-based navigation approaches such that global position estimates can be made that do not drift over time.

Absolute world measurements based on matching the scene to known world points would effectively eliminate the drifting nature of this type of solution.

Bibliography

- [1] Michael J Veth. Fusion of imaging and inertial sensors for navigation. Technical report, Air Force Institute of Technology, 2006.
- [2] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):1052–1067, June 2007.
- [3] M. Smearcheck, D. Marietta, J. Raquet, D. Ruff, and A Herrera. Expandable flight reference data processing software, 2014.
- [4] David Titterton and John L Weston. *Strapdown inertial navigation technology*, volume 17. IET, 2004.
- [5] Pratap Misra and Per Enge. *Global Positioning System: Signals, Measurements and Performance Second Edition*. Lincoln, MA: Ganga-Jamuna Press, 2006.
- [6] Peter S Maybeck. *Stochastic models, estimation, and control*, volume 2. Academic press, 1982.
- [7] Robert Grover Brown, Patrick YC Hwang, et al. *Introduction to random signals and applied Kalman filtering*, volume 3. Wiley New York, 1992.
- [8] Peter S Maybeck. *Stochastic models, estimation, and control*, volume 1. Academic press, 1979.
- [9] Mike Veth, Robert C Anderson, Fred Webber, and Mike Nielsen. Tightly-coupled ins, gps, and imaging sensors for precision geolocation. Technical report, DTIC Document, 2008.
- [10] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [11] Jean-Yves Bouget. Camera calibration toolbox for matlab, dec 2013.
- [12] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [13] Frank Dellaert, Steven M Seitz, Charles E Thorpe, and Sebastian Thrun. Structure from motion without correspondence. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 557–564. IEEE, 2000.
- [14] Noah Snavely, Steven M Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006.

- [15] Noah Snavely, Steven M Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.
- [16] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M Seitz, and Richard Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.
- [17] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment: a modern synthesis. In *Vision algorithms: theory and practice*, pages 298–372. Springer, 2000.
- [18] Henri Gavin. The levenberg-marquardt method for nonlinear least squares curve-fitting problems. *Department of Civil and Environmental Engineering, Duke University*, pages 1–15, 2013.
- [19] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [20] Changchang Wu. Towards linear-time incremental structure from motion. In *3DTV-Conference, 2013 International Conference on*, pages 127–134. IEEE, 2013.
- [21] Kyle J Kauffman. Radar based navigation in unknown terrain. Technical report, DTIC Document, 2012.
- [22] Changchang Wu. Siftgpu: A gpu implementation of scale invariant feature transform (sift), 2007.
- [23] Changchang Wu, Sameer Agarwal, Brian Curless, and Steven M Seitz. Multicore bundle adjustment. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3057–3064. IEEE, 2011.
- [24] Ronald G Driggers, Melvin H Friedman, and Jonathan Nichols. *Introduction to infrared and electro-optical systems*. Artech House, 1999.
- [25] Nathaniel Mitchell. *Electronic warfare and radar systems engineering handbook*, 1992.
- [26] C Garnier, R Collorec, J Flifla, C Mouclier, and F Rousee. Infrared sensor modeling for realistic thermal image synthesis. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 6, pages 3513–3516. IEEE, 1999.
- [27] Nagaraj Nandhakumar and Jake K Aggarwal. Integrated analysis of thermal and visual images for scene interpretation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(4):469–481, 1988.

- [28] Nigel JW Morris, Shai Avidan, Wojciech Matusik, and Hanspeter Pfister. Statistics of infrared images. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–7. IEEE, 2007.
- [29] Kurt Beier and Hans Gemperlein. Simulation of infrared detection range at fog conditions for enhanced vision systems in civil aviation. *Aerospace Science and Technology*, 8(1):63–71, 2004.
- [30] Dean Scribner, Penny Warren, and Jonathan Schuler. Extending color vision methods to bands beyond the visible. In *Computer Vision Beyond the Visible Spectrum: Methods and Applications, 1999.(CVBVS'99) Proceedings. IEEE Workshop on*, pages 33–40. IEEE, 1999.
- [31] AA Alexeenko, NE Gimelshein, DA Levin, RJ Collins, R Rao, GV Candler, SF Gimelshein, JS Hong, and T Schilling. Modeling of flow and radiation in the atlas plume. *Journal of Thermophysics and heat Transfer*, 16(1):50–57, 2002.
- [32] Johannes Kopf, Matt Uyttendaele, Oliver Deussen, and Michael F Cohen. Capturing and viewing gigapixel images. In *ACM Transactions on Graphics (TOG)*, volume 26, page 93. ACM, 2007.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) 26-03-2015		2. REPORT TYPE Master's Thesis		3. DATES COVERED (From — To) Oct 2013–Mar 2015	
4. TITLE AND SUBTITLE On the Integration of Medium Wave Infrared Cameras for Vision-Based Navigation				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
				5d. PROJECT NUMBER N/A	
6. AUTHOR(S) Tharp, Justin S., Second Lieutenant, USAF				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
				7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB, OH 45433-7765	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Richard Van Hook AFRL Sensors Directorate 2241 Avionics Circle Wright-Patterson AFB, OH 45433 312-798-8589 richard.van_hook.1@us.af.mil				8. PERFORMING ORGANIZATION REPORT NUMBER AFIT-ENG-MS-15-M-063	
				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RY	
12. DISTRIBUTION / AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A: APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
				13. SUPPLEMENTARY NOTES This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.	
14. ABSTRACT The ubiquitous nature of GPS has fostered its widespread integration of navigation into a variety of applications, both civilian and military. One alternative to ensure continued flight operations in GPS-denied environments is vision-aided navigation, an approach that combines visual cues from a camera with an inertial measurement unit (IMU) to estimate the navigation states of a moving body. The majority of vision-based navigation research has been conducted in the electro-optical (EO) spectrum, which experiences limited operation in certain environments. The aim of this work is to explore how such approaches extend to infrared imaging sensors. In particular, it examines the ability of medium-wave infrared (MWIR) imagery, which is capable of operating at night and with increased vision through smoke, to expand the breadth of operations that can be supported by vision-aided navigation. The experiments presented here are based on the Minor Area Motion Imagery (MAMI) dataset that recorded GPS data, inertial measurements, EO imagery, and MWIR imagery captured during flights over Wright-Patterson Air Force Base. The approach applied here combines inertial measurements with EO position estimates from the structure from motion (SfM) algorithm. Although precision timing was not available for the MWIR imagery, the EO-based results of the scene demonstrate that trajectory estimates from SfM offer a significant increase in navigation accuracy when combined with inertial data over using an IMU alone. Results also demonstrated that MWIR-based positions solutions provide a similar trajectory reconstruction to EO-based solutions for the same scenes. While the MWIR imagery and the IMU could not be combined directly, through comparison to the combined solution using EO data the conclusion here is that MWIR imagery (with its unique phenomenologies) is capable of expanding the operating envelope of vision-aided navigation.					
15. SUBJECT TERMS Infrared Vision Navigation					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Maj Brian G Woolley (ENG)
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			UU