IMPROVING AUTOMATED AERIAL REFUELING STEREO VISION POSE ESTIMATION USING A SHELLED REFERENCE MODEL

THESIS

Christopher A. Parsons, 2nd LT, USAF
AFIT-ENG-MS-17-M-057

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 document are those of the author and do not reflect the official policy or position of the United States Air Force, the United States 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.
IMPROVING AUTOMATED AERIAL REFUELING STEREO VISION POSE 
ESTIMATION USING A SHELLED REFERENCE MODEL

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 Cyber Operations

Christopher A. Parsons, B.S.C.S
2nd LT, USAF

March 2017

DISTRIBUTION STATEMENT A
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.
Abstract

Automated Aerial Refueling (AAR) of Unmanned Aerial Vehicles (UAVs) is vital to the United States Air Force’s (USAF) continued air superiority. Inspired by the stereo vision system organic to the new KC-46A Pegasus tanker, this research presents a novel solution for computing a relative 6 degree-of-freedom pose between the refueling aircraft and a tanker. The approach relies on a real time 3D virtual simulation environment that models a realistic refueling scenario. Within this virtual setting, a stereo camera system mounted beneath the tanker consumes synthetic imagery of a receiver conducting an aerial refueling approach. This synthetic imagery is processed by computer vision algorithms that calculate the sensed relative-navigation position and orientation. The sensed solution is compared against the virtual environment’s truth data to quantify error and evaluate the stereo vision performance in a deterministic, real time manner. Pose estimation accuracy and computational speed during registration improve though the use of a shelled reference model. The shelled model improves computational speed of pose estimation at the refueling position by 87% and accuracy by 36% when compared with a full reference model. To ensure proper simulation of computer vision concepts, this research quantifies the effect Multi-Sample Anti Aliasing implemented in the virtual stereo cameras on camera calibration and pose estimation. A combined shelled model and Multi-Sample Anti Aliased approach leads to position estimation errors less then 7cm and orientation estimation error less then 1°.
Acknowledgements

Special thanks to our sponsor, AFRL/RQ for their support in this research endeavor.

I would like to thank my family, friends, and mentors for supporting me throughout my military career, especially my parents and grandparents!

Christopher A. Parsons
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Overview/Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Problem Statement</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Research Goals and Hypothesis</td>
<td>2</td>
</tr>
<tr>
<td>1.4 Approach</td>
<td>2</td>
</tr>
<tr>
<td>1.5 Assumptions/Limitations</td>
<td>3</td>
</tr>
<tr>
<td>1.6 Research Contributions</td>
<td>3</td>
</tr>
<tr>
<td>1.7 Thesis Overview</td>
<td>4</td>
</tr>
<tr>
<td>II. Background</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Automated Aerial Refueling</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Computer Vision</td>
<td>6</td>
</tr>
<tr>
<td>2.3 Pinhole Camera Model</td>
<td>7</td>
</tr>
<tr>
<td>2.4 Stereo Computer Vision</td>
<td>7</td>
</tr>
<tr>
<td>2.5 Camera Calibration</td>
<td>9</td>
</tr>
<tr>
<td>2.6 Registration</td>
<td>10</td>
</tr>
<tr>
<td>2.7 Multi-Sample Anti Aliasing</td>
<td>11</td>
</tr>
<tr>
<td>2.8 Previous and Related Work</td>
<td>12</td>
</tr>
<tr>
<td>2.8.1 Non Vision Approaches</td>
<td>12</td>
</tr>
<tr>
<td>2.8.2 Vision Approaches</td>
<td>14</td>
</tr>
<tr>
<td>III. Methodology</td>
<td>17</td>
</tr>
<tr>
<td>3.1 Virtual World</td>
<td>17</td>
</tr>
<tr>
<td>3.2 Environmental Parameters</td>
<td>18</td>
</tr>
<tr>
<td>3.3 Virtual World Axis</td>
<td>19</td>
</tr>
<tr>
<td>3.3.1 CONOPS Flight Path</td>
<td>19</td>
</tr>
<tr>
<td>3.4 Disparity Map Generation</td>
<td>21</td>
</tr>
<tr>
<td>3.5 Re-projection and Point Cloud Generation</td>
<td>22</td>
</tr>
<tr>
<td>3.6 Model Registration and 6DOF Estimation</td>
<td>23</td>
</tr>
<tr>
<td>3.7 Experimental Design</td>
<td>24</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Probe-and-drogue (Left) and USAF Boom (Right) Refueling Methods [1, 2]</td>
</tr>
<tr>
<td>2.</td>
<td>General Camera Model [3]</td>
</tr>
<tr>
<td>4.</td>
<td>Example of Stair Stepping Effect</td>
</tr>
<tr>
<td>5.</td>
<td>Example of MSAAx4 for a Triangle [5]</td>
</tr>
<tr>
<td>6.</td>
<td>High level outline of 3DVW experimental process</td>
</tr>
<tr>
<td>7.</td>
<td>Simulated Tanker and C12 receiver</td>
</tr>
<tr>
<td>8.</td>
<td>Virtual World Axis</td>
</tr>
<tr>
<td>9.</td>
<td>Full (Left) and Shelled (Right) Reference Model</td>
</tr>
<tr>
<td>10.</td>
<td>Virtual World Stereo Calibration</td>
</tr>
<tr>
<td>11.</td>
<td>Filtered Receiver Disparity Map</td>
</tr>
<tr>
<td>12.</td>
<td>Receiver Point Cloud</td>
</tr>
<tr>
<td>13.</td>
<td>3DVW Visualizations ICP Registration</td>
</tr>
<tr>
<td>14.</td>
<td>ICP Characterization Positions in the 3DVW</td>
</tr>
<tr>
<td>15.</td>
<td>Position estimation error for ICP characterization of the full reference model at 30 iterations</td>
</tr>
<tr>
<td>16.</td>
<td>Position estimation error for ICP characterization of the shelled reference model at 30 iterations</td>
</tr>
<tr>
<td>17.</td>
<td>Orientation estimation error for ICP characterization of the full reference model at 30 iterations</td>
</tr>
<tr>
<td>18.</td>
<td>Orientation estimation error for ICP characterization of the shelled reference model at 30 iterations</td>
</tr>
<tr>
<td>19.</td>
<td>ICP Execution Times for ICP Shelled and Full Reference Models</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>20.</td>
<td>Position estimation error for the shelled reference model during a CONOPS flight path</td>
</tr>
<tr>
<td>21.</td>
<td>Orientation estimation error for the shelled reference model during a CONOPS flight path</td>
</tr>
<tr>
<td>22.</td>
<td>Position estimation error for the full reference model during a CONOPS flight path</td>
</tr>
<tr>
<td>23.</td>
<td>Orientation estimation error for the full reference model during a CONOPS flight path</td>
</tr>
<tr>
<td>24.</td>
<td>6DOF Receiver Estimation Error from Tail Exceeding Stereo Camera Viewing area at 38 to 42 meters away from the Cameras</td>
</tr>
<tr>
<td>25.</td>
<td>Virtual Checkerboard With MSAA Disabled (Left) and Enabled (Right)</td>
</tr>
<tr>
<td>26.</td>
<td>Position estimation error for AAR CONOPS Flight path with MSAA disabled</td>
</tr>
<tr>
<td>27.</td>
<td>Orientation estimation error for AAR CONOPS flight path with MSAA disabled</td>
</tr>
<tr>
<td>28.</td>
<td>Position estimation error for AAR CONOPS flight path using MSAA calibration only</td>
</tr>
<tr>
<td>29.</td>
<td>ICP orientation estimation error for AAR CONOPS flight path using MSAA calibration only</td>
</tr>
<tr>
<td>30.</td>
<td>Y Component Error For CONOPS Flight Path with MSAA Disabled at Refueling Envelope</td>
</tr>
<tr>
<td>31.</td>
<td>ICP position estimation error for AAR CONOPS flight path with MSAA enabled</td>
</tr>
<tr>
<td>32.</td>
<td>Orientation estimation error for AAR CONOPS flight path with MSAA enabled</td>
</tr>
<tr>
<td>33.</td>
<td>Standard Deviation of 10 Estimations of Position For Receiver Poses Throughout the AAR CONOPS Flight Path</td>
</tr>
<tr>
<td>Figure</td>
<td>Standard Deviation of 10 Estimations of Orientation For Receiver Poses Throughout the AAR CONOPS Flight Path</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>34.</td>
<td></td>
</tr>
</tbody>
</table>
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Comparison of Average 6DOF Estimation Error For ICP Characterization using Shelled and Full Reference Models</td>
<td>35</td>
</tr>
<tr>
<td>2.</td>
<td>RMS Error of 6DOF Estimation For ICP Characterization using Shelled and Full Reference Models</td>
<td>36</td>
</tr>
<tr>
<td>3.</td>
<td>Average and Standard Deviation for ICP Execution Times using Shelled and Full Reference Models</td>
<td>36</td>
</tr>
<tr>
<td>4.</td>
<td>Comparison of Average 6DOF Estimation Error For CONOPS Flight Path using Shelled and Full Model</td>
<td>41</td>
</tr>
<tr>
<td>5.</td>
<td>RMS Error of 6DOF Estimation For CONOPS Flight Path using Shelled and Full Reference Model</td>
<td>42</td>
</tr>
<tr>
<td>6.</td>
<td>Effect of MSAA on RMS Re-projection and Epipolar Constraint Calibration Error</td>
<td>43</td>
</tr>
<tr>
<td>7.</td>
<td>Elementwise Percent Error from Ideal Re-projection Matrix $Q$</td>
<td>44</td>
</tr>
<tr>
<td>8.</td>
<td>Comparison of Average 6DOF Estimation Error For CONOPS Flight Path With and Without MSAA Calibration</td>
<td>46</td>
</tr>
<tr>
<td>9.</td>
<td>Comparison of Average 6DOF Estimation Error For CONOPS Flight Path between MSAA Calibration Only and MSAA Enabled</td>
<td>50</td>
</tr>
<tr>
<td>10.</td>
<td>Comparison of Average 6DOF Estimation Error For CONOPS Flight Path With and Without MSAA Enabled</td>
<td>56</td>
</tr>
<tr>
<td>11.</td>
<td>RMS Error of 6DOF Estimation For CONOPS Flight Path comparing MSAA Enabled, MSAA Disabled, and MSAA Calibration Only</td>
<td>56</td>
</tr>
</tbody>
</table>
I. Introduction

1.1 Overview/Background

Aerial refueling provides the United States Air Force with aircraft endurance to maintain air superiority and perform humanitarian operations around the world. With the emergence of Unmanned Aerial Vehicles (UAVs), air operations can be done more efficiently and ground operations gain increased support as a result of UAV readiness and endurance. UAVs currently lack the capability to perform mid-flight aerial refueling due to latency in their command and control structure. Although GPS can provide enough accuracy to help aid a UAV in a refueling approach, reliance on a deniable asset can limit the operational capability of automated aerial refueling (AAR). An automated system using a stereo computer vision pipeline can use the available sensors on the tanker to provide 6 degree of freedom (6DOF) measurements of the aircraft receiving fuel (receiver), relative to the tanker. These measurements can be computed more efficiently and provide a higher level of accuracy when using a specialized reference model in the pipeline. Testing these systems traditionally requires multiple flight tests to ensure system accuracy and reliability however by using a virtual world, testing can be performed more rapidly and small errors can be rectified prior to a live flight test. Using a virtual world to simulate and execute AAR algorithms and methods has the potential to lower research costs and produce more robust computer vision solutions.
1.2 Problem Statement

Improve computer vision based relative navigation systems to ensure safe and effective automated refueling operations in GPS denied environments with minimal aircraft modification.

1.3 Research Goals and Hypothesis

- Improve computer vision relative 6DOF estimations according to computational speed and accuracy through the use of a shelled reference model
- Modify a 3D virtual world (3DVW) for simulating, visualizing, and testing AAR computer vision solutions and scenarios
- Provide virtual world results to compare with flight test data when available

1.4 Approach

Experiments to quantify shelled reference model performance are executed in the 3DVW. Synthetic image acquisition and a computer vision pipeline are added to the virtual world to simulate tanker sensors and visualize elements of the computer vision process. A flight path executed in the 3DVW provides receiver truth data for 6DOF error as well as positions to capture realistic virtual imagery. Two sets of deterministic experiments executed in the virtual world provide data for comparing shelled and full reference model performance. A final deterministic experiment provides data representing the effect of Multi-Sample Anti-Aliasing (MSAA) on computer vision simulation in the 3DVW.
1.5 Assumptions/Limitations

The 3DVW presented in this research has yet to be compared with real world results and has been constrained in order to scope the AAR problem. This research does not take into account partial occlusion of the receiver resulting from the refueling boom of the tanker. In addition, the propellers of the C-12 receiver are not present to simplify modeling. The flight path used to execute tests in the 3DVW was generated using BlueMax flight simulation software; however, the virtual world does not internally use a flight dynamics engine to simulate turbulence and other aerodynamic effects on the receiver. Virtual sensors in the 3DVW follow the pinhole camera model such that image distortion was not present and did not need to be removed in the stereo computer vision pipeline.

1.6 Research Contributions

- 3DVW for simulating, visualizing, and testing AAR computer vision applications
- 50% reduction in virtual camera calibration error using MSAAx16 on virtual imagery
- 87% improvement in speed and 35% improvement in accuracy of 6DOF measurements when using a shelled model in the stereo computer vision pipeline
- 6DOF relative position estimations within 7cm in the X and Z components and 2cm in the Y component
1.7 Thesis Overview

Chapter 2 provides a background into the AAR terminology used throughout this research. Foundational concepts for computer vision as well as aspects of the 3DVW are discussed in the background with a review of previous and related work. Chapter 3 provides the methodology for the stereo computer vision pipeline presented in this research along with the tests performed to quantify the performance of the shelled and full reference model as well as the effect of MSAA on stereo computer vision in the 3DVW. Results from the experimental tests outlined in Chapter 3 are presented in Chapter 4 with corresponding analysis. The final chapter provides conclusions based on the results of this research. Future work to improve the results of this AAR research effort will also be included in the final chapter. An included appendix provides additional experimental information in support of the produced results as well as clarification for deterministic assumptions made in this research.
II. Background

2.1 Automated Aerial Refueling

To achieve mission objectives for military and humanitarian operations, refueling capabilities must be available to ensure endurance and capability. While the capability to refuel aircraft has existed for over 50 years, unmanned aerial vehicles (UAVs) are a relatively new aviation platform and cannot refuel with the current tanker fleet. As the dependence on UAVs grows to fulfill requirements for additional combat support, reconnaissance, and surveillance, refueling capabilities must be developed to ensure maximum availability. Command and control latency of UAVs prevents pilots from maneuvering with enough real-time precision to perform a traditional refueling operation. An automated method to estimate the relative position and orientation of the UAV fills the command and control gap and allows a safe approach with minimal latency assuming an available short-range communication link with enough bandwidth. Given a stereo vision system on the tanker aimed at an approaching receiver, AAR can be achieved through a combined approach using computer vision and onboard aircraft inertial measurement units (IMUs). Once AAR can be applied to a refueling platform the need for a boom operator could eventually be eliminated, paving the way towards autonomous tankers.

AAR defines the receiver as the aircraft being refueled by the tanker. Performing AAR with a receiver requires a specific type of flight path to be used to ensure safe operation. The flight path will vary based on aircraft type and environmental conditions and these considerations are defined in a refueling Concept of Operations (CONOPS). The CONOPS flight path for AAR provides guidelines describing an approach that limits occlusion of receiver aircraft structures whenever possible [6]. By maximizing receiver visibility, the computer vision solution can produce more
accurate measurements for position and orientation. The measurements produced from the solution provide relative pose estimation in a six degree-of-freedom (6DOF) format with three values for position and the remaining three for orientation.

There are two primary methods of refueling: probe-and-drogue and USAF boom. Probe-and-drogue involves the use of an inactive drogue that extends behind the tanker and accepts a probe attached to the receiver. During probe-and-drogue refueling the tanker has minimal control over the refueling process except for the overall position of the drogue. In order to gain better control and increase the speed of fuel transfer, the USAF boom method was developed by Boeing [7]. Using a boom allows an operator to control the position of the boom and precisely connect with the refueling receptacle on the receiver. Figure 1 shows both the tanker and receiver in a probe-and-drogue refueling operation (left) and a boom operation (right).

Figure 1. Probe-and-drogue (Left) and USAF Boom (Right) Refueling Methods [1, 2]

2.2 Computer Vision

Computer vision uses images of the world to provide computer systems with data that describes the geometry and attributes of the real world [8]. As humans we have the capability to quickly process imagery and make comparisons. While research into artificial intelligence seeks to provide systems with human-like perception [9] computer systems must use features, colors, as well as other forms of pattern recognition either
together or individually to produce the desired output. Computer vision provides capabilities to recognize, track, match and estimate the range of an object given certain parameters [10] making it a tool for automatic systems that interact with a rapidly changing physical environment.

2.3 Pinhole Camera Model

Cameras are one of the many sensors that can provide raw data to computer vision algorithms. Although cameras do not contain direct range information, performing a calibration and using multiple cameras allows for range and position to be estimated. Figure 2 presents a pinhole camera model as well as the relationship between an object point $P$ and an image point $(u, v)$. The optical center serves as the origin of the camera frame, $F_c$. Image points of a scene lie on the image plane parameterized by $x$ and $y$ extending a certain length from the optical center. The Z or depth axis of the camera frame intersects the image plane at the principal point, $(c_x, c_y)$ [3]. Measuring the distance from the optical center to the principal point provides the focal length of a camera. The color value of the image point $(u, v)$ becomes the color of the first object point encountered along the ray extending from the optical center through $(u, v)$. Many of the computer vision algorithms that use imagery assume a pinhole camera model. Parameters such as focal length, non-center principal point, and unknown optical center can influence the accuracy position estimation and must be carefully measured in real world environments.

2.4 Stereo Computer Vision

Stereo vision uses the geometry of two cameras, specified as epipolar geometry, to compute a depth of objects that lie in both cameras’ field of view [11]. Stereo vision algorithms can mimic how humans estimate depth with their eyes. Figure 3
Figure 2. General Camera Model [3]

displays the epipolar geometry of a general stereo vision system. $O_1$ and $O_2$ represent the optical centers of the left and right cameras and baseline between them intersects each image plane at epipoles $e_1$ and $e_2$. The epipolar plane contains the object point $q$ in 3D space as well as the projections of that point, $p_1$ and $p_2$ on the left and right image planes, respectively. Lines $l_1$ and $l_2$ represent the epipolar lines. An essential matrix condenses the epipolar geometry between the stereo vision system providing a relationship between image points in one image with epipolar lines in the corresponding image. In doing so the essential matrix also provides a direct relationship between the image points $p_1$ and $p_2$ of an object point $q$ known as the epipolar constraint can be represented as the function $p_1^\top E p_2 = 0$ [11]. A fundamental matrix provides a similar encoding of the epipolar geometry without a camera calibration.

Knowing the epipolar geometry allows the 3D coordinates of the object point to
be determined with respect to the optical center of the primary camera. The optical center of the primary camera often serves as the local origin of the stereo vision coordinate frame simplifying transformations into alternative coordinate frames.

2.5 Camera Calibration

In order to produce accurate 3D positions, each camera’s calibration info must be computed. Camera calibration provides important intrinsic and extrinsic parameters of a camera following the pinhole model. Intrinsic information defines the focal length and location of the principal point where as extrinsic information describes each cameras’ orientation. After completing a calibration, re-projection matrices derived from intrinsic and extrinsic information allow each image point to be projected into 3D space. With stereo camera calibration, the re-projection matrix can be combined such that a complete re-projection occurs in the coordinate frame of the primary
camera. In addition, the essential matrix can also be computed by using features obtained through the stereo calibration process.

Camera calibration, as described in [12], performs a camera calibration using a flat object with an affixed pattern. Most applications of Zhang's method use a calibration checkerboard consisting of black and white squares as corners produce well defined features for detection. The method uses the global Levenberg-Marquardt optimization algorithm to minimize the overall re-projection error for each feature [3]. Re-projection error can be computed for each feature image point as the corresponding object points of the checkerboard pattern have known distances from one-another.

2.6 Registration

Registration in computer vision refers to the matching of 2D and 3D point clouds of an object or scene [10]. A sensed model provides the points resulting from computer vision data which often contain noise resulting from unideal conditions in real world scenarios including camera imperfections and errors in camera calibration. A reference model provides points representing the true point cloud of the object or scene being viewed by the computer vision system. By registering a sensed model of the object to a corresponding reference model an estimated pose of the sensed model can be determined.

Points can be registered by using either a rigid or non-rigid registration. A rigid registration does not allow for the model's vertices to be deformed before being matched. Conversely, non-rigid registration allows for points in the model to be warped, providing increased registration flexibility.

The iterative closest point (ICP) algorithm presented in [13] provides the foundational method for most rigid 3D registration algorithms. For each point in the sensed model the algorithm searches for a corresponding closest point in the reference model.
The algorithm then computes the rigid translation and rotation between the corresponding points. Every point in the sensed point cloud receives the translation and rotation before the process restarts using the updated position and orientation of the sensed point cloud. This process of minimization continues until a certain threshold of mean standard error is achieved.

ICP convergence towards a mean squared error of 0 does not always insure correct registration. Local minimums for each closest point used in the initial iterations of ICP can prevent the algorithm from finding better global minimums that would have been found if the model was oriented differently. Although new methods such as those provided in [14] present computational solutions, a good initial guess for position and orientation of the sensed point cloud helps improve the accuracy of ICP. As ICP works iteratively, it can be computationally expensive for large point clouds and have limited capability in real time systems. Improvements presented in [15] provide a modified approach implementing a K-D tree data structure to keep track of points for quicker comparison in ICP. Further variations of ICP presented in [16, 17, 18, 19, 20] provide specific modifications providing optimizations in speed and accuracy.

2.7 Multi-Sample Anti Aliasing

Objects in the 3DVW experience aliasing of edges due to pixel quantization during rasterization. Multi-Sample Anti Aliasing (MSAA) implemented in the virtual sensors attempts to reduce the stair-stepping effect of edges and improve the performance of computer vision systems in the virtual world by making imagery more life-like [21]. Figure 4 provides a visual example of a edge of an object rendered without MSAA.

MSAA works by sub-sampling pixels intersected by edges of a shape. The number of sub-samples that lie within the edges of the shape provide a level of shading to apply to the pixel’s color. Figure 5 displays the differences in shading a triangle using
MSAA with 4 samples per pixels where blue dots show the samples per pixel and red lines present edges of the triangle. Notice that the shading depends on the number of subsamples that lie within the edges of the shape for each pixel.

2.8 Previous and Related Work

Non Vision Approaches.

Differential GPS in [22] was shown in a live flight test to successfully refuel a receiver using the probe and drogue refueling method. Related work by [23] used differential GPS between two aircraft, one acting as the tanker and one as the receiver to perform a CONOPS refueling operation. Although work with GPS shows sufficient position estimation performance for refueling, using it as the only system for position estimation makes AAR operations reliant on GPS availability.

Light Detection and Ranging (LIDAR) provides an alternative method for esti-
mating receiver pose by using lasers to create a point cloud. Work by [24] estimated the relative receiver position by using a scanning LIDAR mounted on the receiver to compute distance from the tanker. Although the method was shown to be viable it uses an active sensor that could help adversaries target the aircraft. In addition, modifications to each receiver aircraft are cost prohibitive and could affect the aero-dynamic capability or stealth features of the aircraft.

To ensure safe refueling operations and improve the overall accuracy of receiver
position estimation, research has primarily focused on combined approaches. These approaches fuse the capabilities of position data from GPS, acceleration and orientation data from inertial measurement units (IMU), and pose estimation data from computer vision solutions to produce a single estimate. Fusion of the aforementioned systems usually involves an Extended Kalman Filter (EKF) in order to combine the results. Work by [25] presents a combined GPS and IMU approach with supporting flight test data. This system provides more flexible operational capability; however losing GPS would still limit the system’s capability to estimate position given drift in IMU measurements. Research presented in [26, 27, 28] uses the EKF to filter GPS information with computer vision in order to improve the overall pose estimations. The computer vision systems combined with the EKF use feature or point matching to determine the relative position of the receiver requiring markers to be painted or affixed to the aircraft. With the exception of [28], these works also use monocular vision with a camera attached to the receiver which is currently unavailable on many operational aircraft. More recent work by [29] uses an EKF that includes GPS, IMU, and computer vision measurements to create a robust tracking. Limited availability of receiver mounted cameras also limits the applicability of the work for military AAR applications.

**Vision Approaches.**

Previous work in computer vision receiver pose estimation focused on using point matching or feature detection. Research by [30] quantifies the difference between feature matching using corner and markers with monocular vision. Work presented in [31] and [32] further expand upon feature matching methods using monocular vision and evaluate them according to their effectiveness in the AAR domain. Other work by [33] uses sensors on the receiver to detect particular wavelengths of light emitted from
Recent work by [34] uses binocular vision to compute receiver pose through the extraction of speeded up robust features (SURF) from a specialized marker added to the aircraft. Comparison of the SURF features with the original marker enables the computation of a receiver pose. Another implementation of receiver pose estimation using binocular vision [35] uses saliency maps to produce easily identifiable features of a specific marker. Once features have been extracted from the saliency maps, a receiver pose can be computed similarly to the method in [34]. All of the previously described methods present solutions to the AAR problem; however, all require modification to the receiver in order to function. To minimize cost and time, additional computer vision solutions should be explored.

The stereo computer vision solution presented in this research extends previous work at AFIT by combining previous approaches [36, 37] to produce more accurate 6DOF measurements. Each approach used a similar stereo computer vision layout with cameras attached to the tanker looking down at the receiver. By doing so, the capabilities already present on soon-to-be operational tankers are leveraged such that external modification to the receiver and tanker are unnecessary. Work done by [37] provided an initial simulation and implementation of the stereo computer vision process. The results were not performed in real time and used lower fidelity virtual models to compare with physical scale models. Werner’s work provided a proof of concept that would be further modified by Denby [36] to optimize and bring the stereo computer vision process of pose estimation closer to real time. While real time rates were achieved, the accuracy of the solution was insufficient to ensure safe refueling operations. This research uses the foundation of AAR work at AFIT to develop a high quality virtual simulation as well as perform a virtual flight test with precise parameters for receiver approach and stereo camera setup.

Virtual environments have been used to test and evaluate AAR pose estimation
methods. MATLAB Simulink and Virtual Reality Toolbox used in previous work provided a basic simulation environment to visualize relative aircraft pose [34] and capture virtual imagery [38, 30, 31, 29]. However, the results of Simulink do not provide high fidelity models or virtual imagery that accurately reflects the real world. Simulation in [34] uses cameras facing computer monitors which display virtual imagery and do not account for issues of resolution, lighting and color quality. Work by [37] used the Ogre 3D graphics engine to simulate virtual imagery and calibrate the cameras using a virtual checkerboard. 6DOF estimates from Werners work showed the effectiveness of virtual imagery but did not mirror the hardware and flight path specifications outlined for UAV AAR. General virtual environments have been proposed for AAR [39, 40] to simulate the refueling process; however, these methods do not allow for computer vision simulation and do not allow computer vision computations to be computed in real-time or alongside visualization.
III. Methodology

To properly compare the shelled and full reference models in the AAR environment, relevant synthetic imagery needs to be generated according to AAR operations. Using a 3DVW provides the flexibility and graphical capability to create scenarios that produce the high fidelity synthetic imagery that is required. Open Source Computer Vision (OpenCV) libraries [3] include verified and tested algorithms to calibrate the virtual sensors and perform requisite stereo vision computations for computing a sensed receiver model. The registration of a sensed model onto a reference model produces the spacial pose data required for AAR. Throughout the experimental process, data can be visualized in the 3DVW to verify and support results. Figure 6 outlines the initial computations and computer vision pipeline executed in the virtual environment. Initial computations are performed before an experiment and used throughout the computer vision pipeline. The computer vision pipeline produces a 6DOF estimation for a receiver following a certain refueling flight path. The pipeline operates in a loop until a 6DOF measurement has been performed for every chosen receiver position/orientation. The results of the initial computations and computer vision pipeline, including associated timing metrics, provide the means by which to measure performance improvements between reference models and 3DVW features.

3.1 Virtual World

The experimental simulation and visualization for this research occurs in the OpenGL based AFTR Burner Engine. Successor to the STEAMiE engine, AFTR Burner uses high fidelity models and textures to create quality simulations representative of the real world [41]. The engine has been shown to interface with real world sensors to compute and visualize output[42]. This work further expands the capabil-
初等计算

(校准)

计算机视觉流程

多视图参数

重投影矩阵

更新接收器姿态

渲染虚拟场景

捕获立体影像

生成视差图

生成点云

ICP注册

输出姿态估计

图6. 3DVW实验过程的高阶概述

3DVW的能力通过展示合成传感器的使用来展示。具有与物理或虚拟传感器接口的能力，3DVW允许当前和未来的研究从实验室测试移动到2017年的实时飞行测试，而无需付出努力。

环境参数。

加油场景在虚拟环境中进行的加油场景由虚拟传感器捕获，这些传感器镜像传感器的规格，包括附件位置和角度，位于油罐机。如图7所示，左侧的大型飞机作为油罐机，右侧的较小飞机（C12）作为C12接收器。虽然C12没有自主功能，但它的使用作为接收器将允许比较使用AAR CONOPS在2017年的计划飞行测试的结果。油罐机的天线和C12接收器的螺旋桨被移除，以便将工作重点放在计算机视觉和注册问题上。工作的
with the boom and propellers in continued research will bring the AAR project closer to application.

**Virtual World Axis.**

The virtual world axes represent the $x$, $y$ and $z$ components of position used in this research. The relative coordinate frame of the virtual world can be seen in Figure 8. The virtual world axes are displayed with the receiver in the same position and orientation as displayed in Figure 7. The red line running from tail to nose represents the positive $x$ component of position. The green line running from the nose towards the left wing of the aircraft represents the positive $y$ component of position. Lastly, the blue line running from the nose perpendicular to the $xy$ plane represents the positive $z$ component of position. Orientation components follow the right-hand rule for the coordinate frame presented in Figure 8.

**CONOPS Flight Path.**

BlueMax flight simulation software provides the position and orientation truth data for the receiver pose in the virtual world [43]. Having the imported flight profile allows the virtual world to calculate the precise difference between the sensed data and truth data for each receiver 6DOF measurement. Error between position and orientation components of the receiver and registered model assume a consistent center point that remains fixed throughout experimentation. Results from the computer vision pipeline in the virtual world are not fully deterministic due to the re-projected point cloud sub-sampling and non-deterministic disparity map generation. However, this variance is negligible on disparity map generation and 6DOF estimation, as explained in Appendix A. The initial computations performed for the computer vision pipeline are deterministic and the same camera calibration parameters are used
Figure 7. Simulated Tanker and C12 receiver

Figure 8. Virtual World Axis
throughout this research.

### 3.2 Registration Models

Our research assumes a known reference model for the receiver aircraft is available. The full and shelled reference model are derivatives of the C12 receiver shown in Figure 7 and Figure 8. The reference model was modified in Blender 3D [44] to create an evenly distributed point set across its exterior surface area as shown in Figure 9. The new point cloud subsequently serves as the full reference model used throughout this research. To produce the shelled model, the viewing angle of the point cloud in Blender was adjusted such that it matched the perspective the tanker's stereo cameras would see as they view an approaching receiver in the refueling envelope. Only the visible surface points of the full model were selected and used to created the shelled reference model. Figure 9 shows different perspectives of the shelled and full reference models used for ICP registration. The full model can be seen on the left with a total of 48,686 points with the shelled reference model with 7772 points on the right.

![Figure 9. Full (Left) and Shelled (Right) Reference Model](image-url)
The models are used in the ICP registration step of the computer vision pipeline to get a position and orientation of the produced point cloud.

3.3 Virtual Sensor Use and Calibration

Two 1024x768 synthetic cameras are added to the tanker replicating a computer vision relative navigation system. Following every receiver pose update, the 3DVW renders a stereo image pair according to the camera orientation and field of view parameters. The virtual world reformats the synthetic imagery into an OpenCV Mat container simplifying the use of OpenCV and providing access to image data.

In order to use the virtual sensors for stereo computer vision, a stereo calibration must be performed. Implementing virtual calibration functionality in the 3DVW aids in examining results between the virtual world and the real world. The virtual calibration board consists of a black and white checkerboard with 16 internal corners on the horizontal axis, 17 internal corners on vertical axis and a 0.5m square size. A white border surrounds the entire board to ensure proper corner detection and stereo calibration. As shown in Figure 10, a set of 10 virtual calibration board poses within the field of view of the virtual sensors allows synthetic stereo calibration pairs to be generated and used as input for the stereo calibration function. The left image of Figure 10 displays the position of the virtual checkerboard behind the tanker and within the viewing area of the stereo cameras. An example of checkerboard corner detection can be seen in the right image of Figure 10. The corners detected by the algorithm can then be used as input for the calibration functions of OpenCV.

Stereo calibration produces a matrix $Q$ which provides the required primary camera information to perform a perspective transformation later in the computer vision pipeline. The OpenCV camera calibration function used in the virtual world also computes the camera matrices for each stereo camera according to the calibration
method proposed in [12]. To categorize the accuracy of the virtual calibration our research uses the standard RMS re-projection error outputted by the OpenCV function in addition to the average epipolar constraint error computed using the essential matrix. Given the computed essential matrix and left and right image points \( p_1, p_2 \) of all detected calibration corners in all 10 image pairs, average epipolar error can be computed between the expected value of 0 and the computed value according to \( p_1^T Ep_2 \).

3.4 Disparity Map Generation

Virtual stereo imagery acts as input to the disparity generation section of the computer vision pipeline. OpenCV provides a block matching function which we use in the virtual world to produce a disparity map.

The disparity map enables the creation of a three dimensional point cloud from stereo imagery by capturing the relative change of an object point between the left and right stereo image. Take, for example, a pair of stereo images \( I_a, I_b \) and a point \( P \) on a physical object \( O \). The difference between image coordinates in \( I_a \) and \( I_b \) of point \( P \) represents the disparity at \( P \). Rectifying images aligns the horizontal rows of pixels in image \( I_a \) and \( I_b \) to speed up the matching process of those pixels.
that represent $P$ by reducing the search space. Disparity maps compute according to the primary camera in the stereo pair in order to maintain consistency with the re-projection matrix $Q$ computed in the camera calibration.

The parameters for the OpenCV block matching function allow for the total number of disparities to be set as well as the window size for block matching. Differing the parameters allows for either accuracy or reliability to be prioritized. The window size of the block matching algorithm defines the size of the pixel block to be matched. Increasing window size smooths the disparity map at the cost of accuracy. Smaller window sizes improves accuracy while increasing the potential for mismatches. Setting the number of disparities gives the maximum range of disparity values that the algorithm can assign to pixels [3]. Our implementation uses a balanced approach with a block size of 9 and 48 total disparities determined by testing and using OpenCV recommendations given sensor resolution and required accuracy.

Once a disparity map has been generated, the virtual world passes it through a speckle filter which removes outliers throughout the disparity. This research uses a speckle filter provided by OpenCV. Outliers are caused by mismatched image points of two different object point and can negatively effect registration. Figure 11 displays a disparity map of the C12 receiver in the refueling position. The disparity has been normalized from white to black in order to visualize the disparity values. Darker pixels represent the objects points that are farther away. Inversely, lighter pixels represent object points that are closer. Missing portions of the aircraft shown as black were filtered out by the speckle filter.

### 3.5 Re-projection and Point Cloud Generation

Disparity maps generated through the stereo block matching process provide the relationship between image points for re-projection. To ensure proper projection
of points into 3D space, a perspective transformation matrix $Q$ must be used as an input. The stereo computer vision module of OpenCV provides access to a re-projection function that computes the $x$, $y$ and $z$ components for each sensed point using the pixel coordinates $x_p$ and $y_p$ according to Equation 1 [3].

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = Q \ast \begin{bmatrix} p_x \\ p_y \\ \text{disparity}(p_x, p_y) \\ 1 \end{bmatrix}^T$$ (1)

Re-projecting with the above function for each pixel in the disparity map produces a matrix of 3D points representing the sensed aircraft. 3D points reprojected behind the camera or at a depth greater than 750 meters are filtered out to produce a filtered point cloud. This step must be done to ensure the point cloud only represents the aircraft and not the background. Following the removal of re-projection artifacts, a point cloud reduction removes points throughout the model to speed up computation. One out of every eight filtered points is extracted from the point cloud array to form the final minimized point cloud. Using one out of every eight points provides a consistent reduction of points throughout all of the sensed surfaces which was proven during early experimentation to reduce total point cloud size while retaining accuracy in registration. The filtered and minimized point cloud must then be transformed from
the camera coordinate frame into the 3DVW coordinate frame for each virtual world coordinate $[x_v \ y_v \ z_v]$ according to Equation 2.

$$[x_v \ y_v \ z_v] = [z \ (-x) \ (-y)] \quad (2)$$

Once the points are transformed into the 3DVW frame of reference they must be transformed to the position and orientation of the primary camera. Transformation is achieved by applying the respective position and orientation of the left camera in the world frame to each point in the point cloud. The corrected 3D points $[x' \ y' \ z']$ receive the following transformation and rotation in Equation 3 given a tanker direction cosine matrix $DCM_{tanker}$, left camera direction cosine matrix $DCM_{camera}$ and left camera position $[x_c \ y_c \ z_c]$.

$$[x' \ y' \ z']^\top = (DCM_{tanker} \ast \left( (DCM_{camera})^\top \ast [x_v \ y_v \ z_v]^\top \right)) + [x_t \ y_t \ z_t]^\top \quad (3)$$

The reduced and transformed 3D points serve as input for ICP registration of the sensed point cloud. Visualizing the points in the virtual world produces a point cloud seen in Figure 12. The leftmost image shows a visualization of the sensed receiver point cloud in the virtual world in the refueling envelope. The images on the right side of Figure 12 provide different views of the point cloud with and without the C-12 model. During early research the visualization serves to aid in ensuring accuracy of point cloud information by making biases and large orientation and position errors more obvious. Point cloud visualization can be disabled to improve performance or be shown to allow for enhanced visual analysis.
3.6 Model Registration and 6DOF Estimation

Registration of the sensed point cloud in the virtual world requires a reference model to use the ICP algorithm. This research uses a point-to-point ICP implementation with a modified KD-Tree approach to converge the reference point cloud to the sensed point cloud [20]. The algorithm tracks translations and rotations at each iteration of the registration to build a combined 6DOF measurement of the sensed point cloud. Figure 13 shows a visualization of the converged reference and sensed point cloud following 30 ICP iterations. ICP matches the reference point cloud (red) to the sensed point cloud (yellow).

The shelled reference model used in this research has a different center than truth model and must be transformed. When using the shelled reference model, the center point transformation is applied to the 6DOF registration measurement to ensure proper pose estimation. The 3DVW writes the final 6DOF measurement to file along with the model truth data for comparison.
3.7 Experimental Design

The stereo computer vision pipeline implemented in the 3DVW provides the data for each experiment performed in this work. The following subsections outline each of the three experiments used to produce the results presented in the following chapter. Unless otherwise noted, all tests were performed using MSAA enabled to ensure virtual imagery consistency. MSAA smooths the edges of objects being rendered in the 3DVW by changing the color of pixels located on the edge of the object as previously described in Section 2.7.

Experiments performed in this work use root-mean-square error (RMS error) to display the accuracy of 6DOF measurement. RMS error is computed according to Equation 4 where \( n \) is the number of samples and \( \hat{v}_p, v_p \) are the truth and estimated values for a particular degree of freedom at position \( p \), respectively [45].
\[RMS\ Error = \sqrt{\frac{\sum_{p=1}^{n}(\hat{v}_p - v_p)^2}{n}}\]  \hspace{1cm} (4)

All experiments using the AAR CONOPS flight path present statistical information for certain regions of the flight path. Selecting regions allows certain critical sections of the refueling envelope to be highlighted with specific performance metrics. Region \(A\) defines the approach of the receiver to a position behind the tanker at ranges between 70m and 175m. At this range relative position estimation computations are more useful than orientation estimations as the IMU of the receiver aircraft can provide more accurate orientation measurements. The intermediary region \(B\) which lies between 38m and 70m contains an initial approach. In region \(B\) at positions 39m to 41m away, the receiver moves such that its tail no longer lies within the camera viewing frusta resulting in position and pitch estimation errors. Partial receiver visibility limits accuracy in the data; however, mitigations and improvements are provided later in the future work section. The final stages of the refueling approach from 32m to 38m defines region \(C\) where accuracy is paramount to ensure safe operations. While in region \(C\), the receiver lies within the viewing frusta of both stereo cameras and moves primarily in the \(x\) direction. During refueling, the receiver will remain in the refueling envelope so approach values can provide insight into accuracy throughout the fuel transfer process.

**Experiment 1: Reference Model Performance Characterization.**

A characterization of the sensed point cloud and subsequent registration by ICP must be available in order to properly assess the accuracy and limitations of the stereo computer vision solution. A characterization of the shelled and full receiver reference models is performed with a maximum of 30 ICP iterations. Using 30 iterations provides a consistent stopping point for all experiments while still providing the ICP
algorithm with enough iterations to converge to a solution. During characterization the computational speed and accuracy of both reference models are computed and represent the performance metrics by which both models can be compared.

ICP registration depends on a good starting point to prevent pose estimates that minimize error yet result in an unrealistic receiver pose. Properly seeding ICP requires the position of the reference model to be transformed to the geometric center for the sensed point cloud. Furthermore, the reference model orientation is seeded to 0 degrees for roll, pitch and yaw as the receiver attitude should mirror that of the tanker during refueling operations. Both seeding methods performed use information available to the stereo computer vision solution and do not assume an a priori position.

Figure 14. ICP Characterization Positions in the 3DVW

Figure 14 shows different perspectives of the 558 receiver positions used for ICP characterization. Positions were selected to ensure even distribution throughout the frusta and full visibility of the receiver in both cameras at a range of 10 to 110 meters away from the stereo cameras. Corresponding orientations at each position are set to 0 degrees for roll, pitch, and yaw. The computer vision pipeline gets executed with the virtual C12 at each of the positions represented by the yellow dots in Figure 14. Both the shelled and full reference are assessed using RMS error as
well as the percent difference of average error magnitude for each degree of freedom. Results are separated into ranges of 20-60 meters and 60-110 meters to target smaller sets of data for comparison. Average execution times for registration within the two ranges show potential computational improvements between the shelled and full reference models.

**Experiment 2: Reference Model Flight Path Performance.**

Differences between the accuracy of the shelled and full reference models are further quantified using the AAR CONOPS flight path. The receiver pose updates according to the truth data for the flight path as the computer vision pipeline processes the synthetic imagery to produce 6DOF receiver pose estimations. Errors represent the mathematical difference between the sensed and truth values for the receiver 6DOF pose. The percent difference of the average error magnitude and RMS error, separated between regions A, B and C, provide metrics comparing 6DOF estimation accuracy between the full and shelled reference models. This experiment focuses exclusively on receiver position and orientation estimation. Additional metrics for orientation estimation are derived due to non-zero orientation truth data in the flight path.

**Experiment 3: Effect of MSAA on Virtual World Computer Vision.**

Effects of MSAA are quantified using camera calibration and the CONOPS flight path. A deterministic calibration in the virtual world using 10 poses produces RMS error, epipolar error and the re-projection matrix $Q$, which act as metrics for calibration performance. 5 tests are executed with the virtual sensors using no MSAA as well as MSAAx4, MSAAx8, MSAAx16, MSAAx32. Percent difference of RMS and epipolar error between calibration with and without MSAA displays improve-
ment in calibration quality. The relationship between an ideally computed $Q$ and the $Q$ resulting from calibration with no MSAA and MSAAx16 provides an accuracy improvement in terms of re-projection.

Pose estimation error using the CONOPS flight path and the shelled reference model also displays improvement in the stereo computer vision pipeline when using MSAA. Three tests are performed using the flight path to quantify the increase in performance using MSAA at different stages in the pipeline. The first test performs a flight path using a non-MSAA calibration and virtual cameras with MSAA disabled. Next, the flight path is executed in the 3DVW using a MSAA stereo calibration and with MSAA disabled in the virtual cameras. To specify the effect of MSAA on virtual imagery, this research performs a final test using a MSAA stereo calibration and with MSAA enabled in the virtual cameras. RMD error and percent change of error magnitude for each of the 6DOF estimation provide the performance between MSAA for AAR stereo computer vision in the 3DVW.

The results obtained from the shelled flight path using a MSAA calibration and MSAA enabled virtual cameras represents the capabilities of the virtual world to simulate computer vision for AAR. In addition, the results provide metrics to compare against flight test data that will be collected in 2017. Following this data collection, visual data from physical sensors can be compared to virtual imagery and the computer vision pipeline can be executed with physical sensor data to compare real-world and virtual world performance. Further analysis of the performance of the stereo computer vision pipeline will be explained in the corresponding results section.
IV. Results

4.1 Reference Model Performance Characterization

Using distributed positions throughout the viewing area of the stereo cameras allows for the performance of the shelled and full models to be compared. Figure 15 and 16 present the error in position estimation using the 3DVW stereo computer vision pipeline with the full and shelled models, respectively. A visual improvement in the accuracy of estimation occurs at ranges less then 70 meters with maximum error magnitudes less then 1 meter that converge towards 0.

![Graph showing position estimation error for ICP characterization of the full reference model at 30 iterations](image)

Figure 15. Position estimation error for ICP characterization of the full reference model at 30 iterations

Using a shelled model for ICP in the stereo computer vision pipeline has a more noticeable graphical difference in orientation estimation at each of the characterization receiver positions. Figure 17 displays the orientation error for the full model and Figure 18 for the shelled model orientation error. Inspection of the error graphs for both models shows improvement in estimation error with the most
improvement occurring in the pitch component. Similar to the results of position estimation, the shelled model also shows smaller deviations in error at all ranges.

6DOF estimation accuracy for both ICP models are presented in Table 1. At
ranges of 20 to 60 meters the shelled model displays at least a 29% improvement in all components of 6DOF estimation error than the same errors obtained using the full model. Comparing $Z$ and pitch error between the two models shows 62% reduction in error with the shelled reference model. The shelled model outperforms the full model at 60 to 110 meters as well providing a 19% or greater improvement in all components of 6DOF estimation error.

Table 1. Comparison of Average 6DOF Estimation Error For ICP Characterization using Shelled and Full Reference Models

<table>
<thead>
<tr>
<th>Range</th>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>Roll (°)</th>
<th>Pitch (°)</th>
<th>Yaw (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20m - 60m</td>
<td>Full Model</td>
<td>0.205</td>
<td>0.106</td>
<td>0.281</td>
<td>0.539</td>
<td>1.6922</td>
</tr>
<tr>
<td></td>
<td>Shelled Model</td>
<td>0.142</td>
<td>0.0458</td>
<td>0.104</td>
<td>0.381</td>
<td>0.641</td>
</tr>
<tr>
<td></td>
<td>Difference (%)</td>
<td>30.64</td>
<td>57.15</td>
<td>63.03</td>
<td>29.33</td>
<td>62.10</td>
</tr>
<tr>
<td>60m - 110m</td>
<td>Full Model</td>
<td>0.658</td>
<td>0.238</td>
<td>0.365</td>
<td>2.966</td>
<td>7.854</td>
</tr>
<tr>
<td></td>
<td>Shelled Model</td>
<td>0.532</td>
<td>0.182</td>
<td>0.230</td>
<td>1.668</td>
<td>2.827</td>
</tr>
<tr>
<td></td>
<td>Difference (%)</td>
<td>19.24</td>
<td>23.58</td>
<td>36.87</td>
<td>43.75</td>
<td>64.00</td>
</tr>
</tbody>
</table>

RMS error provides more insight into the accuracy of each reference model used by ICP. The RMS errors for ICP characterization are provided in Table 2 at each
range and reference model. At each range the shelled model has lower RMS errors for 6DOF receiver pose estimation further solidifying shelled model performance with respect to pose estimation accuracy.

Table 2. RMS Error of 6DOF Estimation For ICP Characterization using Shelled and Full Reference Models

<table>
<thead>
<tr>
<th>Range</th>
<th></th>
<th></th>
<th></th>
<th>Roll (°)</th>
<th>Pitch (°)</th>
<th>Yaw (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X (m)</td>
<td>Y (m)</td>
<td>Z (m)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20m - 60m</td>
<td>Full Model 0.261</td>
<td>0.122</td>
<td>0.291</td>
<td>0.771</td>
<td>2.335</td>
<td>1.339</td>
</tr>
<tr>
<td></td>
<td>Shelled Model 0.188</td>
<td>0.056</td>
<td>0.117</td>
<td>0.517</td>
<td>0.850</td>
<td>0.663</td>
</tr>
<tr>
<td>60m - 110m</td>
<td>Full Model 0.976</td>
<td>0.306</td>
<td>0.478</td>
<td>4.256</td>
<td>9.033</td>
<td>4.342</td>
</tr>
<tr>
<td></td>
<td>Shelled Model 0.780</td>
<td>0.251</td>
<td>0.336</td>
<td>3.094</td>
<td>3.850</td>
<td>2.840</td>
</tr>
</tbody>
</table>

Figure 19 presents the execution time of ICP in seconds using both the shelled and full model in the stereo computer vision pipeline. Shelled model execution times follow a consistent trend for each position with an execution time close to a tenth of a second. Execution time for the full model vary significantly across the characterization range with increases in execution time as the receiver gets closer to the stereo camera pair. Table 3 provides average ICP execution times and standard deviations of the execution time at each range. By using a shelled model for ICP, execution times per position can be reduced by 87% at 20 to 60 meters and by 78% at 60 to 110 meters. Furthermore, shelled model execution times have a standard deviation 28 times lower at 20 to 60 meters and 17 times lower at 60 to 110 meters.

Table 3. Average and Standard Deviation for ICP Execution Times using Shelled and Full Reference Models

<table>
<thead>
<tr>
<th>Range</th>
<th></th>
<th>Average Execution Time (sec)</th>
<th>Standard Deviation (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Model</td>
<td>1.1305</td>
<td>0.3855</td>
</tr>
<tr>
<td></td>
<td>Shelled Model</td>
<td>0.1386</td>
<td>0.0137</td>
</tr>
<tr>
<td>60m - 110m</td>
<td>Full Model</td>
<td>0.4590</td>
<td>0.2981</td>
</tr>
<tr>
<td></td>
<td>Shelled Model</td>
<td>0.1008</td>
<td>0.0175</td>
</tr>
</tbody>
</table>

The shelled model outperforms the full reference model in terms of accuracy and speed of 6DOF receiver pose estimation. It works on the premise that certain portions
of the receiver body will never be seen by the stereo vision system on the tanker. By using surface points there are fewer points to match against during the ICP portion of the stereo computer vision pipeline. In addition, a shelled reference model better reflects what the computer vision solution can compute further enhancing the accuracy of ICP registration.

4.2 Reference Model Flight Path Performance

Given the performance of ICP in general characterization, the shelled and full reference model is next compared further using a CONOPS flight path representative of a real-world operation. The shelled and full reference models are used as input for ICP in the 3DVW stereo computer vision pipeline with position and orientation estimation error presented in Figures 20 - 23.

The flight path position and orientation error provides improved results over general ICP characterization as the flight path maximizes the visibility of the receiver
Figure 20. Position estimation error for the shelled reference model during a CONOPS flight path

Figure 21. Orientation estimation error for the shelled reference model during a CONOPS flight path

with respect to the stereo cameras. Differences in error deviation and overall accuracy are exemplified in the orientation error for the shelled and full reference models presented in Figures 21 and 23. Vertical grouping of markers present in both orien-
Figure 22. Position estimation error for the full reference model during a CONOPS flight path.

Figure 23. Orientation estimation error for the full reference model during a CONOPS flight path.

tation error plots at approximately 49, 34 and 32 meters show computations where the receiver holds position to ensure a proper approach. With sub-centimeter level changes in position at close ranges, the error deviates 2 times greater when using the
the full model opposed to the shelled reference model. At ranges between 42 to 38 meters the tail of the receiver leaves the viewing area as seen in Figure 24 causing position and orientation estimation error with both reference models. Notice in the left image that there are no yellow points representing the tail of the receiver. With no sensed points ICP tries to minimize the error and in the processes tilts the nose of the estimated model up in the positive $Z$ direction as well as forward along the $X$ axis.

![Figure 24. 6DOF Receiver Estimation Error from Tail Exceeding Stereo Camera Viewing area at 38 to 42 meters away from the Cameras](image)

Despite the increase in reference points the full model does not provide any significant improvement when portions of the aircraft cannot be sensed. Position and orientation estimation for the CONOPS flight path improves over general ICP characterization when using the shelled reference model. Table 4 presents the average error magnitude for 6DOF pose estimation for regions $B$ and $C$ of the CONOPS flight path. Orientation values in region $A$ for both models were inconsistent and inaccurate due to the limited point cloud size at distances greater than 125 meters. Although point cloud size affects orientation estimation, position estimates between 175 and 125 meters display average position errors less than 1.5 meters with maximum errors of 7m making the measurements useful for initial position estimation.

The percent differences computed for each component show improvement in pose estimation error in regions $A$, $B$ and $C$. The largest improvement in error reduction
### Table 4. Comparison of Average 6DOF Estimation Error For CONOPS Flight Path using Shelled and Full Model

<table>
<thead>
<tr>
<th>Region</th>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>Roll (°)</th>
<th>Pitch (°)</th>
<th>Yaw (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Model</td>
<td>0.842</td>
<td>0.437</td>
<td>1.357</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shelled Model</td>
<td>0.762</td>
<td>0.315</td>
<td>1.352</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (%)</td>
<td>9.45</td>
<td>27.92</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Model</td>
<td>0.144</td>
<td>0.196</td>
<td>0.303</td>
<td>1.089</td>
<td>1.372</td>
<td>1.376</td>
</tr>
<tr>
<td>Shelled Model</td>
<td>0.115</td>
<td>0.065</td>
<td>0.097</td>
<td>0.365</td>
<td>0.471</td>
<td>0.526</td>
</tr>
<tr>
<td>Difference (%)</td>
<td>20.40</td>
<td>66.88</td>
<td>68.19</td>
<td>66.45</td>
<td>65.65</td>
<td>61.78</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Model</td>
<td>0.100</td>
<td>0.112</td>
<td>0.324</td>
<td>0.399</td>
<td>0.677</td>
<td>0.715</td>
</tr>
<tr>
<td>Shelled Model</td>
<td>0.064</td>
<td>0.028</td>
<td>0.065</td>
<td>0.112</td>
<td>0.161</td>
<td>0.163</td>
</tr>
<tr>
<td>Difference (%)</td>
<td>36.11</td>
<td>75.04</td>
<td>79.88</td>
<td>71.83</td>
<td>76.30</td>
<td>77.14</td>
</tr>
</tbody>
</table>

occurs when using the shelled model for ICP in region C. Shelled model registration reduces error by at least 36% compared to the full model when the receiver makes its final approach to the refueling envelope. In all 6DOF components other than X, the improvement increases to at least 71%. In region C the average magnitude of error is approximately 6 centimeters in X and Z and 3 centimeters in the Y. The X component error differs less between reference models as X error depends on ICP minimization of receiver surfaces, such as the wings and tail, in the X and Y directions. Both the shelled and full reference models have more consistent point densities in the X and Y directions then they do in the Z direction.

RMS errors for each region of the flight path, found in Table 5, provide an additional assessment of 6DOF estimation error resulting from the use of full and shelled models in the 3DVW stereo computer vision pipeline. For all regions, the shelled reference model provides lower RMS errors for estimations. Region C position RMS errors are below 0.07 for the shelled model compared to 0.325 for the full model matching the reduction in estimation error shown in Table 4. The smallest amount of error, especially in region C, ensures safe and effective refueling operations making the shelled model a better choice when using the stereo computer vision pipeline.
Table 5. RMS Error of 6DOF Estimation For CONOPS Flight Path using Shelled and Full Reference Model

<table>
<thead>
<tr>
<th>Region</th>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>Roll (°)</th>
<th>Pitch (°)</th>
<th>Yaw (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Model</td>
<td>1.111</td>
<td>0.606</td>
<td>1.877</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shelled Model</td>
<td>1.053</td>
<td>0.484</td>
<td>1.876</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region B</td>
<td>Full Model</td>
<td>0.216</td>
<td>0.213</td>
<td>0.331</td>
<td>1.397</td>
<td>1.840</td>
</tr>
<tr>
<td></td>
<td>Shelled Model</td>
<td>0.162</td>
<td>0.085</td>
<td>0.122</td>
<td>0.503</td>
<td>0.754</td>
</tr>
<tr>
<td>Region C</td>
<td>Full Model</td>
<td>0.111</td>
<td>0.118</td>
<td>0.325</td>
<td>0.502</td>
<td>0.773</td>
</tr>
<tr>
<td></td>
<td>Shelled Model</td>
<td>0.069</td>
<td>0.032</td>
<td>0.067</td>
<td>0.135</td>
<td>0.188</td>
</tr>
</tbody>
</table>

4.3 Effect of MSAA

Virtual imagery contains aliasing artifacts that result from the rendering process of the virtual world. The aliasing artifacts shown in the left image of Figure 25 affect the camera calibration as well as the stereo computer vision process. Using MSAA when rendering from the virtual cameras removes the artifacts as seen in the right image of Figure 25.

![Figure 25. Virtual Checkerboard With MSAA Disabled (Left) and Enabled (Right)](image)

Improving the precision of checkerboard corners makes virtual calibration imagery more representative of the real world. Table 6 displays the RMS and epipolar errors for calibrations using different levels of MSAA sampling.

With the exception of MSAAx4 and MSAAx8, every MSAA improves both cali-
Table 6. Effect of MSAA on RMS Re-projection and Epipolar Constraint Calibration Error

<table>
<thead>
<tr>
<th>Type of error (pixels)</th>
<th>No MSAA</th>
<th>MSAAx4</th>
<th>MSAAx8</th>
<th>MSAAx16</th>
<th>MSAAx32</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS Re-projection</td>
<td>0.243865</td>
<td>0.243687</td>
<td>0.242856</td>
<td>0.124402</td>
<td>0.123115</td>
</tr>
<tr>
<td>Epipolar Constraint</td>
<td>0.145751</td>
<td>0.147311</td>
<td>0.146923</td>
<td>0.053891</td>
<td>0.053862</td>
</tr>
</tbody>
</table>

Calibration error metrics with the largest jump in error reduction occurring with the use of MSAAx16. Using 4 or 8 samples for MSAA does not produce imagery superior to that produced without MSAA due to the multi pixel gap between aliased checkerboard corners. MSAAx4 and MSAAx8 produces larger errors by shifting aliased corners farther from their original rendered positions. MSAAx16 reduces the RMS error from unmodified virtual imagery by 48% and the epipolar error by 63%. No significant improvement in calibration error occurs between MSAAx32 and MSAAx16. Due to the computational overhead of adding more samples, MSAAx16 provides the best performance to cost improvement over unmodified imagery.

The re-projection matrix $Q$ produced from the calibration sees improvement from MSAA as well. Table 7 displays the percent error between the ideal $Q$ and those generated by the calibrations. Computing an ideal $Q$ uses the camera focal length and principal point provided by the manufacturer of the physical cameras to produce a truth value for each element. $Q$ produced through the MSAAx16 calibration better reflects the ideal values than the calibration without MSAA. Although percent errors for the $Q$ matrix resulting from unmodified imagery has errors less than 2% for all elements, the error will be exacerbated in re-projection of points at farther ranges from the camera. To insure the best possible results for the estimation of 6DOF receiver measurements in the stereo computer vision pipeline, the calibration and $Q$ matrix generated via MSAA imagery are used in the 3DVW.

Flight tests executed in the 3DVW with and without an MSAA calibration present the effect of an incorrect $Q$ on the 6DOF estimation capabilities on the stereo com-
Table 7. Elementwise Percent Error from Ideal Re-projection Matrix $Q$

<table>
<thead>
<tr>
<th></th>
<th>No MSAA $Q$</th>
<th>MSAAx16 $Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0</td>
<td>1.615%</td>
<td>0.129%</td>
</tr>
<tr>
<td>0 0 0</td>
<td>0.083%</td>
<td>0.001%</td>
</tr>
<tr>
<td>0 0 0</td>
<td>0.120%</td>
<td>0.091%</td>
</tr>
<tr>
<td>0 0 0.525%</td>
<td>0</td>
<td>0.274%</td>
</tr>
</tbody>
</table>

puter vision pipeline. Figures 26 and 27 present the error in position and orientation, respectively, for a CONOPS flight path with MSAA disabled. Figure 28 and 29 show the corresponding errors for the same flight path with MSAA enabled for calibration. A visual improvement can be seen in the consistency and accuracy of position estimation when using the calibration as opposed to unmodified virtual imagery especially in region $C$. Orientation estimates are graphically similar with small improvements in accuracy in region $C$.

Figure 26. Position estimation error for AAR CONOPS Flight path with MSAA disabled

Table 8 presents the numerical difference between the two methods in terms of average magnitude of 6DOF estimation error in each flight path region. Percent
change is measured in terms of 6DOF error reduction with respect to no MSAA estimation error. A calibration performed with MSAA reduces orientation error in every region except for A in the roll component. Position estimation error in the Y
component decreased significantly by 81% in Region B and 91% in Region C.

Table 8. Comparison of Average 6DOF Estimation Error For CONOPS Flight Path With and Without MSAA Calibration

<table>
<thead>
<tr>
<th></th>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>Roll (°)</th>
<th>Pitch (°)</th>
<th>Yaw (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSAA Calib</td>
<td>1.055</td>
<td>0.529</td>
<td>2.170</td>
<td>5.820</td>
<td>4.859</td>
<td>5.055</td>
</tr>
<tr>
<td>No MSAA</td>
<td>1.297</td>
<td>0.754</td>
<td>2.202</td>
<td>5.548</td>
<td>5.933</td>
<td>5.215</td>
</tr>
<tr>
<td>Difference (%)</td>
<td>18.66</td>
<td>29.91</td>
<td>1.46</td>
<td>-4.92</td>
<td>18.10</td>
<td>3.06</td>
</tr>
<tr>
<td>Region B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSAA Calib</td>
<td>0.124</td>
<td>0.065</td>
<td>0.116</td>
<td>0.355</td>
<td>0.551</td>
<td>0.484</td>
</tr>
<tr>
<td>No MSAA</td>
<td>0.156</td>
<td>0.348</td>
<td>0.114</td>
<td>0.544</td>
<td>0.554</td>
<td>0.815</td>
</tr>
<tr>
<td>Difference (%)</td>
<td>20.44</td>
<td>81.28</td>
<td>-2.49</td>
<td>34.69</td>
<td>0.41</td>
<td>40.63</td>
</tr>
<tr>
<td>Region C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSAA Calib</td>
<td>0.040</td>
<td>0.017</td>
<td>0.038</td>
<td>0.121</td>
<td>0.142</td>
<td>0.158</td>
</tr>
<tr>
<td>No MSAA</td>
<td>0.037</td>
<td>0.200</td>
<td>0.064</td>
<td>0.300</td>
<td>0.166</td>
<td>0.499</td>
</tr>
<tr>
<td>Difference (%)</td>
<td>-7.98</td>
<td>91.63</td>
<td>40.13</td>
<td>59.68</td>
<td>14.10</td>
<td>68.27</td>
</tr>
</tbody>
</table>

A visual representation of the $Y$ component error can be seen in Figure 30 with different perspectives of the reference point cloud and sensed receiver point cloud.

Large $Y$ error stems from the re-projection process and a less accurate calibration. Higher calibration error previously presented in Table 6 results in a less accurate re-projection matrix $Q$. When the depth values are re-projected into 3D space from the
primary (left) camera the error manifests in a combined vector pointing primarily in the positive $Y$ as image points of the receiver lie right of the image center. As the receiver position from the cameras increases, so does the error in the $Y$ direction. The more accurate calibration computed using MSAA diminishes the errors making estimation more accurate especially in regions $B$ and $C$.

MSAA used to enhance camera calibration also effects the stereo computer vision pipeline when applied to virtual imagery of the receiver prior to disparity map computation. Figures 31 and 32 represent position and orientation errors, respectively, of the stereo computer vision pipeline when using MSAA for both calibration and virtual imagery. For clarity, MSAA flight paths use MSAA for both calibration and virtual imagery acquisition. Partial MSAA flight paths only use MSAA for a camera calibration.

Results for the MSAA flight path show similar orientation estimation errors to those using only a MSAA calibration. Graphical improvements in the MSAA flight path manifest in region $C$ with improved error consistency between $X$, $Y$ and $Z$. Table 8 provides quantifiable differences between the MSAA and partial MSAA flight path 6DOF estimation errors in terms of average error magnitude per region.

Virtual imagery processed by MSAA improves 6DOF measurement in region $A$ by at least 20% over unmodified virtual imagery. Improvement in error likely results
from the addition of image points to smooth the edge of the receiver model at ranges greater than 70 meters. For region B, MSAA decreases error by at least 9% for X, Z and pitch while increasing estimation error by 12% for yaw and under 6% for Y and roll. MSAA provides no improvement over partial MSAA with a maximum increase in orientation error by 16% and a maximum increase in position estimation error of 74% for region C. Although a MSAA calibration by itself improves pose estimation, its use does not reflect physical imagery that does not contain similar 3DVW aliasing. Furthermore, the differences in average position estimation error only changes by 2.5 centimeters between partial MSAA and MSAA in the X and Z components. Ensuring the applicability of these results to real-world data and imagery improves the effect of the 3DVW and the stereo computer vision pipeline on AAR research and future development.

The pose estimation errors for the MSAA flight path presented in Table 9 as well as Figures 31 and 32 represents the capability of the stereo computer vision pipeline
to estimate the position of the C12 receiver in the 3DVW. The results of this research are significant with average error magnitudes less than 7 centimeters for position estimation and 0.15° for orientation estimation in region $C$. A comparison of errors between no MSAA and MSAA for calibration and virtual imagery can be found in Appendix B along with RMS errors between no MSAA, partial MSAA and MSAA.

Pose estimation errors for the stereo computer vision pipeline in the 3DVW will be compared with results using imagery from a flight test scheduled for 2017. Given the performance of the pipeline, the effectiveness of the 3DVW to simulate AAR can be quantified and improved to provide a more realistic and effective simulation tool.
Table 9. Comparison of Average 6DOF Estimation Error For CONOPS Flight Path between MSAA Calibration Only and MSAA Enabled

<table>
<thead>
<tr>
<th>Region</th>
<th></th>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>Roll (°)</th>
<th>Pitch (°)</th>
<th>Yaw (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSAA</td>
<td>0.656</td>
<td>0.314</td>
<td>1.125</td>
<td>3.938</td>
<td>3.780</td>
<td>3.343</td>
</tr>
<tr>
<td></td>
<td>MSAA Calib</td>
<td>1.055</td>
<td>0.529</td>
<td>2.170</td>
<td>5.820</td>
<td>4.859</td>
<td>5.055</td>
</tr>
<tr>
<td></td>
<td>Difference (%)</td>
<td>37.84</td>
<td>40.56</td>
<td>48.16</td>
<td>32.34</td>
<td>22.21</td>
<td>33.87</td>
</tr>
<tr>
<td></td>
<td>MSAA</td>
<td>0.113</td>
<td>0.066</td>
<td>0.100</td>
<td>0.378</td>
<td>0.477</td>
<td>0.541</td>
</tr>
<tr>
<td></td>
<td>MSAA Calib</td>
<td>0.124</td>
<td>0.065</td>
<td>0.116</td>
<td>0.355</td>
<td>0.551</td>
<td>0.484</td>
</tr>
<tr>
<td></td>
<td>Difference (%)</td>
<td>9.06</td>
<td>-0.72</td>
<td>13.68</td>
<td>-6.40</td>
<td>13.45</td>
<td>-11.88</td>
</tr>
<tr>
<td></td>
<td>MSAA</td>
<td>0.061</td>
<td>0.029</td>
<td>0.063</td>
<td>0.123</td>
<td>0.153</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>MSAA Calib</td>
<td>0.040</td>
<td>0.017</td>
<td>0.038</td>
<td>0.121</td>
<td>0.142</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>Difference (%)</td>
<td>51.41</td>
<td>-74.83</td>
<td>-66.09</td>
<td>-2.19</td>
<td>-7.82</td>
<td>-17.61</td>
</tr>
</tbody>
</table>
V. Conclusion

5.1 State of AAR

UAVs currently lack the ability to refuel using traditional methods due to command and control latency between pilot input and aircraft pose adjustment. The ability for enemies to prevent automated refueling operations by denying GPS leads to research using alternate sensors to estimate position and aid receiver IMUs. Additional methods of computer vision and lasers require the modification of either the tanker or receiver to perform relative receiver pose estimation.

5.2 Research Conclusions

By using stereo computer vision cameras available on the tanker, position and orientation of a receiver can be determined without modifying the exterior of either the tanker or receiver. The 3DVW has been shown as an effective tool to visualize the AAR process and simulate AAR computer vision. Using the 6DOF pose estimation error provided in this research the ability for the virtual world to simulate the AAR process can be compared with flight test data in 2017. With further modification following its comparison to real flight test imagery and data, the 3DVW can also be an effective tool to enhance further AAR research by accelerating the research and development process and providing visualizations for AAR concepts and errors.

Using a shelled receiver reference model yields at least a 29% improvement in 6DOF estimation error for characterization and a 36% improvement in 6DOF estimation error for the AAR CONOPS flight path compared to the full reference model. Real-time requirements for AAR benefit from using the shelled model through the reduction of ICP computational time by 87% at 20 to 60 meters and by 78% at 60 to 110 meters.
Using MSAA makes virtual imagery more representative of the real world and improves stereo computer vision simulation in the 3DVW with improved calibration error and pose estimation in region A. Performing a better calibration using imagery modified by MSAAx16 results in a re-projection matrix $Q$ that better represents the ideal values for the stereo camera pair.

5.3 Research Contributions

This research presents 6DOF relative receiver pose estimation algorithm showing accuracy on the order of 6cm at the refueling position. Performance improvements through the use of a shelled reference model are quantified with respect to pose estimation accuracy and computational speed. The effects of MSAA on the 3DVW are assessed in terms of calibration, re-projection matrix accuracy and flight path performance. Comparison of the 3DVW with results from a live AAR CONOPS flight test scheduled for late 2017 will further evaluate the simulation capability of the 3DVW to help improve further AAR research.

5.4 Recommendations for Future Work

This work make many assumptions to scope the AAR problem. The tanker boom causes occlusions of the receiver throughout many portions of the CONOPS AAR flight path. Occlusion can severely affect the results of pose estimation and must be taken into account to develop a full AAR pose estimation solution. Implementing the stereo computer vision pipeline with another version of ICP could improve occluded registration while receiving the performance boosts of the shelled model.

Propellers found on many UAVs must be taken into account when using computer vision, especially in a simulated environment. This research currently ignores propellers since static propellers occlude the wings and engines which are primary
matching surfaces for ICP. A proper implementation of spinning propellers in the 3DVW as well as any relevant visible effects must be implemented with advanced motion blurring techniques to better match real-world flight test results.

Adding distortion to virtual imagery will further improve the realism of the 3DVW and produce synthetic imagery that is better representative of imagery from physical sensors. The stereo computer vision pipeline will also be effected by camera distortion. Rectifying this distortion properly must also be implemented before a system moves to hardware and the tanker aircraft.

Calibration results from the virtual world many not be easily obtained in the real-world due to physical sensors, distortion and other factors. Developing a system to continuously update camera parameters could help aid in operational effectiveness for a computer vision solution. Using the GPS, when available, and on board IMUs the stereo computer vision system could continuously calibrate and improve estimation capabilities through a flight. Moreover, using neural networks to train a camera calibration could use virtual imagery and modify the calibration to produce better results depending on receiver aircraft type and refueling conditions.
Appendix A. Virtual World Determinism

Although the 3DVW is deterministic, the disparity map produced by OpenCV is not deterministic such that disparity computations using the same input image can produce a slightly different disparity map each execution. In addition, the ordering of the OpenCV array representing the data prevents a deterministic sub-sampling of the disparity values for re-projection. Point clouds projected into the 3DVW change but have minimal effects on registration with ICP. At further distances the effect increases slightly. The standard deviations presented in Figure 33 and 34 are computed by computing the receiver pose 10 times for multiple positions throughout the CONOPS flight path. With exception of one outlier at approximately 150 meters from the stereo cameras, every position standard deviation shown in Figure 33 lies at or below 5 centimeters with standard deviations less then 2.5 centimeters at distances closer than 50 meters. The deviations do not take into account the re-seeding of ICP with the previously computed receiver position. Ignoring this effect would drop the standard deviation further. Orientation standard deviations fluctuate more throughout the receiver positions but with no standard deviations greater then 0.4 degrees. Based on the current accuracy of the stereo computer vision pipeline in the virtual world the expected errors will have minimal effect on the results. Further work should seek to more accurately quantify the determinism of the virtual world as well as minimize the effect of non deterministic libraries on the 3DVW itself.
Figure 33. Standard Deviation of 10 Estimations of Position For Receiver Poses Throughout the AAR CONOPS Flight Path

Figure 34. Standard Deviation of 10 Estimations of Orientation For Receiver Poses Throughout the AAR CONOPS Flight Path
## Appendix B. MSAA Additional Data

Table 10 provides differences between error magnitude for a flight path with MSAA enabled and disabled.

**Table 10. Comparison of Average 6DOF Estimation Error For CONOPS Flight Path With and Without MSAA Enabled**

<table>
<thead>
<tr>
<th>Region</th>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>Roll (°)</th>
<th>Pitch (°)</th>
<th>Yaw (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>MSAA</td>
<td>0.656</td>
<td>0.314</td>
<td>1.125</td>
<td>3.938</td>
<td>3.780</td>
</tr>
<tr>
<td></td>
<td>No MSAA</td>
<td>1.297</td>
<td>0.754</td>
<td>2.202</td>
<td>5.548</td>
<td>5.933</td>
</tr>
<tr>
<td></td>
<td>Difference (%)</td>
<td>49.439</td>
<td>58.342</td>
<td>48.911</td>
<td>29.014</td>
<td>36.288</td>
</tr>
<tr>
<td>A</td>
<td>MSAA</td>
<td>0.113</td>
<td>0.066</td>
<td>0.100</td>
<td>0.378</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>No MSAA</td>
<td>0.156</td>
<td>0.348</td>
<td>0.114</td>
<td>0.544</td>
<td>0.554</td>
</tr>
<tr>
<td></td>
<td>Difference (%)</td>
<td>27.649</td>
<td>81.141</td>
<td>11.532</td>
<td>30.510</td>
<td>13.802</td>
</tr>
<tr>
<td>A</td>
<td>MSAA</td>
<td>0.061</td>
<td>0.029</td>
<td>0.063</td>
<td>0.123</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>No MSAA</td>
<td>0.037</td>
<td>0.200</td>
<td>0.064</td>
<td>0.300</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>Difference (%)</td>
<td>-63.489</td>
<td>85.364</td>
<td>0.557</td>
<td>58.799</td>
<td>7.376</td>
</tr>
</tbody>
</table>

Table 11 provides the RMS errors associated with each MSAA flight path test. Smaller RMS errors are associated with more accurate estimations for 6DOF receiver pose.

**Table 11. RMS Error of 6DOF Estimation For CONOPS Flight Path comparing MSAA Enabled, MSAA Disabled, and MSAA Calibration Only**

<table>
<thead>
<tr>
<th>Region</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Roll</th>
<th>Pitch</th>
<th>Yaw</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>MSAA</td>
<td>0.656</td>
<td>0.314</td>
<td>1.125</td>
<td>3.938</td>
<td>3.780</td>
</tr>
<tr>
<td></td>
<td>MSAA Calib</td>
<td>1.499</td>
<td>0.803</td>
<td>3.116</td>
<td>9.509</td>
<td>7.312</td>
</tr>
<tr>
<td></td>
<td>No MSAA</td>
<td>1.721</td>
<td>0.917</td>
<td>3.102</td>
<td>8.984</td>
<td>9.183</td>
</tr>
<tr>
<td>Region B</td>
<td>MSAA</td>
<td>0.113</td>
<td>0.066</td>
<td>0.100</td>
<td>0.378</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>MSAA Calib</td>
<td>0.175</td>
<td>0.090</td>
<td>0.176</td>
<td>0.468</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>No MSAA</td>
<td>0.204</td>
<td>0.357</td>
<td>0.158</td>
<td>0.657</td>
<td>1.012</td>
</tr>
<tr>
<td>Region C</td>
<td>MSAA</td>
<td>0.061</td>
<td>0.029</td>
<td>0.063</td>
<td>0.123</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>MSAA Calib</td>
<td>0.047</td>
<td>0.020</td>
<td>0.042</td>
<td>0.149</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>No MSAA</td>
<td>0.045</td>
<td>0.201</td>
<td>0.067</td>
<td>0.314</td>
<td>0.192</td>
</tr>
</tbody>
</table>
Bibliography


33. John Valasek, Kiran Gunnam, Jennifer Kimmett, Monish D Tandale, John L Junkins, and Declan Hughes. Vision-based sensor and navigation system for


44. Blender Online Community. *Blender - a 3D modelling and rendering package.* Blender Foundation, Blender Institute, Amsterdam,

Improving Automated Aerial Refueling Stereo Vision Pose Estimation Using A Shelled Reference Model

Parsons, Christopher A., 2nd LT, USAF

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

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

Automated Aerial Refueling of Unmanned Aerial Vehicles is vital to the United States Air Force’s continued air superiority. This research presents a novel solution for computing a relative 6 degree-of-freedom pose between the refueling aircraft and a tanker. The approach relies on a real time 3D virtual simulation environment that models a realistic refueling scenario. Synthetic imagery is processed by computer vision algorithms that calculate the sensed relative-navigation position and orientation. Pose estimation accuracy and computational speed during registration improve through the use of a shelled reference model. The shelled model improves computational speed of pose estimation at the refueling position by 87% and accuracy by 36% when compared with a full reference model. To ensure proper simulation of computer vision concepts, this research quantifies the effect Multi-Sample Anti Aliasing implemented in the virtual stereo cameras on camera calibration and pose estimation. A combined shelled model and Multi-Sample Anti Aliased approach leads to position estimation errors less then 7cm and orientation estimation error less then 1°.


Dr. Scott Nykl, PhD, AFIT/ENG
(937) 255-3636, x4395; scott.nykl@afit.edu