VISION-BASED VIBRATION MONITORING OF CIVIL STRUCTURAL DAMAGE DETECTION USING A LOW-COST AUTONOMOUS UAV

A Thesis

by

JAVIER BECERRIL

Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Major Subject: Computer Science

The University of Texas Rio Grande Valley

May 2026

VISION-BASED VIBRATION MONITORING OF

CIVIL STRUCTURAL DAMAGE DETECTION

USING A LOW-COST AUTONOMOUS UAV

A Thesis
by
JAVIER BECERRIL

COMMITTEE MEMBERS

Dr. Qi Lu

Chair of Committee

Dr. Constantine Mohammad Tarawneh

Committee Co-Chair

Dr. Jinghao Yang

Committee Member

Dr. Dongchul Kim

Committee Member

May 2026

ABSTRACT

Becerril, Javier. Vision-Based Vibration Monitoring of Civil Structural Damage Detection Using a Low-Cost Autonomous UAV. Master of Science (MS), May 2026, 41 pp., 11 tables, 14 figures, 38 references.

This thesis presents the development and experimental validation of a low-cost autonomous unmanned aerial vehicle (UAV) for non-contact vibration-based structural damage detection. We extract structural vibration signals directly from video recordings captured by the onboard camera during flight. The extracted vibration signals are analyzed in the frequency domain to identify natural frequency shifts associated with structural degradation and damage.

To evaluate the effectiveness of the proposed approach, a laboratory-scale steel frame structure is tested under both healthy and simulated damage conditions. In contrast to advanced commercial UAV inspection systems, the developed UAV achieves comparable inspection capability at a significantly lower cost of approximately $1,000, making it suitable for scalable and cost-sensitive monitoring applications. The UAV-collected vibration signals are also compared with signals obtained from multiple stationary sensing devices, including contact accelerometers, smartphones, and a USB camera. The reference signal is captured from a stationary board used to isolate and remove UAV-induced vibration, resulting in a more accurate estimation of the true structural vibration. The corrected vibration measurements exhibit a higher percentage error of up to 5.7%; the results remain sufficiently accurate for reliable structural damage identification.

The findings indicate that our low-cost UAV provides a practical and flexible solution for structural health monitoring, particularly in environments where traditional contact-based sensing is difficult or impractical. Furthermore, this work highlights the potential for future deployment of multiple cooperative UAVs to improve inspection coverage, robustness, and efficiency in large-scale structural monitoring applications.

DEDICATION

The completion of this thesis would not be possible without the support of my family and friends. Sin el apoyo de mi familia, alcanzar mis sueños y metas no sería posible.

ACKNOWLEDGMENTS

I want to express my sincere gratitude to my advisor, Dr. Lu, for his guidance and supervision, and to the MARS (Multiple Autonomous Robot Systems) lab for their support. Thanks to my committee members for their comments and guidance. In addition, thanks to our collaborators who worked on this project, Dr. Mohsen Amjadian, and his students Maximiliano Vargas and Jennifer Herrera, in Civil Engineering at UTRGV, and my lab mates Joanna Gutierrez and Jorge Rios, in MARS lab. I also acknowledge the financial support provided by the NSF Expand AI ARISE project (No. 2434916), the NSF CREST Center for Multidisciplinary Research Excellence in Cyber-Physical Infrastructure Systems (MECIS) (No. 2112650), and the NSF MSI program (No. 2318682).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

Routine close-range inspection of civil infrastructure is among the most labor-intensive and dangerous tasks in infrastructure maintenance. Many critical inspection points are physically inaccessible without scaffolding, rope access, or specialized equipment. Inspecting the underside of a bridge requires working over water or traffic, examining electrical towers presents a safety risk that limits direct contact, and contaminated industrial zones require a distanced analysis [35]. For each of these scenarios, the ability to deploy a sensor to a precise location remotely without physical access or a trained operator manually positioning a platform represents a meaningful reduction in both cost and risk [10].

Conventional contact sensors such as accelerometers provide accurate dynamic response data but require physical installation, wiring, and maintenance at each measurement point. For large-scale or difficult-to-access structures, deploying and maintaining a contact sensor network is both logistically challenging and expensive, limiting the scalability of traditional monitoring approaches [5, 31]. Non-contact vision systems based on fixed cameras address the installation problem but introduce spatial limitations of their own. Stationary platforms such as smartphones and USB cameras require a direct line of sight to the inspection target from a fixed position. Testing demonstrated that these platforms achieve high frequency identification accuracy on a laboratory steel frame [3], but are limited to their stationary positions. They cannot access elevated, hazardous, or interior inspection points and must be manually repositioned for each new measurement location. For structures with many distributed inspection points, this constraint limits scalability.

UAVs extend the spatial reach of fixed cameras by accessing those elevated or hazardous locations without permanent sensor installation. However, prior UAV-based structural monitoring work has relied on commercial platforms such as the DJI Mavic 3 series and the DJI Matrice

210, which range from approximately $2,000 to $10,000 [10, 15]. Beyond cost, these systems require an operator to manually position the UAV before each data collection session and run on proprietary platforms that cannot be adapted to carry different sensor payloads. Existing work has identified the need for approaches capable of improving measurement consistency while separating structural motion from UAV-induced disturbances during data collection [15].

To address these limitations, this work presents a low-cost autonomous UAV system for vibration-based structural health monitoring (SHM). The system autonomously aligns at a defined distance from the inspection target without GPS or manual input. As the UAV hovers in place, the onboard camera records structural motion to extract vibration signals and analyze natural frequency shifts to identify damage-induced changes in structural health. A stationary reference point is used to remove the UAV-induced motion from the vibration signal, leaving only the structural vibration response. The UAV measurement accuracy is compared against contact accelerometers, smartphones, and a USB camera, with all platforms tested on the same structure under the same conditions. The contributions of this thesis are the development of a low-cost UAV platform that combines autonomous positioning and removal of UAV-induced motion to extract vibration signals to assess structural health, tested against multiple platforms.

Chapter 2 reviews related work in autonomous UAV inspection, visual servoing, and UAV-based structural displacement measurement, identifying the gap that this work addresses. Chapter 3 presents the methodology and describes the UAV hardware platform, onboard components, the SE(3) pose estimation, autonomous controller, signal processing, and experimental structure. Chapter 4 presents the results, the position lock accuracy from stationary tags, and the natural frequency identification accuracy from the vibration experiment. Chapter 5 discusses accuracy and comparison to prior work. Chapter 6 concludes with contributions, limitations, and future work.

CHAPTER II

RELATED WORK

2.1 Vibration-Based Structural Health Monitoring

Vibration-based structural health monitoring (SHM) exploits the fundamental relationship between a structure's dynamic properties and its physical condition. Structural damage, whether from fatigue cracking, connection loosening, or redistribution of mass alters the stiffness or mass distribution of the system, producing measurable shifts in its natural frequencies, damping ratios, and mode shapes [7, 17]. Monitoring these shifts provides a non-invasive indicator of structural degradation without requiring physical access to the interior or surface of the structure.

Conventional vibration monitoring relies on contact sensors, primarily accelerometers, physically mounted to the structure. These sensors deliver accurate response data but require direct access for installation and maintenance. For large-scale or remote infrastructure such as long-span bridges, elevated towers, wind turbines, deploying a sensor network is both logistically challenging and expensive, motivating the development of non-contact alternatives [5, 31, 1].

2.2 Vision-Based Vibration Measurement

Computer vision methods extract structural displacement and vibration signatures from video recordings without physical contact. Digital Image Correlation (DIC) tracks the intensity pattern of natural surface texture across frames [30, 24]. Optical flow algorithms estimate dense motion fields by tracking pixel intensity patterns between frames. Both techniques can produce displacement time histories from which spectral analysis identifies natural frequencies [7, 14]. Recent work validated DIC-based analysis for detecting damage-induced frequency shifts in laboratory civil structures, confirming that image-based displacement measurement can resolve the small amplitude changes associated with structural condition changes [2, 18].

Vision-based pose estimation methods offer a distinct advantage over feature tracking. A marker's geometry is precisely known, allowing full 6-DOF pose estimation via the Perspective-n-Point (PnP) algorithm rather than 2D pixel displacement alone. This provides physical-unit displacement directly, without a separate pixel-to-distance calibration step. Studies have validated pose-estimation-based methods for structural vibration identification, achieving modal frequency errors below 2% on laboratory specimens [23]. A comparison of visual target families ARTag, AprilTag, ArUco, and STag established that AprilTag and ArUco represent the current state of the art for detection speed and robustness [13]. AprilTag's faster detection throughput makes it preferable for real-time embedded applications, which is why it was selected for this work.

Vision-based methods carry well-documented limitations. Camera noise, motion blur, and illumination sensitivity degrade accuracy at large standoff distances [7, 38]. Double-differentiating displacement to obtain acceleration amplifies noise by $1/\Delta t^2$, which at low frame rates (15 fps) dominates the acceleration signal [14, 26]. For frequency identification, this does not affect the result. The PSD peak location is invariant to whether displacement or acceleration is analyzed, but it prevents reliable amplitude measurement at the frame rates available on embedded hardware.

2.3 Autonomous UAV Positioning for Close-Range Inspection

Autonomous close-range inspection requires the UAV to hold a stable, precisely defined position relative to an inspection target and not just hover in the vicinity. This capability is distinct from GPS waypoint navigation, which provides meter-level accuracy insufficient for sensor-target alignment, and from optical flow station-keeping, which maintains zero ground-relative velocity but does not constrain the drone's position relative to a specific target point.

Image-based visual servoing (IBVS) is the general framework for controlling a robot using visual feedback to drive image features toward a desired configuration [6]. In previous work by Cho et al. [8], it was demonstrated that IBVS for ship-deck landing of a quadrotor UAV with the controller driving the visual target image position toward the frame center, confirming the viability of vision-based feedback for precision UAV positioning. The approach of this thesis is to use a second stationary reference marker. Platform motion is canceled algebraically before the control signal is computed, making the feedback invariant to platform motion rather than subject to it. Visual targets for infrastructure inspection were studied by Lynch et al. [21], who pre-deployed targets at structural inspection points and used a nonlinear model predictive controller (NMPC) for autonomous UAV alignment, validating the premise that fiducial targets enable autonomous inspection positioning. The key difference is controller complexity. NMPC requires a full nonlinear model of the UAV dynamics and solves an optimization problem at each control step. This thesis achieves comparable alignment behavior using a model-free proportional controller, and further introduces a two-tag SE(3) relative pose formulation that algebraically cancels platform motion, a step not present in Lynch et al. Optical flow proximity control for wall inspection was studied by Lippiello and Siciliano [19] (Wall inspection control of a VTOL UAV), who demonstrated station-keeping at a defined standoff from a wall surface using optical flow from a forward-facing camera. This approach provides passive proximity maintenance but does not enforce alignment with a specific target point on the surface.

2.4 UAV-Mounted Cameras for Structural Monitoring

UAV platforms extend the spatial reach of vision-based SHM by providing mobile, elevated viewpoints that can access structures inaccessible to ground-based cameras. Several research groups have demonstrated natural frequency extraction from UAV video [36, 12, 34, 20, 4]. Yoon et al. [36] demonstrated an early UAV-based structural system identification approach using cross-correlation of responses measured simultaneously from a hovering UAV and a fixed reference sensor. Hoskere et al. [12] presented an end-to-end UAV vibration measurement framework, validating it on a shaking table, a shear building, and a full-scale suspension bridge, and addressed platform motion by incorporating scene reference points. Weng et al. [34] separated structural displacement from UAV platform motion using planar homography transformation with neural-network-guided feature selection. Luo et al. [20] proposed a time-varying homography computed from multiple calibration points to reduce signal drift in operational modal responses. Bolognini et al. [4] demonstrated that three UAVs hovering simultaneously at 2.5 m from a structure achieved natural frequency errors below 0.2% relative to accelerometers, and showed that surface references are not strictly necessary to obtain accurate displacement measurements. Kim and Kim [15] conducted a systematic reliability assessment of UAV-based displacement measurement using a commercial drone and homography-based ego-motion correction. Their conclusion explicitly identified the need for methods for removing the movement of UAVs by finding key points that can correct such movements on the same plane as the measurement target, as future work. The two-tag SE(3) relative pose method presented in this thesis is the direct answer to this open problem, generalized to full 3-D. The reference tag can be placed on any surface within the camera view, and the cancellation holds for arbitrary 3-D platform motion, not just planar homographic perturbations.

A key challenge common to all UAV-based measurement methods is separating structural motion from platform motion. A hovering multirotor is subject to translational drift, roll/pitch perturbations, yaw oscillations, and high-frequency rotor vibration. Approaches that measure absolute pose cannot distinguish between structural displacement and camera displacement. All platform motion appears as structural motion. The two-tag SE(3) method cancels UAV motion algebraically, requires no planar reference region, no neural network, and no inertial sensor fusion, and the cancellation is exact rather than approximate.

2.5 Laser Doppler Vibrometry and UAV-Mounted Sensing

Laser Doppler Vibrometry (LDV) is the established non-contact technology for precision structural vibration measurement, capable of nanometer-scaled displacement resolution, far exceeding the camera-based approach used in this thesis. Rembe et al. [28] conducted a comprehensive review of UAV-mounted LDV for large civil structures and identified UAV positioning, stability, laser beam pointing, and motion-induced measurement disturbance as the primary issues preventing practical UAV-LDV deployment. Specifically, the review identifies the inability of low-cost UAVs to maintain the pointing alignment required for LDV operation as the critical missing capability.

This thesis addresses the issues presented. The autonomous alignment and station-keeping framework demonstrated here provides the stable, pre-defined positioning that UAV-mounted LDV systems require. The platform separates alignment from measurement. The visual positioning system developed in this thesis is used exclusively for positioning feedback, without the need to be a measurement sensor. While this work uses the onboard camera as the measurement instrument, an LDV could be mounted alongside it to take measurements instead, with no changes required to the alignment controller.

2.6 Collaborative Work

This thesis directly advances collaborative work [3], a vibration measurement study accepted to the 2026 International Conference on Unmanned Aircraft Systems (ICUAS). That paper tested four platforms: a Samsung Galaxy S21 smartphone, an iPhone 15 Pro, a USB camera, and the same custom UAV used in this work, on the same scaled steel frame structure using Motion Tracker Beta software [9]. Motion Tracker Beta tracks a single circular marker via 2D pixel centroid detection, applying 2D pixel subtraction as its motion reference.

Smartphone results achieved frequency errors of 0–0.6% (healthy) and 0–0.2% (damaged), USB camera results were 1.6–1.8% (healthy) and 1.3–1.8% (damaged). The UAV result approximately yielded 5.5% error on both conditions. These results were attributed to rotor-induced vibration contaminating the single-tag absolute pose measurement, lower resolution, or constant drone drift and misalignment. The paper identified AprilTag-based autonomous positioning as explicit future work, which this thesis implements.

The current thesis replaces Motion Tracker Beta with the two-tag SE(3) relative pose system and adds the autonomous controller. The experimental structure, excitation method, and ground truth reference values are used in this thesis as a direct comparison between the two methods on identical hardware and structure to test the validity of the measurement system.

CHAPTER III

METHODOLOGY

3.1 Vision-Based Motion Tracking

Vision-based motion tracking is a non-contact optical technique used to measure structural displacements by analyzing image sequences captured during motion. In this approach, identifiable surface features are tracked between a reference image and subsequent frames to estimate displacement over time. In this work, the method is implemented using AprilTag fiducial markers [27] detected in each camera frame, from which the pose of a stationary reference tag and a vibrating target tag are estimated. The relative pose between these two tags is then used to recover structural displacement while canceling rigid-body motion of the UAV.

3.2 Hardware Development

The UAV was assembled from commercial off-the-shelf components at a total cost of approximately $1,000. The airframe is a DJI FlameWheel F450 quadrotor (450 mm diagonal), the flight controller is a Pixhawk running ArduPilot firmware in GUIDED mode, and a Raspberry Pi 4 Model B companion computer responsible for perception and high-level control. The flight stabilization is handled by the hierarchical flight controller (see Fig. 3.1), while visual perception and relative pose estimation are executed on the companion processor (see Fig. 3.2).

The flight controller serves as the stabilization unit using the onboard Inertial Measurement Unit (IMU), composed of accelerometers and gyroscopes, to estimate vehicle attitude and angular rates. These measurements are processed internally by ArduPilot's control loops to generate motor commands that regulate roll, pitch, yaw rate, and thrust.

The companion computer is a Raspberry Pi 4 Model B (4 GB RAM, 4× Cortex-A55 @ 1.8 GHz) running ROS 2 Jazzy [22] on Ubuntu 24.04. A forward-facing Intel RealSense D455 depth camera is mounted on the frame and operated in color-only mode at 640×480 resolution and 15 fps to remain within the Raspberry Pi 4 CPU budget. An MTF-01 optical flow sensor is

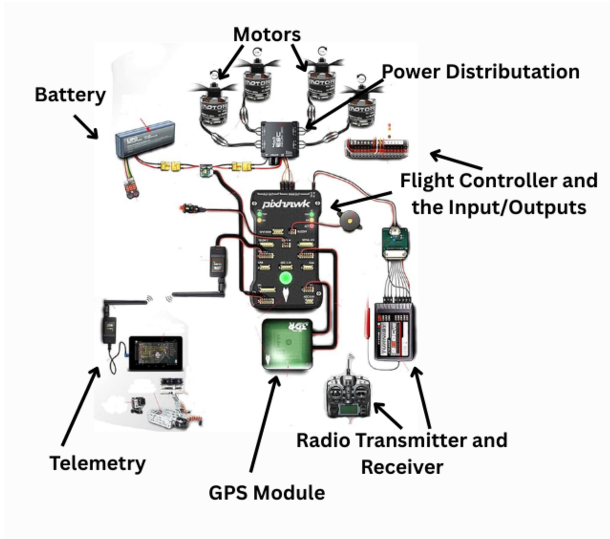


Figure 3.1: UAV system architecture.

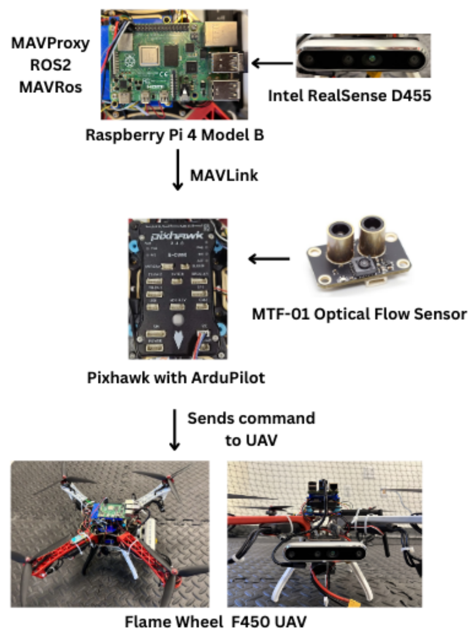


Figure 3.2: Visual perception architecture.

connected directly to the flight controller. It provides low-level ground-relative velocity feedback for hover stabilization, supplementing barometric altitude hold in GPS-denied operation(see Fig. 3.2). The hardware couples the RealSense D455 and Raspberry Pi 4 for real-time AprilTag-based pose estimation [27, 37] with the MTF-01 optical flow sensor for GPS-denied hover stabilization, enabling fully autonomous structural inspection positioning without GPS or manual pilot input.

In experimental trials, the forward-facing camera is used to acquire the vibration signal. The UAV detects AprilTags attached to the structure and wall and computes the relative pose between the camera and the visual markers [27, 37]. Once the UAV detects the AprilTags, it autonomously aligns to the target configuration and maintains the camera facing the structural frame for data acquisition. The camera frame rate of 15 fps was chosen as the lowest valid D455 profile that satisfies the Nyquist criterion for the 3–8 Hz structural frequency range of interest. The Nyquist limit at 15 fps is 7.5 Hz, covering both target structural modes with a minimum margin of over 45%.

The total hardware, as listed in (see Table 3.1), has a cost of approximately $1,000 at the time of assembly includes the DJI FlameWheel F450 frame ($400), the Intel RealSense D455 camera ($450), the Raspberry Pi 4 ($60), and the MTF-01 optical flow sensor ($30). For comparison, commercially available consumer drones, such as in [15], the DJI Mavic 3 series range from approximately $2,000 to $2,500, while higher-end platforms such as the DJI Matrice 210 in [10] retail around $10,000. Although commercial systems provide integrated stabilization and imaging hardware, they lack the level of extensibility and modularity required for research-oriented experimentation. The custom platform, therefore, provides a cost-effective and modular alternative that enables the SE(3) relative-pose software contributions of this thesis. In addition, the platform can be easily upgraded and adapted for deployment in other application domains, further enhancing its versatility and long-term utility.

Table 3.1: Component and Specification Table

Component	**Specification**
Airframe	DJI FlameWheel F450, 450 mm diagonal
Flight controller	Pixhawk, ArduPilot GUIDED
Companion computer	Raspberry Pi 4 Model B, 4 GB, ROS 2 Jazzy
Camera	Intel RealSense D455, 640×480 @ 15 fps
Optical flow	MTF-01
Tag family	AprilTag tag36h11
Tag size (physical)	165.1 mm (outer black border)
Standoff distance	$\sim$ 1.5 m
Approximate total cost	$\sim$$1,000 USD

3.3 Software Systems

The full ROS 2 stack running on the Raspberry Pi 4 consists of five nodes (see Table 3.2). The system is conceptually organized into three layers, stabilization, perception, and command. The stabilization layer is implemented onboard the Pixhawk and is responsible for low-level state estimation and control using the onboard inertial measurement unit and optical-flow-assisted velocity feedback. All motor commands are generated by the flight controller. The perception layer runs on the companion computer using ROS 2. This layer processes camera imagery, detects AprilTags, and computes relative pose measurements between the UAV and the visual reference [27, 37]. The command layer uses these pose estimates to generate velocity setpoints during autonomous alignment. Camera IMU streaming and frame synchronization were disabled to reduce CPU load on the Pi.

Table 3.2: ROS 2 stack nodes

Node	**Purpose**
MAVROS	MAVLink to ROS 2 bridge. Publishes odometry and accepts velocity setpoints
realsense2_camera_node	Intel D455 driver. Publishes color image stream
apriltag_ros	AprilTag detector. Outputs corner pixel coordinates per frame
apriltag_pnp_broadcaster	PnP pose estimator; publishes SE(3) transforms
hover_guided_hold	Autonomous state-machine flight controller

3.4 Autonomous Flight Controller

The autonomous flight controller is implemented as a four-state machine. The system consists of SEARCH, COARSE_ALIGN, SETTLE, and HOLD. In SEARCH, the UAV yaws in place at a fixed angular rate until both AprilTag markers are detected simultaneously in the camera frame. No translational motion is commanded during this phase. During COARSE_ALIGN a proportional controller drives the UAV toward the target configuration: centered laterally on the midpoint of the two tags, positioned at a forward standoff distance of approximately 1.5 m, and aligned in yaw with the target. Velocity setpoints are published to '/mavros/setpoint_velocity', and ArduPilot in GUIDED mode interprets these as target velocities for the inner-loop controller.

The use of vision-derived target error for UAV positioning is consistent with image-based visual servoing approaches reported in the literature [8, 21].

The outer-loop controller is modeled as a first-order integrator plant. The proportional gain k_p maps pose error to commanded velocity, yielding the closed-loop error dynamics with time constant $\tau = 1/k_p$.

$$e(t) = e_0 \exp(-k_p t) \tag{3.1}$$

Gains were selected to give a 95% settling time between approximately 6 and 10 seconds for the primary translational and yaw axes, while maintaining clear bandwidth separation from the faster flight-controller inner loop as summarized in Table 3.3 and shown in Fig. 3.3.

Table 3.3: Proportional controller gains and closed-loop settling characteristics per axis.

Axis	k_p	τ	**95% settling (s)**
Y	0.5–0.7	1.4–2.0	4–6
Forward (distance)	0.35	2.9	8.6
Lateral	0.35	2.9	8.6
Vertical	0.15	6.7	20

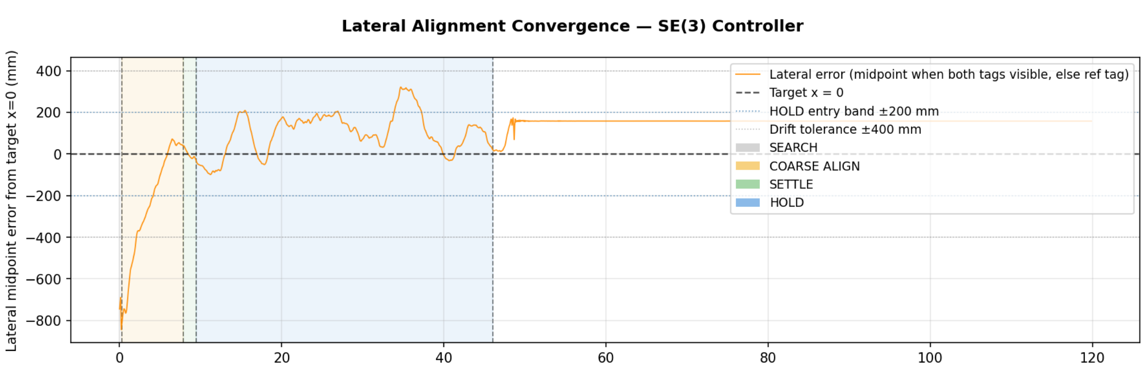


Figure 3.3: Convergence plot from a hardware trial showing lateral error for each state.

Velocity commands are clamped to safe maximums of 0.3 m/s in the horizontal plane, 0.2 m/s vertically, and 0.5 rad/s in yaw. Once all position errors fall within the threshold, the controller waits for a dwell period to allow transient oscillations to decay before leaving the SETTLE state and entering the HOLD state. The controller sends zero velocity commands, entering a passive measurement window. During this phase, the MTF-01 optical flow sensor provides ground-relative velocity feedback to the flight controller, reducing slow drift. The topic ‘/measurement_hold_active‘ is published ‘True‘ on the HOLD entry, providing a verifiable record that the autonomous controller reached measurement-ready state before data acquisition.

3.5 UAV Simulation

The simulation environment replicates the physical system architecture and was used to validate the UAV design before hardware testing. The simulation consists of ArduPilot Software-in-the-Loop (SITL) running alongside Gazebo [16] under ROS 2 Jazzy, connected through MAVROS. The implementation used in this thesis is available online [1]. The full software pipeline used in hardware for AprilTag detection, PnP pose estimation, and the autonomous flight controller runs identically in simulation, with only the camera source being different.

Controller gains were initially tuned in simulation(see Fig. 3.4) by observing settling time and alignment performance against the target conditions defined. State transitions were verified by placing AprilTag models at varying positions in the Gazebo world and confirming proper progression through SEARCH, COARSE_ALIGN, SETTLE, and HOLD. A separate ‘tag_oscillator‘ node, which moves an AprilTag sinusoidally at a known frequency, was used to validate the relative-pose measurement pipeline and PSD analysis before physical experiments.

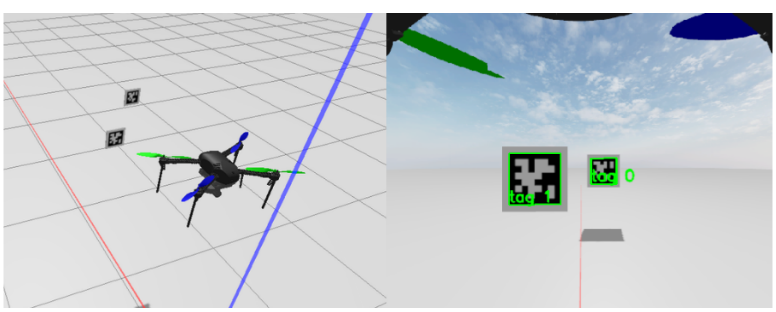


Figure 3.4: UAV simulation in Gazebo. The simulated UAV (left) and the detected AprilTags in the camera view (right).

[1] https://github.com/MARSLab-UTRGV/mars-drone-development

3.6 AprilTag Pose Estimation

Each video frame is processed by the AprilTag detector [27] using the ‘tag36h11‘ family [32]. The detector returns the four corner pixel coordinates of each observed tag, as shown in Fig. 3.8(b). For frames in which both tags are detected, OpenCV’s ‘solvePnP‘ is called separately for each tag with option ‘SOLVEPNP_ITERATIVE‘.

The PnP problem estimates the rotation matrix R and translation vector t that minimize the squared reprojection error between the known three-dimensional tag corner locations and the observed two-dimensional image points [11]. The result for each tag is a 4×4 rigid-body transformation matrix $\mathbf{T}_{\text{tag}}^{\text{cam}} \in \mathrm{SE}(3)$, expressing the tag pose in the camera frame [11, 25]. The physical tag size used for this computation is 165.1 mm measured across the outer black border. Frames with reprojection error greater than 8 pixels are rejected. The ‘SOLVEPNP_IPPE_SQUARE‘ solver was evaluated but found to occasionally return geometrically invalid poses without raising an exception. By contrast, ‘SOLVEPNP_ITERATIVE‘ produced stable solutions and a validated reprojection error of approximately 0.26 pixels on representative test frames. Intel RealSense D455 factory intrinsics are loaded from a calibration YAML file corresponding to the 640×480 recording resolution.

3.7 Two-Tag Relative Pose and UAV Motion Cancellation

For each frame where both tags are successfully detected, the relative pose between the vibrating target tag and the stationary reference tag is computed as:

$$\mathbf{T}_{\text{rel}} = (\mathbf{T}_{\text{ref}}^{\text{cam}})^{-1}\,(\mathbf{T}_{\text{vib}}^{\text{cam}}) \tag{3.2}$$

This transformation expresses the target tag pose in the coordinate frame of the reference tag. The lateral translation component $x(t) = T_{\text{rel}}[0,3]$ is taken as the structural displacement time history in meters. This two-tag SE(3) formulation cancels UAV rigid-body motion algebraically. If the UAV undergoes any rigid motion M between frames, then both tag poses in the camera frame are transformed by the same motion.

$$(\mathbf{M}\,\mathbf{T}^{\text{cam}}_{\text{ref}})^{-1}(\mathbf{M}\,\mathbf{T}^{\text{cam}}_{\text{vib}}) = (\mathbf{T}^{\text{cam}}_{\text{ref}})^{-1}\mathbf{M}^{-1}\mathbf{M}\,\mathbf{T}^{\text{cam}}_{\text{vib}} \qquad = (\mathbf{T}^{\text{cam}}_{\text{ref}})^{-1}\mathbf{T}^{\text{cam}}_{\text{vib}} \tag{3.3}$$

The UAV motion term cancels exactly. This result holds for any rigid-body translation, rotation, or combined motion and does not depend on a filter or approximate compensation step. The cancellation is a direct geometric property of the relative transform [11]. Relative AprilTag pose estimation for UAV applications has also been demonstrated in prior work [37].

3.8 Signal Processing with Welch PSD

The relative displacement time series $x(t)$ is sampled at approximately 15 Hz, with occasional gaps when one or both tags are not detected. The dominant structural natural frequency is identified using Welch power spectral density estimation [33]. The signal is divided into overlapping segments of length $N/4$ samples with $N/8$ overlap, a Hann window is applied to each segment, the discrete Fourier transform is computed, and the resulting squared magnitudes are averaged [33]. The dominant peak in the 3–8 Hz structural frequency band is identified as the natural frequency. No numerical differentiation is applied, since differentiating the displacement signal at 15 fps would amplify high-frequency noise by a factor proportional to $1/\Delta t^2$ [14, 26]. For the approximately 12-second clips used in these experiments, the frequency resolution is $\Delta f \approx 0.333$ Hz. The analysis axis is fixed to the lateral x-direction on physical grounds because lateral bending is the first mode of the cantilever frame and the excitation direction is structurally prescribed.

3.9 Experimental Setup

Two distinct excitation methods were used across the experiments. The stationary camera platforms were tested under controlled harmonic excitation using an electrodynamic shaker, which allows precise frequency control and repeatable loading conditions. The UAV experiments used free-vibration excitation initiated by a manual tap due to flight safety constraints and logistical limitations. However, both experiments lead to the same result since the natural frequency of the structure remains constant. Both experiments used the same steel frame structure and the same 240 g proof mass to simulate the damaged condition. The experimental structure is a scaled one-story steel frame consisting of four stainless-steel columns (18 in tall, 1.5 in wide, 1/16 in thick) supporting an aluminum roof diaphragm (12 × 6 × 1/8 in) as shown in Fig. 3.5. The structure and UAV hardware are the same as those used in [3], enabling direct comparison between the two-tag SE(3) method and the earlier single-tag Motion Tracker Beta approach on identical hardware and structure. Ground-truth reference values are the contact accelerometer frequencies reported in [3]. 5.09 Hz for the healthy condition and 4.51 Hz for the damaged condition.

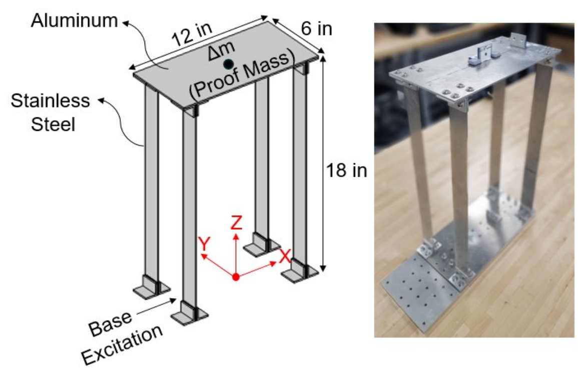


Figure 3.5: Measurements of the scaled one-story frame building for testing.

3.9.1 Forced Vibration Setup

Three stationary camera platforms were evaluated on this structure under controlled harmonic excitation, as shown in Fig. 3.6. Two PCB accelerometers were mounted at the roof and base of the frame to record the dynamic response, serving as the ground truth reference for frequency identification.

Three camera platforms were positioned at a stand-off distance of approximately 4 ft from the frame (see Fig. 3.7) a Samsung Galaxy S21 Ultra 5G smartphone, an iPhone 15 Pro smartphone, and an IEights USB camera. Each platform recorded at multiple configurations. The smartphones recorded at UHD (3840 × 2160) resolution at both 30 fps and 60 fps. The USB camera recorded at 1080p/60 fps and 720p/120 fps. A large black circle on a white background was fixed to the center of the roof diaphragm as the optical tracking target.

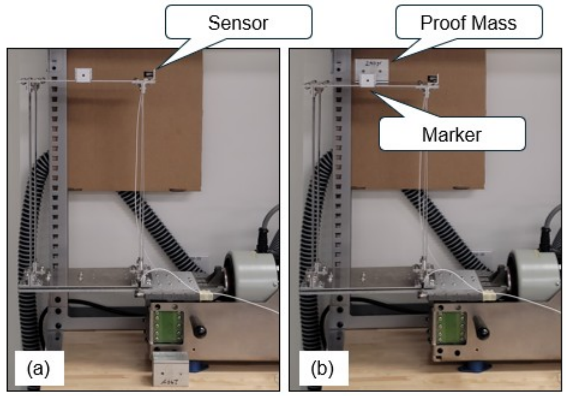


Figure 3.6: Frame building testing: (a) healthy and (b) damaged.

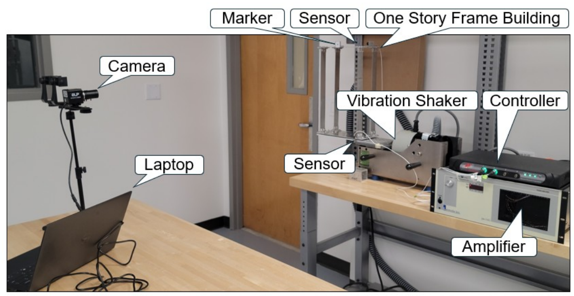


Figure 3.7: Forced vibration experimental setup for testing non-contact camera methods.

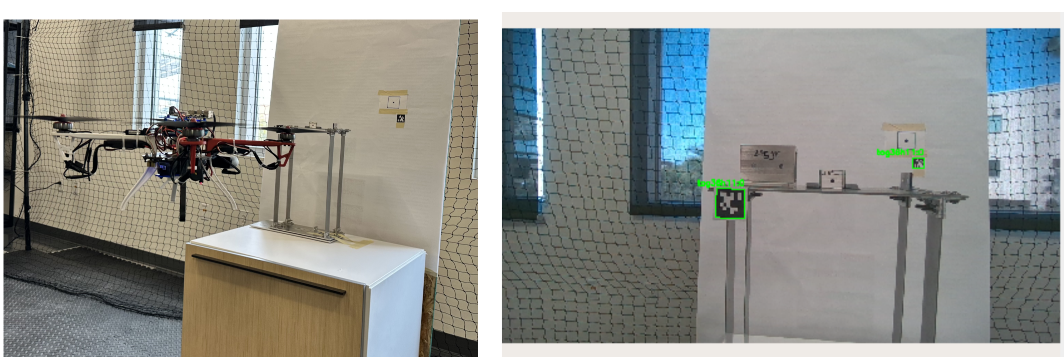
Displacement time histories were extracted from each video using Motion Tracker Beta [9], an open-source image-based tracking tool that tracks the pixel centroid of the circular target across frames. The extracted displacements were analyzed in the frequency domain to identify the fundamental natural frequency under healthy and simulated damage conditions.

3.9.2 Free Vibration Setup

The UAV free vibration experiment evaluated the system's ability to identify structural natural frequencies for damage detection under autonomous flight conditions. The autonomous alignment controller positioned the UAV at the defined standoff distance before each measurement window, providing the stable imaging conditions required for vibration signal extraction.

AprilTag ID=0 (165.1 mm, tag36h11) was fixed to the background wall and remained stationary throughout the tests. AprilTag ID=1 of the same size and family was attached to the roof diaphragm and moved with the vibrating structure (see Fig. 3.8). A 240 g proof mass was attached to the center of the roof diaphragm to represent the damaged condition. The added-mass method is an established structural health monitoring laboratory technique for producing reproducible and reversible frequency shifts without permanently modifying the structure [29]. It replicates the dynamic effect of mass redistribution, such as corrosion-driven deposition or ice accumulation, but does not replicate stiffness-degradation damage, such as cracking or connection loosening. Accordingly, the conclusions of this validation apply to mass-change scenarios, and generalization to stiffness-degradation damage requires additional experimentation.

The structure was excited by a free-hand tap-and-release at the roof diaphragm edge, initiating free-decay vibration. Three independent tap events were analyzed for each condition. Video clips were trimmed to an approximately 12-second window containing each decay event. Free-vibration responses decay over time, reducing the effective signal window available for frequency estimation and contributing to variance in the Welch PSD estimate. All data collection was managed by the autonomous controller. Before each measurement window, the UAV transitioned through SEARCH, COARSE_ALIGN, SETTLE, and HOLD. The testing environment can be observed in Fig. 3.8. The '/measurement_hold_active' topic recorded in the ROS bags confirms that HOLD was reached for every analyzed trial.



(a) (b)

Figure 3.8: Experimental setup. (a) The UAV is collecting data; (b) The UAV's camera detects two AprilTags.

CHAPTER IV

RESULTS

4.1 Controller Alignment and Station-Keeping Results

Two complementary analysis characterize the autonomous controller's performance. One is a convergence analysis that measures alignment time from a cold start, and the other is a stability comparison that quantifies inspection positioning accuracy during the HOLD window.

4.1.1 Convergence Analysis

The convergence analysis used the full 2-minute recording of an autonomous hardware flight (640×480 @ 15 fps, 1799 frames). Both Tag 0 (reference) and Tag 1 (inspection target) were detected simultaneously in 1648 of 1650 sampled frames, a 99.9% detection rate. The midpoint of the two tag positions in the image was used as the controller's centering target. State transitions were annotated from the hardware terminal log and confirmed by comparing the offline PnP analysis against the onboard estimates.

The controller reduced lateral alignment error from −819 mm to ±36 mm in 9.5 seconds, 96% reduction, fully autonomously from SEARCH mode to HOLD. 7.6 seconds were spent during the alignment of the drone and once alignment was successful, the controller spent 1.6 seconds settling and moving into a HOLD state (see Table 4.1). Times are elapsed seconds from the start of the recording. Phase durations shown. COARSE_ALIGN is an active proportional control. SETTLE is a fixed 1.6 s velocity-zeroing buffer before HOLD entry.

Table 4.1: Convergence of UAV State Transition

State	Time (s)	Lateral error	Depth (ref_z)
SEARCH to COARSE_ALIGN	0.26	-819 mm	1.54 m
COARSE_ALIGN to SETTLE	7.62	+38 mm	1.62 m
SETTLE to HOLD	1.58	-36 mm	1.52 m
HOLD end (landing)	36.6	+18 mm	1.80 m

During the HOLD window (t = 9.46 s to t = 46.06 s, 549 frames), the drone's lateral position remained within ±400 mm from the center (with most time being spent under 200 mm) and the standoff depth remained within ±800 mm of the 1.5 m target (see Table 4.2. Both values are within the controller's defined drift tolerances, which were chosen to prevent the tags from leaving the camera field of view during the passive hold.

Table 4.2: Location-Lock Performance During HOLD Phase

HOLD metric	**Value**	**Drift tolerance**
Lateral RMSE from center	141.6 mm	±400 mm
Lateral mean offset	±102.9 mm	—
Depth RMSE from 1.5 m	265 mm	±800 mm
Depth mean	1.747 m	—

HOLD is passive by design. The controller sends zero velocity commands during measurement to avoid disturbing the vibrating structure. All HOLD drift comes from the MTF-01 optical flow sensor alone.

4.1.2 Stability Comparison

To characterize the inspection quality improvement provided by the controller, three trimmed 20-second clips at 640×480 @ 30 fps were compared: two uncontrolled hover trials (optical flow only) and one controller trial clipped to the HOLD window.

The primary figure of importance is the pixel RMS from the frame center and the horizontal pixel offset from the camera frame center, which is how close the inspection target is to the center of the camera frame. This directly answers whether the controller places the drone in the correct position for close-range inspection, independent of jitter. Standard deviation of position (σ_x) is not used as the primary metric here because HOLD is passive. The drone goes silent during measurement, so jitter during HOLD reflects the optical flow hardware, not the controller. A centered but slightly noisy drone is preferable to a stable but misaligned one for inspection purposes. The SE(3) controller maintained the inspection target at 175 px RMS from the camera frame center, compared to 325–327 px for uncontrolled optical-flow hover (see Table 4.3 and Fig. 4.1), a 1.85× improvement. The jitter metric (σ_x) shows comparable performance between conditions, which is expected since the passive HOLD relies on the same optical flow sensor used in the uncontrolled case. The controller's value is alignment before measurement, not reduced jitter during it.

Table 4.3: Stability Comparison: SE(3) Controller vs. Uncontrolled Hover

Metric	**No Controller (1)**	**No Controller (2)**	**SE(3) Controller**
3D position stability			
σ_x lateral (mm)	71.8	126.4	86.2
σ_y vertical (mm)	21.2	18.8	31.7
σ_z depth (mm)	46.5	60.0	68.3
RMS 3D (mm)	88.1	141.2	114.5
Image-plane centration			
Pixel RMS from center (px)	**327**	**325**	**175**
Pixel mean dist from center (px)	324	318	172
Pixel max dist from center (px)	393	482	256
AprilTag detection rate	100%	100%	100%

These alignment results were achieved autonomously, without the use of a GPS or pilot input on a platform assembled at approximately $1,000. Commercial inspection UAVs used in comparable close-range structural monitoring work, such as the DJI Matrice 210 [10], retail at approximately $10,000, and the DJI Mavic 3 series [15] ranges from $2,000 to $2,500. Both require GPS for autonomous operation and cannot perform indoor positioning without additional external infrastructure. The custom platform presented here demonstrates autonomous indoor alignment at approximately one-tenth the cost of high-end commercial inspection UAVs.

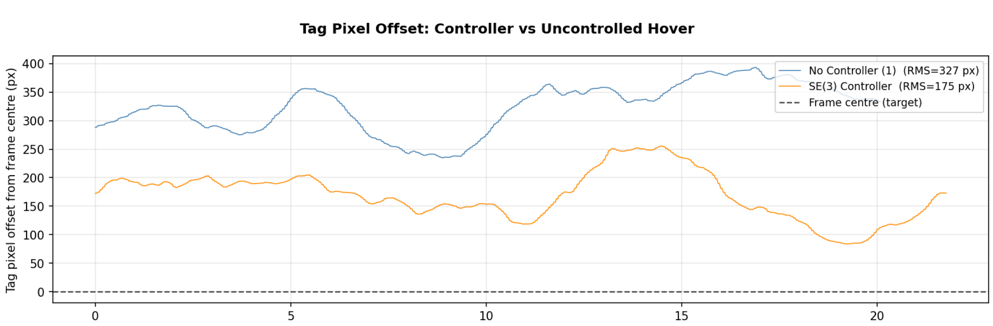


Figure 4.1: Stability comparison: pixel offset from frame center.

4.2 Vibration Frequency Identification

The vibration measurement experiment serves as a bounding precision test. It demonstrates that the platform's alignment accuracy is sufficient to resolve small structural displacements using only the onboard camera, a conservative sensor choice. If the system can correctly identify a structural natural frequency from camera-derived displacement at a 1.5 m distance, it is capable of supporting focused sensors (laser Doppler vibrometers, ultrasonic transducers) at the same or closer distances.

Six independent trials were conducted across healthy and damaged structural conditions. All trials were recorded with the drone in HOLD state after completing the full controller sequence. Natural frequency was identified from the Welch PSD of the lateral (X-axis) relative displacement time series. Ground truth reference values are from contact accelerometers reported in [3]: 5.09 Hz (healthy), 4.51 Hz (damaged).

Contact accelerometers mounted at the roof and base of the frame serve as the control reference for all frequency comparisons. Fig. 4.2 shows the acceleration time histories and corresponding PSDs for both the healthy and damaged frames. The healthy frame produces a dominant peak at 5.09 Hz, and the addition of the 240 g mass shifts this peak to 4.51 Hz.

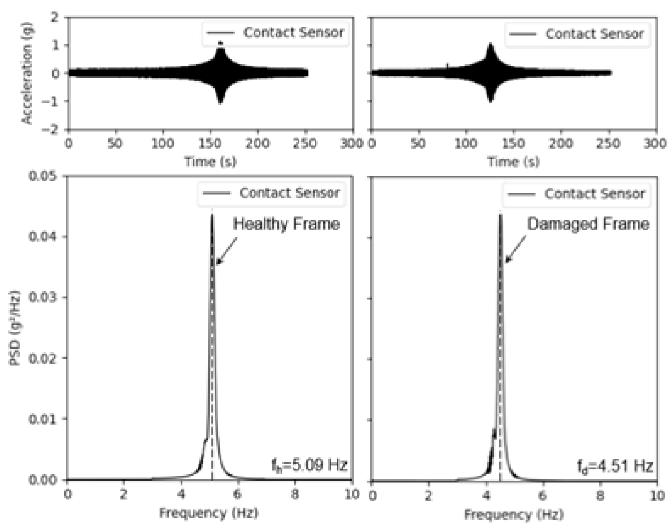


Figure 4.2: Time history of roof accelerations and corresponding PSDs measured by contact accelerometers for the healthy (5.09 Hz) and damaged (4.51 Hz) frame [3].

All three healthy trials identify a clear dominant spectral peak in the 5.0 to 5.3 Hz range(example can be seen in Table 4.4 and Fig. 4.3). The mean identified frequency is 5.131 Hz, yielding a mean error of 0.8% against the 5.09 Hz ground truth.

Table 4.4: Healthy Structure Vibration Analysis (No Added Mass)

Trial	**Analysis window**	**Detection rate**	**Identified frequency**	**Error vs 5.09 Hz**
Test 1	28–40 s	92.8%	5.136 Hz	0.9%
Test 2	41–53 s	95%	5.247 Hz	3.07%
Test 3	54–66 s	100%	5.013 Hz	1.51%
Mean	-	-	5.131 Hz	0.8%

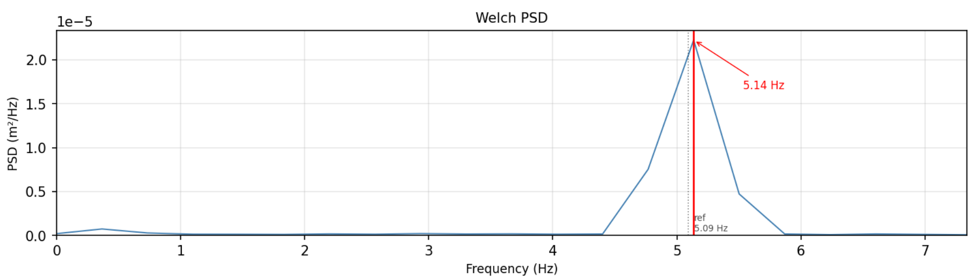


Figure 4.3: PSD analysis of healthy structure.

The added mass reduces the structure's lateral vibration amplitude. Despite the weaker signal, all three trials return frequencies in the 4.3–4.7 Hz range(as seen in Table 4.5 and Fig. 4.4), consistent with the 4.51 Hz ground truth. The mean identified frequency is 4.547 Hz, yielding an error of 0.82%.

Table 4.5: Damaged Structure Vibration Analysis (240 g Proof Mass)

Trial	**Analysis window**	**Detection rate**	**Identified frequency**	**Error vs 4.51 Hz**
Test 1	10–23 s	100%	4.65 Hz	3.1%
Test 2	24–37 s	100%	4.34 Hz	3.77%
Test 3	38–50 s	87%	4.65 Hz	3.1%
Mean	-	-	4.547 Hz	0.82%

Fig. 4.5 shows acceleration time histories and corresponding PSDs obtained from UAV camera video processed through vision-based motion tracking [3]. Although the time histories

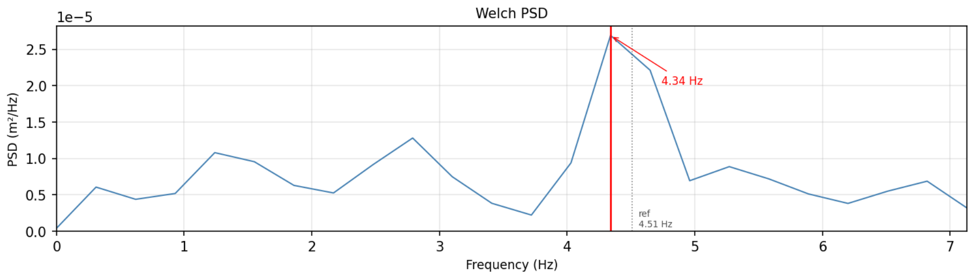


Figure 4.4: PSD analysis of damaged structure.

from the UAV camera differ in amplitude from the contact sensor due to camera resolution and platform-induced vibration, the frequency analysis reliably captures the dominant structural peaks under both healthy and damaged conditions. This confirms that the UAV successfully identifies the damage-induced frequency shift.

Both conditions achieve below 1% mean frequency error (see Table 4.6. The prior-work column represents the same UAV hardware tested in [3] using Motion Tracker Beta software with single-marker 2D tracking. The hardware, structure, and accelerometer ground reading frequency are identical; only the measurement method differs. The two-tag SE(3) relative pose approach achieves significant improvement.

The frequency shift between healthy (5.131 Hz) and damaged (4.547 Hz) conditions is 11.4%, matching the shift of 11.4% observed in the accelerometer frequency results (from 5.09 Hz to 4.51 Hz). The frequency error (0.8–0.82%) is an order of magnitude smaller than the detectable frequency shift, providing a detection margin well above the measurement uncertainty. Structural damage was successfully detected in all six, confirming that the autonomous alignment framework positions the UAV accurately enough for onboard sensors to resolve damage-induced structural changes.

The forced vibration experiments using stationary cameras produced frequency identification errors below 2% of the contact accelerometer reference values. Both smartphones achieved errors of 0 to 0.6% under healthy conditions and 0 to 0.2% under damaged conditions, due to their high resolution and stable fixed mounting. The USB camera produced slightly higher errors of 1.3 to 1.8%, attributable to its lower resolution limiting displacement tracking accuracy. Despite this difference, all stationary platforms successfully detected the damage-induced frequency shift from 5.09 Hz to 4.51 Hz.

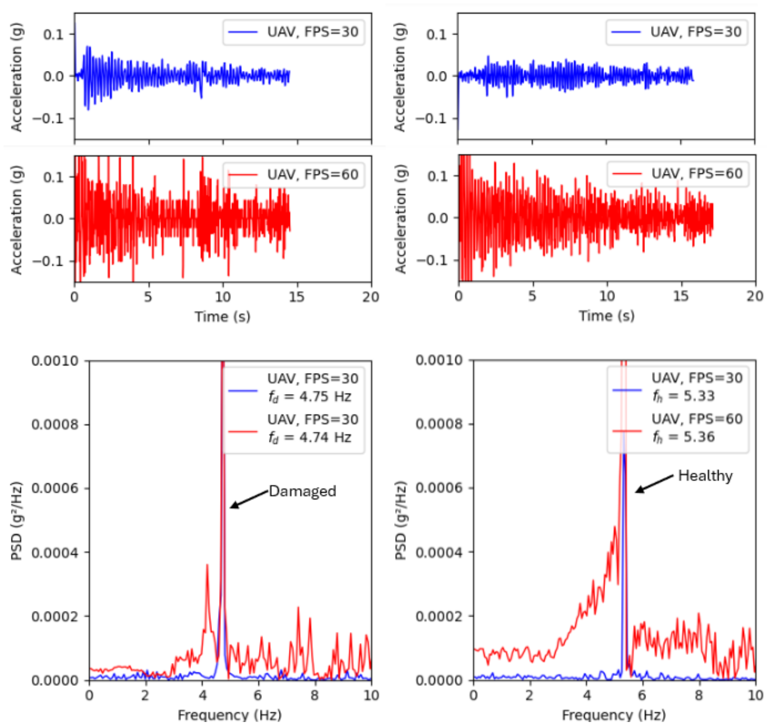


Figure 4.5: Time history of roof accelerations and corresponding PSDs measured by the UAV camera under healthy and damaged conditions.

Table 4.6: Structure Vibration Analysis Summary and Comparison to Prior Work

Condition	Mean identified freq	Ground truth	Error	Same UAV, no AprilTags
Healthy (no mass)	5.131 Hz	5.09 Hz	0.8%	∼5.5%
Damaged (240 g)	4.547 Hz	4.51 Hz	0.82%	∼5.5%

Table 4.7 places the SE(3) result in context of the full multi-platform study from previous collaborative work [3], conducted on the same steel frame with the same contact accelerometer references. The FE model (5.20 Hz) and contact sensor (5.09 Hz healthy, 4.51 Hz damaged) rows provide the structural reference values established by the Civil Engineering team. The forced vibration rows show results from stationary cameras during shaker excitation. Smartphones achieved the lowest errors (0.0 to 0.6%) due to their high spatial resolution and fixed mounting, while the USB camera followed at 1.3 to 1.8%. The free vibration rows show UAV results with-

out shaker excitation. The single-marker UAV yielded 5.1 to 5.7% error due to platform-induced motion contaminating the measurement signal. The two-tag SE(3) method reduces UAV error to a mean of 0.8% and 0.82%, a greater than six-fold improvement on identical hardware, bringing the autonomous UAV into the same accuracy range as the stationary USB camera. All platforms across both vibration conditions successfully detected the damage-induced frequency shift.

Table 4.7: Frequency and Error Comparison Across All Platforms

	Platform	f_h **(Hz)**	ε_h **(%)**	f_d **(Hz)**	ε_d **(%)**
FE Model		5.20	—	—	—
Contact Sensor		5.09	—	4.51	—
Forced Vibration	Smartphone (Samsung)	5.10	0.2	4.51	0.0
		5.12	0.6	4.52	0.2
	Smartphone (iPhone)	5.10	0.2	4.51	0.0
		5.10	0.2	4.52	0.2
	USB Camera	5.01	1.6	4.43	1.8
		5.00	1.8	4.45	1.3
Free Vibration	UAV, Single-Marker	5.38	5.7	4.75	5.3
		5.35	5.1	4.76	5.5
	UAV, SE(3) Two-Tag	5.136	0.90	4.650	3.10
		5.247	3.07	4.340	3.77
		5.013	1.51	4.650	3.10
	Mean	**5.131**	**0.80**	**4.547**	**0.82**

4.3 Controller Operation Verification

Autonomous operation during the vibration trials was verified through two independent records. Both ROS bags contain 12 published messages on the /measurement_hold_active topic, which is published exclusively by hover_guided_hold on HOLD entry and exit. The presence of this topic in every bag confirms that the autonomous controller reached measurement-ready state before each analyzed trial, no trial was processed from an uncontrolled hover.

CHAPTER V

DISCUSSION

5.1 Platform Positioning Accuracy

The convergence results establish that the autonomous controller achieves useful inspection positioning from arbitrary starting conditions. A 96% reduction in lateral alignment error (±819 mm to ±36 mm) in under 10 seconds, using only the onboard AprilTag detections and with no GPS or inertial navigation. The time-constant analysis predicted a settlement time of 96% of 6 to 9 seconds for the horizontal axes. The observed COARSE_ALIGN duration of 7.6 seconds falls within this range, which is when the controller directly contributes to the alignment of the drone. The additional 1.6 seconds of SETTLE is a fixed velocity, where it is independent of the controller.

The controller places the inspection target 1.85× closer to the camera frame center compared to uncontrolled hover (175 px vs. 325–327 px RMS). This matters because the camera center is where any inspection sensor mounted on the drone is pointed. If the target tag appears far from the frame center, the drone is not aimed at the inspection point, and whichever sensor is onboard will not be measuring the right location.

At 327 px from the center of a 640 px wide image, the target is sitting near the edge of the frame. A laser Doppler vibrometer, for example, requires its beam to land on a specific measurement point. If the drone is this far off-center, the beam misses. At 175 px, the target is within the central region of the frame, and the drone is close enough to the correct position for a mounted sensor to collect valid data. In addition, as the distance from the target increases, this distance from the center of the frame will yield a greater distance in actuality.

The HOLD phase results show that the platform maintains the target within the drift tolerances over a 36-second passive hold, demonstrating operational robustness for a measurement window of practical duration. The mean lateral offset of +103 mm during HOLD reflects a small steady-state bias from optical flow, which is expected. Optical flow stabilizes ground-relative velocity to zero but cannot enforce a specific absolute position. This offset does not affect the relative SE(3) measurement, which cancels platform motion algebraically.

5.2 Sensor-Agnostic Inspection Platform

The primary contribution of this thesis is the autonomous alignment and hold framework. The two-tag SE(3) relative pose formulation makes the platform sensor-agnostic. Once the controller has aligned the drone to the inspection point and entered HOLD, any inspection sensor mounted on the drone is in a stable, centered position relative to the target, allowing it to use onboard specialized sensors to record data. The vibration experiment validates this indirectly. The onboard camera is among the least sensitive sensors that could plausibly be used for structural measurement, yet it achieves sub-1% frequency error. Any sensor with better precision will perform at least as well once alignment is achieved.

UAV-mounted LDV is an active research direction for non-contact structural vibration measurement of large civil structures [28]. Rembe et al. identify UAV positioning stability, laser beam pointing, and motion-induced measurement disturbance as the primary open research challenges, with autonomous precise alignment at a defined standoff explicitly identified as an unsolved requirement. The alignment framework presented here directly addresses that problem: pre-deploy AprilTag markers at the measurement points, and the UAV autonomously aligns and holds without GPS or a pilot.

The experimental validation presented in this thesis is the product of an interdisciplinary collaboration between the computer science and civil engineering departments at UTRGV. The scaled steel frame structure, the finite element model developed in COMSOL Multiphysics, and the contact accelerometer measurements used as ground-truth frequency references were contributed by the UTRGV Civil Engineering team. Their structural fabrication and instrumentation work established the validated reference frequencies for the healthy condition and damaged condition, against which the UAV measurement errors of 0.8% and 0.82% are reported.

The accuracy of the SE(3) method is further contextualized by the multi-platform study in the previous work [3], conducted on the same scaled steel frame with the same contact accelerometer references. That study evaluated two high-resolution smartphones, a USB camera, and the same UAV hardware used in this thesis, all measuring the same healthy and damaged structural conditions. The SE(3) method developed in this thesis reduces UAV error to 0.8% and 0.82%, placing the UAV on par with the USB camera results and within a factor of two of the high-resolution smartphones. This improvement is achieved on identical hardware, confirming that the dominant source of error in the earlier UAV result was the inability to separate platform motion from structural displacement. The cost-effective UAV platform, assembled at approximately $1,000, thus achieves measurement accuracy comparable to stationary camera setups while adding the capability to autonomously reach and hold at inaccessible inspection points.

CHAPTER VI

CONCLUSION

This thesis validated a low-cost autonomous UAV platform for non-contact vibration-based structural damage detection. The system autonomously aligns itself to a structural inspection target and holds position at a defined distance for close-range inspection without the use of GPS or manual intervention, and identifies natural frequency shifts indicative of structural damage from onboard camera video. The system is built from commercial off-the-shelf components with an approximate cost of $1000 and runs entirely on open-source software (ROS 2, ArduPilot, OpenCV). This was achieved through three contributions. An autonomous alignment and station-keeping controller, a vision-based relative pose estimation method that cancels UAV rigid-body motion from the measurement signal, and experimental validation on a laboratory steel frame under healthy and simulated damage conditions. The cancellation holds for any combination of translation, rotation, or vibration the drone undergoes. This approach directly answers the problem identified by Kim & Kim [15], who called for a method to remove UAV motion using reference points on the same plane as the measurement target.

Two experiments validate the system. In the station-keeping test, with both tags mounted stationary, the controller achieved a 96% reduction in lateral alignment error (-819 mm to 36 mm) in under 10 seconds and maintained the target 1.85× closer to the camera center compared to uncontrolled hover (177 px vs. 327 px RMS). In the vibration precision test, applied to a scaled steel frame under healthy (5.09 Hz) and damaged (4.51 Hz) conditions, the system identified natural frequencies with mean errors of 0.8% and 0.82%. This is a greater than six-fold improvement over the 5.5% error from single-marker tracking on the same hardware [3]. For broader context, the multi-platform study conducted in collaboration with this work [3] tested the same structure with smartphones (0-0.6% error), a USB camera (1.3-1.8%), and the same UAV hardware using single-marker tracking (5.1-5.7%). These results confirm that non-contact vision-based platforms across a range of costs and mobility can reliably detect damage-induced natural frequency shifts. This brings UAV accuracy in line with stationary camera systems at the same hardware cost. The system successfully detected the damage-induced frequency shift in all six trials. For SHM, it is critical to detect substantial shifts in natural frequency rather than just measuring a change in amplitudes, and this platform does so, confirming that an $1,000 autonomous UAV can achieve reliable inspection that a commercial system can achieve at a fraction of the cost.

This study was conducted in an indoor environment with simulated mass-redistribution damage. This does not truly replicate real outdoor environments and variables such as wind disturbance or cracking damage. Real structural vibration amplitudes are typically sub-millimeter, which may challenge camera-based systems and promote the application of mounted specialized sensors. Technical improvements include developing a closed-loop hold controller to further reduce drift, higher-end recording and processing via upgraded Raspberry Pi 5, expanded sensor capabilities, and multi-point inspection routing. The most critical next step for deployment is outdoor evaluation by simulating wind disturbance in a controlled environment to assess the stability and accuracy of the system. The platform also allows for the future development and deployment of multiple cooperative UAVs to improve the inspection coverage and efficiency on large structures, providing the foundation for scalable swarm UAV structural health monitoring.

This work represents a step toward a scalable, low-cost autonomous UAV inspection system. The pose estimation method used to determine shifts in natural frequency is applicable to any camera platform. As UAV hardware and UAV collaborative communication are improved, a multi-UAV system built on this framework could enable comprehensive, non-contact inspection of large civil structures, delivering the reliability and coverage of commercial inspection services while maintaining a significantly lower cost.

REFERENCES

[1] Mohammad Abedin and Armin B. Mehrabi. Health monitoring of steel box girder bridges using non-contact sensors. *Structures*, 34:4012–4024, 2021.

[2] Mohsen Amjadian, Quoc-Bao Ta, Blanca Capistran, Qi Lu, Gasser Ali, and Constantine Tarawneh. Experimental study on damage detection of civil structures using digital image correlation (DIC) method. In *ASCE CI and CRC Joint Conference*, March 2026.

[3] Javier Becerril, M. Vargas, J. Herrera, J. Gutierrez, J. Rios, M. Amjadian, C. Tarawneh, J. Yang, and Q. Lu. Non-contact vibration-based damage detection of civil structures using a cost-effective autonomous UAV. In *2026 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2026.

[4] Michele Bolognini, Giovanni Izzo, Daniele Marchisotti, Lorenzo Fagiano, Maria Pina Limongelli, and Emanuele Zappa. Vision-based modal analysis of built environment structures with multiple drones. *Automation in Construction*, 143:104550, 2022.

[5] James M. W. Brownjohn, Alessandro De Stefano, You-Lin Xu, Helmut Wenzel, and A. Emin Aktan. Vibration-based monitoring of civil infrastructure: challenges and successes. *Journal of Civil Structural Health Monitoring*, 1(3–4):79–95, 2011.

[6] Francois Chaumette and Seth Hutchinson. Visual servo control. i. basic approaches. *IEEE Robotics & Automation Magazine*, 13(4):82–90, 2006.

[7] Yuansheng Cheng, Zhe Tian, Donghong Ning, Ke Feng, Zhixiong Li, Sumika Chauhan, and Govind Vashishtha. Computer vision-based non-contact structural vibration measurement: Methods, challenges and opportunities. *Measurement*, 243:116426, 2025.

[8] Gangik Cho, Joonwon Choi, Geunsik Bae, and Hyondong Oh. Autonomous ship deck landing of a quadrotor UAV using feed-forward image-based visual servoing. *Aerospace Science and Technology*, 130:107869, 2022.

[9] Kristóf Floch and Attila Kossa. Motion Tracker Beta: A GUI based open-source motion tracking application. *SoftwareX*, 23:101424, 2023.

[10] Habeene Habeenzu, Patrick McGetrick, Su Taylor, and David Hester. UAS-based bridge displacement measurement using two cameras with non-overlapping fields of view. *Automation in Construction*, 167:105687, 2024.

[11] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2 edition, 2004.

[12] Vedhus Hoskere, Jong-Woong Park, Hyungchul Yoon, and Billie F. Spencer. Vision-based modal survey of civil infrastructure using unmanned aerial vehicles. *Journal of Structural Engineering*, 145(7):04019062, 2019.

[13] Michail Kalaitzakis, Brennan Cain, Sabrina Carroll, Anand Ambrosi, Camden Whitehead, and Nikolaos Vitzilaios. Fiducial markers for pose estimation. *Journal of Intelligent & Robotic Systems*, 101(4):71, 2021.

[14] Maksat Kalybek, Mateusz Bocian, and Nikolaos Nikitas. Performance of optical structural vibration monitoring systems in experimental modal analysis. *Sensors*, 21(4):1239, 2021.

[15] Hongjin Kim and Guyeon Kim. Reliability assessment of a vision-based dynamic displacement measurement system using an unmanned aerial vehicle. *Sensors*, 23(6):3232, 2023.

[16] N. Koenig and A. Howard. Design and use paradigms for Gazebo, an open-source multi-robot simulator. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2149 – 2154 vol.3, 04 2004.

[17] Maria Pina Limongelli. Vibration-based structural health monitoring: Challenges and opportunities. In Alphose Zingoni, editor, *Advances in Engineering Materials, Structures and Systems*, pages 1999–2004. Taylor & Francis / CRC Press, 2019.

[18] Yuankun Lin, Pinbo Huang, Zihao Ni, Shengli Xie, Yulei Bai, and Bo Dong. Full-field vibration measurements by using high-speed two-dimensional digital image correlation. *Applied Sciences*, 13(7), 2023.

[19] Vincenzo Lippiello and Bruno Siciliano. Wall inspection control of a VTOL unmanned aerial vehicle based on a stereo optical flow. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4296–4302, 2012.

[20] Jun Luo, Kaisen Tang, Yuan Hu, Yongli Zhong, Xinpeng Liu, and Zhitao Yan. UAV-based operational modal analysis method using improved homography-based perspective rectification method. *Journal of Vibration and Control*, 30(7–8):1829–1840, 2024.

[21] Angus Lynch, Joao Buzzatto, Minas Liarokapis, and Shahab Heshmati-Alamdari. Robust, fiducial markers based predictive control scheme for infrastructure inspection with unmanned aerial vehicles. In *2024 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 681–687, 2024.

[22] Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. Robot operating system 2: Design, architecture, and uses in the wild. *Science Robotics*, 7, 05 2022.

[23] Shivank Mittal and Ayan Sadhu. Fiducial marker-based decentralized computer vision for structural modal identification. *Journal of Sound and Vibration*, 612:119152, 2025.

[24] Mohammed Abbas Mousa, Mustafasanie M. Yussof, Ufuoma Joseph Udi, Fadzli Mohamed Nazri, Mohd Khairul Kamarudin, Gerard A. R. Parke, Lateef N. Assi, and Seyed Ali Ghahari. Application of digital image correlation in structural health monitoring of bridge infrastructures: A review. *Infrastructures*, 6(12):176, 2021.

[25] Richard M. Murray, Zexiang Li, and S. Shankar Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, Boca Raton, FL, 1994.

[26] Guang-Yu Nie, Saran Srikanth Bodda, Harleen Kaur Sandhu, Kevin Han, and Abhinav Gupta. Computer-vision-based vibration tracking using a digital camera: A sparse-optical-flow-based target tracking method. *Sensors*, 22(18):6869, 2022.

[27] Edwin Olson. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3400–3407, 2011.

[28] C. Rembe, Benjamin Halkon, and Mohamed A. A. Ismail. Measuring vibrations in large structures with laser-Doppler vibrometry and unmanned aerial systems: A review and outlook. *Advanced Devices & Instrumentation*, 6, 2025.

[29] O. S. Salawu. Detection of structural damage through changes in frequency: a review. *Engineering Structures*, 19(9):718–723, 1997.

[30] Hubert Schreier, Jean-José Orteu, and Michael A. Sutton. *Image Correlation for Shape, Motion and Deformation Measurements: Basic Concepts, Theory and Applications*. Springer, New York, NY, USA, 2009.

[31] Haojie Wang, Jin-Kun Guo, Han Mo, Xikang Zhou, and Yiping Han. Fiber optic sensing technology and vision sensing technology for structural health monitoring. *Sensors*, 23(9):4334, 2023.

[32] John Wang and Edwin Olson. Apriltag 2: Efficient and robust fiducial detection. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4193–4198, 2016.

[33] Peter D. Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[34] Yufeng Weng, Jiazeng Shan, Zheng Lu, Xilin Lu, and Billie F. Spencer. Homography-based structural displacement measurement for large structures using unmanned aerial vehicles. *Computer-Aided Civil and Infrastructure Engineering*, 36(9):1114–1128, 2021.

[35] K. Y. Wong, D. W. Allen, H. Sohn, and C. R. Farrar. Design of a structural health monitoring system for long-span bridges. *Structure and Infrastructure Engineering*, 3(4):283–293, 2007.

[36] Hyungchul Yoon, Vedhus Hoskere, Jong-Woong Park, and Billie F. Spencer. Cross-correlation-based structural system identification using unmanned aerial vehicles. *Sensors*, 17(9):2075, 2017.

[37] Boxin Zhao, Zongzhe Li, Jun Jiang, and Xiaolin Zhao. Relative localization for UAVs based on AprilTags. In *2020 Chinese Control and Decision Conference (CCDC)*, pages 444–449, 2020.

[38] Yizhou Zhuang, Weimin Chen, Tao Jin, Bin Chen, He Zhang, and Wen Zhang. A review of computer vision-based structural deformation monitoring in field environments. *Sensors*, 22(10):3789, 2022.

APPENDIX

APPENDIX

AN APRILTAG MEASUREMENT AND PNP DERIVATION

A.1 AprilTag Physical Measurement and PnP Parameter Derivation

The PnP solver requires the physical side length of the AprilTag to recover metric depth from image observations. For a square tag of physical side length s (metres), the four 3D corner positions are defined as

$$P_1 = \left(-\tfrac{s}{2},\ \tfrac{s}{2},\ 0\right), \quad P_2 = \left(\tfrac{s}{2},\ \tfrac{s}{2},\ 0\right), \quad P_3 = \left(\tfrac{s}{2},\ -\tfrac{s}{2},\ 0\right), \quad P_4 = \left(-\tfrac{s}{2},\ -\tfrac{s}{2},\ 0\right) \tag{A.1}$$

Under the pinhole camera model, the estimated depth z to a tag of apparent pixel width w_{px} is

$$z = \frac{f_x \cdot s}{w_{\text{px}}} \tag{A.2}$$

where f_x is the horizontal focal length in pixels. A 1 mm error in the declared tag size propagates directly into depth error. At a standoff of 1.5 m, a 1 mm size error produces

$$\delta z = z \cdot \frac{\delta s}{s} = 1500 \text{ mm} \times \frac{1 \text{ mm}}{67.3 \text{ mm}} \approx 22 \text{ mm} \tag{A.3}$$

This is comparable to the COARSE_ALIGN convergence threshold of ±50 mm and is large enough to affect standoff control. For this reason, the tag size was physically measured rather than taken from a nominal print specification.

The tag36h11 markers used in all experiments were printed at a nominal size of 165.1 mm measured across the outer black border. The inner data pattern, which defines the four corner pixel coordinates returned by the AprilTag detector, spans the inner 81% of the outer border. The physically measured inner side length is

$$s = 165.1 \text{ mm} \times 0.408 = 67.3 \text{ mm} = 0.0673 \text{ m} \tag{A.4}$$

Table A.1: AprilTag size measurement history and resolution

Value (m)	Source	Status
0.127	Simulation model (SDF geometry)	Simulation only
0.165	Outer border, hardware nominal	Incorrect for PnP
0.0673	Inner pattern, physically measured 2026-03-18	**Used in all hardware experiments**

The value 0.0673 m is used without rounding in all PnP nodes. Rounding to the nearest 5 mm (e.g., 0.070 m) would introduce a systematic 90 mm depth bias at 1.5 m standoff, exceeding the controller's convergence tolerance.

A.2 Effect of Tag Size Error on Vibration Frequency Identification

This section addresses whether a tag size miscalibration could corrupt the natural frequency identified from the relative pose time series.

If the declared tag size contains a scale error factor k such that $s_{\text{declared}} = k \cdot s_{\text{true}}$, then `solvePnP` returns poses scaled by k:

$$\mathbf{T}^{\text{cam}}_{\text{ref,measured}} = k \cdot \mathbf{T}^{\text{cam}}_{\text{ref,true}}, \qquad \mathbf{T}^{\text{cam}}_{\text{vib,measured}} = k \cdot \mathbf{T}^{\text{cam}}_{\text{vib,true}} \tag{A.5}$$

The relative pose computed from these scaled measurements is

$$\begin{aligned}
\mathbf{T}_{\text{rel,measured}} &= \left(k\,\mathbf{T}^{\text{cam}}_{\text{ref,true}}\right)^{-1}\left(k\,\mathbf{T}^{\text{cam}}_{\text{vib,true}}\right) && \text{(A.6)}\\
&= \frac{1}{k}\,(\mathbf{T}^{\text{cam}}_{\text{ref,true}})^{-1}\cdot k\,\mathbf{T}^{\text{cam}}_{\text{vib,true}} && \text{(A.7)}\\
&= (\mathbf{T}^{\text{cam}}_{\text{ref,true}})^{-1}\,\mathbf{T}^{\text{cam}}_{\text{vib,true}} && \text{(A.8)}\\
&= \mathbf{T}_{\text{rel,true}} && \text{(A.9)}
\end{aligned}$$

The scale factor k cancels exactly in the relative transform. A tag size error therefore has no effect on the relative displacement time series $x(t) = \mathbf{T}_{\text{rel}}[0, 3]$, and consequently no effect on the Welch PSD peak frequency identified from $x(t)$. The tag size error affects only absolute pose amplitudes, not the relative structural displacement or its spectral content.

This result confirms that the natural frequency measurements reported in Chapter 5 are valid regardless of tag size calibration precision, provided both tags are declared with the same size value in all PnP nodes.

VITA

Javier Becerril earned a Bachelor of Science degree in Mechanical Engineering, in May 2023. He further pursued his academic interests by earning a Master of Science in Computer Science from The University of Texas Rio Grande Valley in May 2026.

Javier's research focused on UAV development and simulations. These areas reflect his strengths in robotics, hardware, and his willingness to learn. Through his college career Javier has held internship positions within mechanical and software engineering. Javier has also participated in entrepreneurship programs and competitions. The author may be reached at becerjav777@gmail.com