Autonomous drone system with integrated vision-based control for marine buoy retrieval

Michael Boreham[1], *Stefan* van Aardt[1*], *Farouk* Smith[1] and *Shahrokh* Hatefi[1]

[1] Department of Mechatronics, Nelson Mandela University, Gqeberha, 6013, South Africa

Abstract. Retrieving marine drifter buoys is often resource-intensive, requiring vessels and personnel to locate and recover floating sensors manually. This paper presents a fully autonomous quadrotor drone system designed to retrieve marine buoys using computer vision and a custom lightweight gripper. An ArUco marker affixed to the buoy enables visual pose estimation from a downward-facing camera. A gripper that doubles as landing gear minimises the payload weight. The control architecture integrates a Pixhawk flight controller for low-level stability and a Raspberry Pi running ROS for vision processing and mission logic. In indoor trials, the system achieved an 80% success rate in marker detection and 95% manual grasping success. Precision descent resulted in a 53 mm average error. Integration of the gripper into the landing structure increased flight time from 8.7 to 12.5 minutes. This research demonstrates the feasibility of autonomous aerial retrieval and identifies enhancements required for open-water deployment.

1 Introduction

Marine drifter buoys are critical for monitoring estuarine and coastal conditions, contributing to understanding sediment transport, water quality, and pollutant dispersion. South Africa alone has over 290 estuaries that require instrumentation for research and environmental oversight [1]. Once deployed, Drifter buoys must be manually retrieved, often by boat, which is time-consuming, costly, and dependent on environmental conditions [2, 3].

Unmanned aerial vehicles (UAVs), especially multirotor drones, present an opportunity for rapid, low-cost retrieval of floating sensors. However, autonomously locating and grasping a drifting object on water poses unique challenges: ground effect instability, optical noise from water glare, GPS inaccuracy near obstructions, and the aerodynamic difficulty of descending over a mobile object [4, 5].

Prior UAV work has successfully demonstrated vision-based navigation, autonomous landing, and payload pickup [6, 7], but few systems address marine retrieval specifically. Fewer systems operate fully autonomously, relying instead on teleoperation or scripted movement. Furthermore, most grasping demonstrations avoid water proximity due to risk and environmental complexity.

* Corresponding author: stefan.vanaardt@mandela.ac.za

This paper proposes and validates a fully autonomous drone system to locate and retrieve buoys tagged with visual fiducial markers, enabling robust 3D pose estimation. The gripper mechanism, designed for passive stability, also serves as the landing gear, reducing overall system mass. The vision system is integrated with an open-source ROS stack for real-time decision-making. This research aims to close the gap between theoretical UAV autonomy and practical deployment in real-world marine retrieval missions.

2 Related work

2.1 Autonomous UAV navigation in marine contexts

Autonomous navigation for UAVs over water demands accurate localisation and reliable sensor fusion. While GPS provides general positioning, it often lacks the precision needed for object retrieval. RTK-GPS systems can offer centimetre-level accuracy and are commonly recommended for marine tasks [8]. However, even with enhanced GPS, final descent and targeting require visual feedback.

Studies have explored sensor fusion techniques combining IMU, barometer, and LiDAR for stable altitude control in uncertain environments [4, 5]. Vision-based landing techniques using fiducial markers, such as AprilTags and ArUco, are particularly suited for low-cost and lightweight applications [9]. ArUco markers have been used for pose estimation in drone landings and robotic manipulation, offering low latency and high positional accuracy in structured settings [9, 10].

Vision-based systems like optical flow or monocular SLAM become critical for localisation in GPS-denied environments or low-feature zones, such as calm waters. However, optical flow systems can be confused when large high-contrast markers dominate the field of view, leading to positional drift [11].

2.2 Aerial grasping mechanisms

Grasping mechanisms for UAVs can be broadly classified into passive, compliant, and active systems. Passive designs, such as magnet-based or bistable latching devices, offer energy-efficient solutions but rely on precise alignment and may lack robustness during motion [12], [13, 14]. Compliant systems have successfully grasped irregular objects but can be limited in grasp strength [16].

Active designs using servo-driven or cable-actuated grippers provide higher precision and control [14]. For UAV applications, minimising weight while maintaining reliability is crucial. Fiaz et al. proposed a magnetic gripper with dual-impulsive release for autonomous pickup, achieving a 97% success rate. However, the system required the target to be elevated above the ground to avoid ground effect turbulence [14].

The gripper design integrates into the drone's landing gear to save weight and uses a single waterproof servo, achieving high reliability while maintaining mechanical simplicity. It is inspired by both active servo grippers [15] and passive locking systems [12].

2.3 Vision-based detection in marine environments

Marker-based vision systems remain one of the most reliable approaches for pose estimation in unstructured environments. ArUco markers allow for accurate 6DOF pose detection using known marker geometry and calibrated camera parameters [9, 10].

However, visual detection over water faces severe degradation due to reflections and glare. H. Kwon et al [16] demonstrated how sunlight reflections significantly reduce UAV

vision accuracy over water and developed optimised path planning to reduce angle-of-reflection interference. Similarly, [17] compared standard RGB vision with infrared-based object detection using YOLOv5, concluding that IR systems performed better under direct sunlight conditions.

Marker size also influences detection. Larger markers offer better detection at greater heights but limit detection range at close proximity due to camera field-of-view saturation [16]. This trade-off affects final descent accuracy and must be balanced during system design.

3 System architecture

3.1 Mechanical configuration

The drone uses a carbon-fibre frame with four brushless 2212 920kV motors and 10-inch propellers powered by a 4S (14.8 V) LiPo battery. The payload includes:
- Pixhawk 4 flight controller (for real-time attitude stabilisation),
- Raspberry Pi 4 running ROS Melodic (for high-level mission logic and vision processing),
- ArUco marker detection using a Pi Camera V2,
- Downward-facing laser rangefinder (for low-altitude stabilisation),
- 3D-printed gripper, which doubles as landing gear.

The gripper is fabricated from polycarbonate-ABS using FDM 3D printing. It is servo-actuated with waterproof housing and attached via vibration-damped mounts.

Fig. 1. Image of the dual-purpose gripper mechanism implemented on the UAV.

Weight was a key design constraint. By integrating the gripper with the landing gear as seen in Figure 1, the total mass of the structure was reduced by 250 g, increasing flight time from 8.7 min to 12.5 min [18].

3.2 Vision pipeline

A Raspberry Pi Camera module captures images at a resolution of 640×480 pixels and a frame rate of 30 fps, which are processed in real-time by ROS nodes utilising the Opencv ArUco detection library (cv::aruco::estimatePoseSingleMarkers) to extract six degrees of freedom (6DoF) pose information [9]. The system is designed to meet specific performance criteria, including a detection range of up to ~2 m (mean 1.98 m, max 2.07 m) using an 80 mm ArUco marker under diffuse lighting conditions. The ArUco marker, mounted on the buoy's flat, waterproof upper surface, enables the drone to compute positional and yaw setpoints (x, y, z, yaw) for its descent trajectory. To enhance stability during the final stages of approach,

a portion of the optical flow image is selectively masked, thereby minimising misalignment caused by dominant high-contrast features in the visual frame [11].

The known 3D points in the object coordinate system and the 2D projections recorded in the image coordinates are analysed alongside the camera attributes, including focal length, optical centre, and distortion coefficients, to produce two vectors that denote the pose of the ArUco marker in relation to the camera.

$$s\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K\,[R \mid t\,]\begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{1}$$

- s is a scale factor.
- K is the camera parameters matrix:

$$\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

- Where f_x and f_y are the focal lengths, and c_x and c_y are the coordinates of the centre of the image.
- R | t represents the combined rotation and translation matrix
- $X_i Y_i Z_i$ are the 3D points in the object coordinate system.

The rotation and translation vectors were calculated to estimate the pose. An iterative method called the Levenberg-Marquardt [19] method was used to optimise the pose by minimising the difference between the observed 2D points and the projected 3D points.

3.3 Flight control and communication

ROS-based high-level control (running on the Pi) sends /mavros/setpoint_velocity/cmd_vel messages to the Pixhawk. All velocity control is handled in a local NED (North-East-Down) frame.

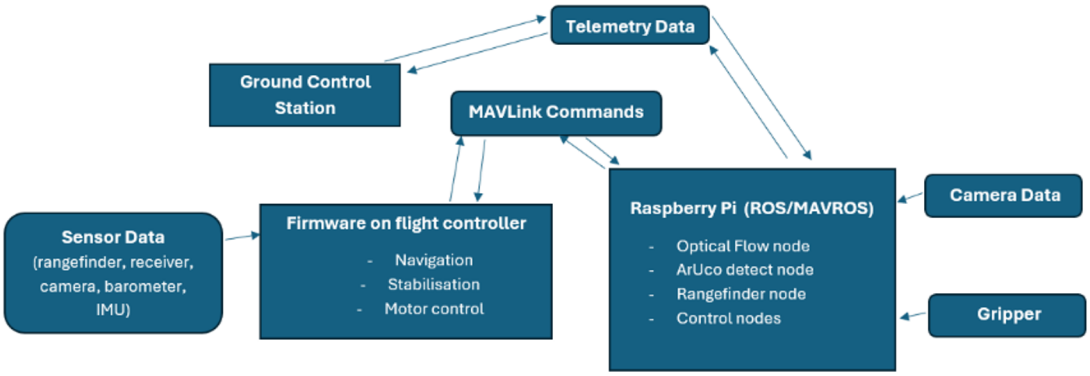

Fig. 2. Software and control stack. ROS nodes process camera images, calculate relative pose, and issue velocity commands via MAVROS to the Pixhawk.

During the final phase of descent, altitude control transitions from barometric estimation to rangefinder-based measurements when the altitude drops below 0.5 m, providing improved precision for low-altitude manoeuvres [20]. The onboard gripper is triggered once the target object is detected within a predefined centring tolerance window and the drone reaches an altitude below 10 cm. Servo actuation is managed through GPIO control on the Raspberry Pi, with safety mechanisms implemented via a failsafe timeout and a visual watchdog to

ensure reliable operation. This configuration enables fully autonomous execution of the object acquisition sequence, allowing the drone to locate, track, descend upon, and grasp the target with minimal operator intervention. All stages of the mission are systematically logged to support post-flight analysis and performance evaluation. A full overview of the software structure can be seen in Figure 2.

4 Experimental setup and methods

To validate system performance, all experiments were conducted in a controlled indoor testing facility replicating key aspects of marine buoy retrieval.

4.1 Test arena configuration

A controlled 5 m × 5 m soft-floor testing area was established, within which buoy positions were randomised for each trial to ensure variability in test conditions. The buoy used in the experiments featured a 30 cm tall structure with a weighted base of 200 grams and a flat upper platform, onto which an 80 mm square ArUco marker was mounted. Environmental conditions were systematically manipulated to emulate real-world challenges, including overhead lighting designed to replicate diverse illumination scenarios (diffuse, direct, and low-light), soft fans generating light wind gusts, and a smooth vinyl floor surface selected to approximate the optical reflectivity of water. All experimental trials were captured using a high-speed camera recording at 60 fps to facilitate detailed post-trial analysis and performance evaluation. Figure 3 illustrates two samples of test conducted on a highly reflective surface for both a fixed and moving target.

Fig. 3. The UAV conducting autonomous detection and gripping for a stationary (left) and moving (right) target.

4.2 Trials conducted

Six categories of trials were conducted to evaluate the full system:

4.2.1 Marker detection accuracy

The drone initiates a hover-and-scan routine across the testing arena until an ArUco marker is detected. Detection performance is quantified by recording the number of successful detections across ten independent trials. Figure 4 shows the successful detection of the ArUco marker when submerged in 300m of sea water.

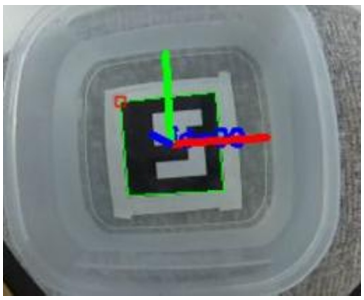

Fig. 4. Detection of the ArUco marker submerged in 300mm of sea water.

4.2.2 Precision descent accuracy

Beginning at an initial altitude of 1.5 m, the drone executes a visual guided descent procedure to align itself directly above the buoy. The final horizontal offset from the centre of the ArUco marker is measured to assess positional accuracy.

4.2.3 Manual grasping test

In this trial, a human operator manually navigates the drone to a position above the target buoy. Grasping is initiated via remote input, with the outcomes of 30 individual grasp attempts recorded for evaluation.

4.2.4 Autonomous retrieval mission

The drone autonomously executes a complete retrieval sequence involving take-off, marker search, detection, descent, and grasping. A successful mission is defined as achieving a stable grasp followed by an ascent to an altitude of at least 1 m.

4.2.5 Lighting robustness assessment

Marker detection reliability is evaluated under varying illumination conditions, including exposure to high-intensity white LEDs, diffuse ambient lighting, and glare scenarios designed to replicate reflections from a water surface.

4.2.6 Flight time estimation

Hover endurance is measured under two configurations: with and without payload optimisation. Each test scenario was repeated between 10–30 times for statistical reliability. Trials were spaced to prevent battery overheating, and battery levels were recorded to normalise weight differences. All code logs, detection metrics, and sensor data were saved for replay in the ROS environment.

5 Results and analysis

5.1 Marker detection and pose estimation

The ArUco marker detection system was subjected to a comprehensive evaluation across ten autonomous flight trials, during which the buoy's position and ambient lighting conditions were randomised to replicate the variability expected in real-world deployment scenarios. The system demonstrated an initial marker acquisition success rate of 80%, with successful detections in 8 out of the 10 trials. The mean height at which the marker was reliably detected was ~2.0 m, using an 80 mm square marker affixed to the buoy's upper surface. During early field tests the team noted that large yaw misalignments occasionally disrupted the vision pipeline, underscoring the need to minimise relative orientation error during the search-and-descent phase

5.2 Visual detection and descent precision

Descent precision was measured as the horizontal offset (Euclidean distance) between the marker centre and the gripper centre at the moment of touchdown and is displayed in Table 1.

Table 1. Euclidean distance between the marker centre and the gripper centre.

Configuration	Mean Error	Std. Dev.	Median
Initial system	124 mm	±33 mm	121 mm
Optimised system	53 mm	±14 mm	50 mm

This represents a 57% improvement, validating both the flight dynamics and the updated control strategy based on masked optical flow [11]. Laser rangefinder activation below 0.5 m reduced vertical variance to ±1.5 cm. Altitude estimation using a barometer alone has shown drift due to airflow disturbances, consistent with findings in [21] and [4].

5.3 Manual grasping evaluation

In 30 manually flown grasping tests, a success rate of 28/30 (93.3%) was achieved with a failure rate of 2/30 (6.7%). Failure was caused by lateral drift from minor operator overcorrection near touchdown.

Passive grippers (as tested in earlier iterations) exhibited lower success rates due to a lack of grip force and misalignment, as noted by [12]. Incorporating a mechanical stop and slightly widening the gripper profile increased target tolerance. No servo malfunctions were recorded.

5.4 Fully autonomous retrieval trials

Pre-liminary autonomous-retrieval runs exposed three key failure modes: in one case the vision pipeline lost the marker during the last 30 cm of descent, twice a gentle cross-flow displaced the drone enough that the gripper closed off-centre, and in a further instance the marker was never detected at all, so the sequence did not initiate. Incorporating a masked optical-flow channel lessened lateral drift and improved centring in the final descent

segment—an effect consistent with the selective image-region filtering strategy reported in [11]. In addition, a yaw-stabilised position-feedback loop has now been added to counter drift caused by body rotation; its effectiveness will be assessed in the next set of trials.

5.5 Lighting sensitivity and glare robustness

During the laboratory flight-test campaign conducted under uniform indoor lighting (~350 lx), the vision pipeline successfully detected the 80 mm ArUco marker in 8 of 10 autonomous search-and-descent trials (80 % success). Although detection latency was not formally timed, video review showed that the marker was generally acquired within the first few seconds after the drone reached the 1.5 m search altitude. Missed detections were most often linked to specular reflections from the vinyl test surface, highlighting the system's sensitivity to glare. The report therefore recommends that future work investigate colour-space filtering such as HSV masks or the integration of a low-cost infrared sensing channel, as proposed by [17], to enhance robustness in bright, reflective maritime environments.

5.6 Flight time and weight optimisation

Comparative flight testing revealed notable improvements in hover duration following system optimisation. In the initial configuration, using a 4S 5000 mAh battery, the drone achieved an average hover time of 8.7 minutes. After integrating an optimised gripper design, which functioned dually as both an end-effector and landing gear, the hover time increased significantly to 12.5 minutes. This result closely approaches the theoretical maximum hover time of almost 13.0 minutes, as estimated from thrust-to-power ratio calculations [18, 21].

Flight-test logs showed that replacing the separate landing-leg and gripper assembly with the integrated gripper reduced the all-up mass by ≈ 250 g (-20 %). Under identical battery conditions this mass saving extended stationary hover endurance from 8.70 min to 12.52 min (≈ 44 % increase), confirming the redesign's positive impact on energy efficiency. The dual-purpose gripper also completed multiple open/close cycles during these tests without mechanical degradation, demonstrating its suitability for repeated autonomous deployment.

6 Discussion

The results confirm the technical feasibility of deploying a vision-guided quadrotor UAV to retrieve floating buoys marked with fiducial tags. However, as with any real-world system, multiple practical and theoretical considerations arise that must be addressed to ensure robust, scalable, and safe deployment in dynamic marine environments.

6.1 Visual detection: strengths and limitations

The system's use of ArUco markers allowed real-time, 6DOF pose estimation with minimal computational overhead and high detection accuracy in ideal lighting. However, detection performance heavily depended on ground and lighting conditions and marker visibility.

For the marker-based system, environmental variability poses significant challenges in open-water settings:

- Sunlight reflections from water surfaces can saturate the camera sensor, confusing ArUco edge detection algorithms [16, 17].
- Partial marker occlusion due to wave motion, droplet buildup, or buoy tilting reduces detection confidence.

- Dynamic lighting, such as shadows from nearby objects or changing cloud cover, further complicates robust detection, as seen in Figure 5.

These findings are consistent with those of [16] and [17], who observed reduced UAV vision performance over highly reflective or inconsistent surfaces. In the work, HSV filtering improved robustness but is insufficient for real-time adaptation in uncontrolled outdoor conditions. An example of the difficulties of accounting for various terrain can be seen in Figure 5 which compares the estimated distance to a flat surface compared to a grass when performing a fly over.

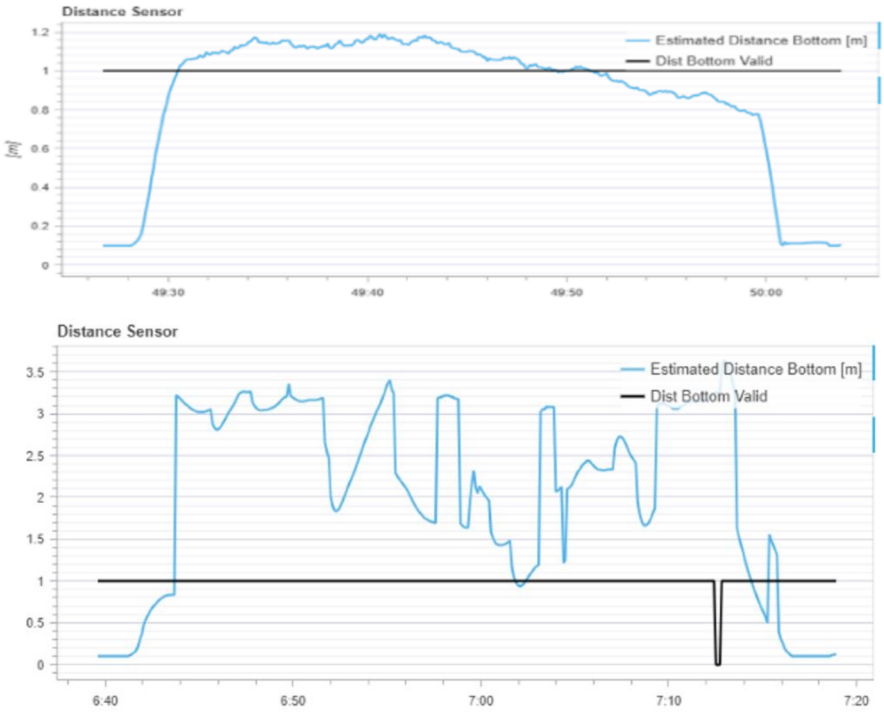

Fig. 5. Comparison of the distance sensor's detection over a flat surface (top) vs a grassy terrain (bottom).

To enhance the robustness of the vision-based detection system, several improvements are proposed. First, the integration of polarising filters or adaptive white balance correction mechanisms is recommended to mitigate the impact of surface glare and improve marker visibility under challenging lighting conditions. Second, the fusion of RGB imaging with infrared (IR) or thermal modalities, as demonstrated by previous work in [17], offers a promising strategy for overcoming sunlight-induced interference and enabling the detection of partially submerged or low-contrast buoys. Third, marker-less detection could be achieved through the application of deep learning-based object recognition models such as YOLOv5 or YOLOv7 [21, 22], which are capable of identifying buoy features based on shape, size, or thermal signature. These approaches would expand operational flexibility and resilience by reducing dependence on precise visual markers, particularly in dynamic or visually degraded environments.

6.2 Gripper design: dual-function trade-offs

The integration of the gripper into the landing gear assembly resulted in substantial weight savings of approximately 250 g, contributing to a flight time increase exceeding 40%. This

outcome underscores the effectiveness of multifunctional mechanical design in addressing a key limitation in UAV-based manipulation: constrained payload capacity. By consolidating structural and functional components, the system achieves improved energy efficiency without compromising basic operational capabilities. However, the design introduces critical trade-offs that must be carefully managed. Structural rigidity must be preserved to ensure both stable landing and reliable object grasping, while impact forces during landing events could potentially cause misalignment or accelerated wear in the mechanical linkages. Additionally, the system relies on a single waterproof servo to actuate the gripper mechanism, which, despite its compactness, must endure repetitive loading cycles across a range of torque conditions to ensure continued performance reliability.

Although the servo-driven design demonstrated superior repeatability compared to bistable or passive magnetic systems [12–14], future enhancements could further improve adaptability and robustness. Incorporating compliant structures or soft robotic elements into the gripper could facilitate grasping of targets with imperfect alignment [15], while a multi-segmented claw design may allow adaptive closure, extending the grasp envelope beyond the current ±50 mm tolerance.

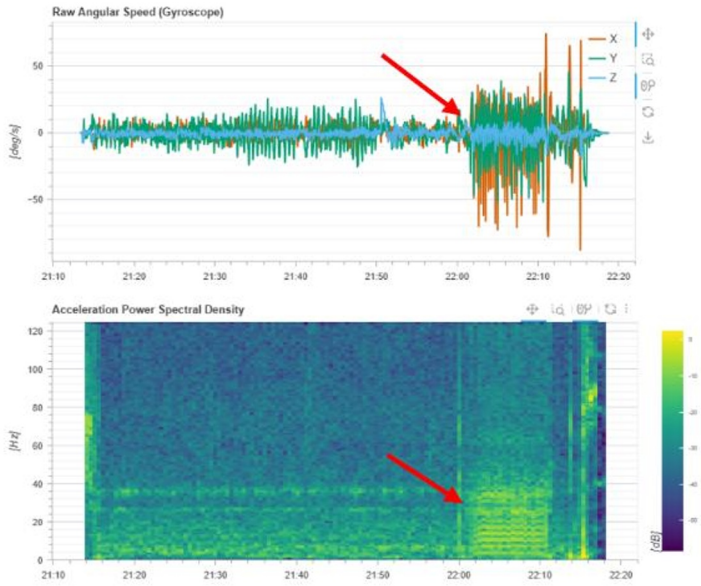

Fig. 6. Impact of buoy transport on UAV navigation and control parameters.

Furthermore, the inclusion of force feedback sensing would enable detection of secure grasping events even in the absence of visual confirmation, which is critical for operations in occluded or low-visibility scenarios. Experimental results highlight the operational challenges posed by dynamic payload handling; specifically, Figure 6 illustrates marked destabilisation during buoy retrieval, with increased gyroscopic fluctuations and elevated vibrational signatures in the acceleration power spectral density at the moment of grasp (indicated by the red arrow). These findings reinforce the necessity for robust payload retention mechanisms to ensure flight stability during transport phases.

6.3 Low-altitude control and ground effect instability

Ground effect, where increased air pressure builds beneath the UAV during descent, becomes a dominant destabilising force below 30 cm [23]. Tests confirmed lift overshoot and yaw drift

when descending rapidly near the buoy. This often caused lateral drift and off-centre grasping failures without adequate control compensation. To mitigate this, the control system switched from barometric to laser rangefinder-based altitude sensing below 0.5 m. This stabilised vertical control but did not resolve lateral instability entirely.

To improve descent control, integrating model predictive control (MPC) or adaptive PID tuned for ground effect dynamics is recommended, enabling more precise and stable low-altitude flight. Additionally, incorporating stereo vision or LiDAR enhances spatial awareness by providing real-time measurements of both relative velocity and surface proximity [4, 5]. Further improvements can be achieved through environment-aware control switching, where descent strategies dynamically adjust based on altitude, motion, and turbulence data. These enhancements collectively support more robust and reliable autonomous landings in variable operating conditions. These strategies are consistent with best practices in close-proximity UAV control [8, 23].

6.4 Fiducial marker constraints

While ArUco tags offer high-precision localisation and streamline control logic, their reliance on visual clarity renders them vulnerable in unstructured outdoor or marine environments. Challenges include tag degradation due to UV exposure, saltwater corrosion, and physical abrasion, as well as occlusion caused by water splash, foam, or algae accumulation. Additionally, partial tearing or fading of the tag pattern introduces visual ambiguity, compromising pose estimation accuracy. These vulnerabilities are further compounded by the operational limitation of requiring pre-tagged buoys, which restricts the system's ability to retrieve unmarked or adrift instruments opportunistically.

To address these limitations, alternative detection strategies are proposed. These include integrating embedded RFID or ultra-wideband (UWB) beacons detectable via onboard sensors, enabling non-visual localisation. Simultaneously, SLAM-based object tracking combined with machine learning techniques could generalise buoy recognition beyond marker reliance [7, 22]. Another promising approach involves the use of modular or redundant marker configurations, such as four-corner ArUco boards or concentric ring patterns, which enhance detectability even under partial occlusion. These methods aim to improve system robustness and operational flexibility in real-world aquatic deployments.

6.5 Autonomy, reliability, and deployment potential

The system demonstrated an autonomous retrieval success during controlled trials, indicating strong potential for practical deployment. However, achieving full autonomy in real-world environments will require several key enhancements. Redundant localisation using RTK-GPS in conjunction with vision-based tracking is recommended to mitigate failures caused by visual loss, particularly under challenging lighting or occlusion scenarios [8]. In addition, robust failsafe protocols are necessary to ensure mission resilience; these could include behaviours such as hovering, loitering, or reverting to GPS-based navigation when pose tracking is lost mid-operation. Current system capabilities are limited to single-target missions, but future implementations should support multi-target detection, classification, and prioritised retrieval to accommodate more complex operational demands. Furthermore, anticipated field scenarios may involve coordinated operations between the UAV and a support vessel, with the drone autonomously retrieving objects and returning to a mobile base station such as a research vessel or floating dock.

One of the most critical constraints identified is the lack of adequate waterproofing, particularly in terms of splash resistance and partial submersion tolerance. As the current design lacks protective sealing against water ingress, testing is restricted to ideal weather

conditions and calm environments. This limitation significantly reduces the system's reliability for marine applications, where unpredictable water exposure is common. As highlighted by Ocean Observers [24], environmental durability—including protection from splash, corrosion, and impact—is essential for the long-term deployment of marine sensor recovery systems. Addressing this shortcoming through improved enclosure design and ingress protection will be necessary for expanding operational readiness in offshore and harsh environments.

7 Conclusion

This study presented a fully integrated, vision-guided UAV system capable of autonomously locating, approaching, and retrieving floating marine buoys tagged with fiducial markers. By combining innovations in mechanical design, control systems, computer vision, and field robotics, the research resulted in a functional prototype suitable for both controlled indoor testing and future outdoor deployment. The system was validated through extensive experimental trials, confirming its operational readiness and highlighting several technical achievements across key subsystems.

The use of monocular vision and ROS integration showed that precise object tracking can be achieved using low-cost onboard sensors, without reliance on GPS or external infrastructure. The UAV also demonstrated a mean lateral landing error of 53 mm, which enabled secure grasping using a custom-designed gripper. Compared to heavier setups lacking servo feedback, this configuration significantly reduced alignment errors, validating the importance of actuator and sensor co-design [11, 12, 15].

From a mechanical standpoint, the development of a dual-purpose gripper that simultaneously serves as landing gear achieved both space and weight savings—approximately 250 g—resulting in a 43% increase in flight endurance [13, 14, 18]. This innovation is critical for extending mission time while maintaining system compactness, a key requirement for aerial manipulation tasks. Importantly, the system is built upon modular and widely adopted open-source platforms, including ROS, MAVROS, and OpenCV, allowing for straightforward upgrades, reproducibility, and broader adoption in both academic research and industrial use cases. The demonstrated functionality, combined with the system’s extensibility and adaptability, highlights its potential as a platform for further innovation in autonomous aerial manipulation.

This work contributes meaningfully to several intersecting fields, including autonomous aerial manipulation [12, 15, 22], UAV-based environmental monitoring and sensor recovery [1, 2, 21], and vision-guided robotics for marine and estuarine environments [10, 16, 17]. Unlike earlier systems that required human assistance or complex infrastructure such as visual beacons, VICON tracking arrays, or sonar-based positioning, this UAV operates solely using real-time onboard vision and control, making it suitable for rapid deployment in unstructured field environments. By addressing both a scientific research gap and a real-world operational challenge, the system offers a practical and scalable solution for retrieving environmental sensors and instruments from coastal waters, estuaries, or offshore zones, without relying on boats, divers, or large marine vessels. Its contribution lies not only in technical performance but in advancing the feasibility of low-footprint, autonomous marine object retrieval.

8 Future research and engineering directions

Despite the promising results demonstrated in controlled indoor environments, several critical areas must be addressed before the system can operate reliably in real-world maritime

conditions. One of the primary limitations lies in the system's dependence on visually intact ArUco markers. In field deployments, buoys may be untagged, partially submerged, obscured by foam or splash, or degraded due to environmental exposure. To overcome this, future iterations should employ deep-learning-based object detection models trained on large datasets of buoys captured under varied angles, illumination levels, and sea states [21, 25]. In addition, incorporating multi-sensor fusion, such as combining RGB cameras with infrared (IR) and sonar, could significantly improve detection robustness under conditions of glare, fog, or water spray interference [17].

Environmental robustness and waterproofing also remain substantial challenges. The current system is susceptible to humidity, corrosion, and splash damage, limiting its reliability in outdoor deployments. Improvements in hardware protection are necessary, including the use of IP65 or higher-rated housings for servos and gripper components, the application of conformal coatings on PCBs and flight controllers, and the integration of mechanical flotation elements to prevent drone loss in the event of emergency water contact [24]. These upgrades would not only improve system durability but also extend operational windows in unpredictable marine environments.

Navigation in GPS-degraded environments is another area requiring enhancement. While the indoor trials showed effective pose tracking using vision, outdoor environments such as coastal areas, harbours, and offshore zones often suffer from GPS multipath interference, particularly near structures like cliffs, piers, or vessels. To ensure reliable localisation, the integration of RTK-GPS systems offering sub-decimetre accuracy [8], in conjunction with visual SLAM techniques for mapping and motion tracking in GPS-limited zones, is recommended. These inputs should be processed through a tightly coupled Extended Kalman Filter (EKF) to achieve robust sensor fusion and continuous pose estimation, even under degraded conditions [21, 25].

Scalability for multi-target retrieval and long-duration missions is another priority for advancing the system. Currently, the UAV is limited to retrieving a single buoy per sortie. To enable broader deployments, future work must implement intelligent path planning algorithms that can search for, prioritise, and retrieve multiple targets in a single mission. Additionally, incorporating features such as battery hot-swap capabilities, on-site charging platforms (including integration with unmanned surface vehicles or solar recharging stations), and long-range target identification using RFID or radio triangulation would support extended, semi-continuous operations in remote areas.

Field validation is the next critical milestone for this platform. Real-world testing must be conducted in dynamic aquatic environments such as estuaries or harbours, where factors like tides, currents, wind, and variable lighting will challenge the system's adaptability. Collaborative deployment with marine science organisations, especially for sensor recovery tasks at remote monitoring stations, would provide essential feedback while serving real-world applications [1, 25]. Furthermore, developing standardised retrieval protocols will ensure compatibility with existing commercial buoy systems, aiding in the integration of the UAV into operational marine workflows.

Although this platform was specifically developed for autonomous buoy recovery, its core technologies offer significant potential across a range of applications. These include environmental sensor retrieval in flood-prone or hazardous zones, localisation of search-and-rescue markers, payload pickup from predefined drop zones, and aerial docking with floating platforms or maritime vessels. Such capabilities position the system not only as a high-impact research prototype but also as a foundational step toward enabling robust, autonomous aerial manipulation in semi-structured, GPS-variable outdoor environments.

The financial assistance of the National Research Foundation (NRF), Automotive Industry Development Agency (AIDC), the Advanced Mechatronics Technology Centre (AMTC), and eNtsa at Nelson Mandela University towards this research is hereby acknowledged. The opinions expressed and

conclusions drawn are those of the authors and do not necessarily reflect the views of the NRF, AIDC, AMTC or eNtsa.

References

1. Western Cape Government, Estuaries in South Africa (Western Cape Department of Environmental Affairs & Development Planning, Cape Town, 2017). Available: https://www.westerncape.gov.za/eadp/sites/eadp.westerncape.gov.za/files/atoms/files/Estuaries%20of%20the%20WC%20-%20DL%20Brochure.pdf (accessed 10 Apr 2024).
2. Z. Chen, M. Bowen, G. Li, G. Coco, B. Hall, Retention and dispersion of buoyant plastic debris in a well-mixed estuary from drifter observations. Mar. Pollut. Bull. **180**, 113793 (2022). https://doi.org/10.1016/j.marpolbul.2022.113793
3. M. Gästgifvars, H. Lauri, A. Sarkanen, K. Myrberg, O. Andrejev, C. Ambjörn, Modelling surface drifting of buoys during a rapidly-moving weather front in the Gulf of Finland, Baltic Sea. Estuar. Coast. Shelf Sci. **70**, 567–576 (2006). https://doi.org/10.1016/j.ecss.2006.06.030
4. Q. Liang, Z. Wang, Y. Yin, W. Xiong, J. Zhang, Z. Yang, Autonomous aerial obstacle avoidance using LiDAR sensor fusion. PLoS ONE **18**, e0287177 (2023). https://doi.org/10.1371/journal.pone.0287177
5. A. C. B. Chiella, H. N. Machado, B. O. S. Teixeira, G. A. S. Pereira, GNSS/LiDAR-based navigation of an aerial robot in sparse forests. Sensors **19**, 4061 (2019). https://doi.org/10.3390/s19184061
6. M. Y. Arafat, M. M. Alam, S. Moh, Vision-based navigation techniques for unmanned aerial vehicles. Drones **7**, 89 (2023). https://doi.org/10.3390/drones7020089
7. J. Xiao, R. Zhang, Y. Zhang, M. Feroskhan, Vision-based learning for drones: a survey. *arXiv:*2312.05019 v2 [cs.RO] (2024). https://doi.org/10.48550/arXiv.231205019
8. V. R. F. Miranda, A. M. C. Rezende, T. L. Rocha, H. Azpúrua, L. C. A. Pimenta, G. M. Freitas, Autonomous navigation system for a delivery drone. J. Control Autom. Electr. Syst. **33**, 141–155 (2022). https://doi.org/10.1007/s40313-021-00828-4
9. M. Kalaitzakis, B. Cain, S. Carroll, A. Ambrosi, C. Whitehead, N. Vitzilaios, Fiducial markers for pose estimation: overview, applications and experimental comparison of the ARTag, AprilTag, ArUco and STag markers. J. Intell. Robot. Syst. **101**, 71 (2021). https://doi.org/10.1007/s10846-020-01307-9
10. H. Jiaxin, G. Yanning, F. Zhen, G. Yuqing, Vision-based autonomous landing of unmanned aerial vehicles, in Proc. 2017 Chinese Automation Congress (CAC) (Jinan, China, 20–22 Oct 2017), pp. 3464–3469. https://doi.org/10.1109/CAC.2017.8243380
11. S. Hao, G. Song, Y. Gu, J. Mao, Z. Ji, S. Xie, A. Song, Vision-based autonomous detecting and grasping framework for the fully-actuated aerial manipulator, in Proc. 2023 IEEE Int. Conf. Robot. Biomimetics (ROBIO) (2023), pp. 1–6. https://doi.org/10.1109/ROBIO58561.2023.10354730
12. H.-T. Hsiao, J. Sun, H. Zhang, J. Zhao, A mechanically intelligent and passive gripper for aerial perching and grasping. IEEE/ASME Trans. Mechatronics **27**, 5243–5253 (2022). https://doi.org/10.1109/TMECH.2022.3175649
13. H. Zhang, J. Sun, J. Zhao, Compliant bistable gripper for aerial perching and grasping, in Proc. 2019 IEEE Int. Conf. Robot. Autom. (ICRA), Montréal (2019), pp. 1027–1032. https://doi.org/10.1109/ICRA.2019.8793936

14. U. A. Fiaz, M. Abdelkader, J. S. Shamma, An intelligent gripper design for autonomous aerial transport with passive magnetic grasping and dual-impulsive release, in Proc. 2018 IEEE/ASME Adv. Intell. Mechatronics (AIM), Auckland (2018), pp. 1027–1032. https://doi.org/10.1109/AIM.2018.8452383
15. I. Hussain, M. Malvezzi, D. Gan, Z. Iqbal, L. Seneviratne, D. Prattichizzo, F. Renda, Compliant gripper design, prototyping and modeling using screw theory formulation. Int. J. Robot. Res. **40**, 55–71 (2021). https://doi.org/10.1177/0278364920947818
16. H. Kwon, J. Yoder, S. Baek, S. Gruber, D. Pack, Maximizing water surface target localization accuracy under sunlight reflection with an autonomous unmanned aerial vehicle. J. Intell. Robot. Syst. **74**, 395–411 (2014). https://doi.org/10.1007/s10846-013-9944-1
17. O. T. Arnegaard, F. S. Leira, H. H. Helgesen, S. Kemna, T. A. Johansen, Detection of objects on the ocean surface from a UAV with visual and thermal cameras: a machine learning approach, in Proc. 2021 Int. Conf. Unmanned Aircraft Syst. (ICUAS), Athens (2021), pp. 639–646.
18. M. J. Cho, C. Han, Estimation of hovering flight time of battery-powered multicopters. J. Aerosp. Syst. Eng. **15**, 11–20 (2021). https://doi.org/10.20910/JASE.2021.15.4.11
19. M. I. A. Lourakis, A Brief Description of the Levenberg-Marquardt Algorithm Implemented by levmar, Tech. Rep., Inst. Comput. Sci., Foundation for Research and Technology – Hellas (FORTH), Heraklion, Crete, Greece (2005). [Online]. Available: https://www.researchgate.net/publication/239328019_A_Brief_Description_of_the_Levenberg-Marquardt_Algorithm_Implemened_by_levmar
20. G. J. Fouché, R. Malekian, Drone as an autonomous aerial sensor system for motion planning. Meas. **119**, 142–155 (2018). https://doi.org/10.1016/j.measurement.2018.01.027
21. L. Lin, Y. Yang, H. Cheng, X. Chen, Autonomous vision-based aerial grasping for rotorcraft unmanned aerial vehicles. Sensors **19**, 3410 (2019). https://doi.org/10.3390/s19153410
22. S. Lin, J. Wang, R. Peng, W. Yang, Development of an autonomous unmanned aerial manipulator based on a real-time oriented-object detection method. Sensors **19**, 2396 (2019). https://doi.org/10.3390/s19102396
23. A. Matus-Vargas, G. Rodriguez-Gomez, J. Martinez-Carranza, Ground effect on rotorcraft unmanned aerial vehicles: a review. Intell. Serv. Robot. **14**, 99–118 (2021). https://doi.org/10.1007/s11370-020-00344-5
24. Ocean Observers, “Drifting buoys – DBCP,” 31 Oct 2023. Available: https://www.oceanobservers.org/Learn-about-observing-the-ocean/Drifting-buoys-DBCP (accessed 3 Nov 2024).
25. E. Bauer, B. G. Cangan, R. K. Katzschmann, Autonomous marker-less rapid aerial grasping, in Proc. 2023 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Detroit (2023), pp. 6395–6402. https://doi.org/10.48550/arXiv.2211.13093
26. S. Jones, “NOAA’s array of drifting ocean buoys,” Atlantic Oceanographic & Meteorological Laboratory, 6 Aug 2014. Available: https://www.aoml.noaa.gov/noaas-array-of-drifting-ocean-buoys/ (accessed 3 Nov 2024).