IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Autonomous Drone-Ground Robot Alignment Through Ground Robot Visual Servo Control With Drone Detection and Tilt Correction

Sean Clark Dominguez, Jeanette Pao, Immanuel Paradela, *Graduate Student Member, IEEE*, John Mel Bolaybolay, Earl Ryan Aleluya, Francis Jann Alagon, Sherwin Guirnaldo, Carl John Salaan, Kazunori Ohno, *Member, IEEE*, Yoshito Okada, Argel Bandala, and Abu Ubaidah Bin Shamsudin

***Abstract*—Retrieving ground robots from dangerous environments after their operation is a challenging task that poses risks for the personnel. Some researchers often employ drones for retrieval, which makes operations safer. However, this set-up requires an accurate method that guarantees drone and ground robot alignment due to inaccuracies in standard GPS devices, drone drifts, and wind gusts. Hence, this research article introduces simultaneous object detection and tilt correction as part of visual servoing to achieve precise drone-rover alignment. Drone detection using YOLOv8 and a tilt correction algorithm was integrated for the proposed visual servo of the ground robot. The study collected 3024 images as a data set for drone detection. The experimental results show that the trained instance segmentation model detected and captured drone objects. The study conducted an initial test for visual servo control of the ground robot in various surface terrains, resulting in a maximum alignment error on rough surfaces. Furthermore, the study conducted drone-ground robot alignment real test in an outdoor field setting. The alignment between the drone and the ground robot produced a maximum alignment error of 20.3 cm, below the threshold error. The open field experiments verified the effectiveness of the ground robot's visual servo control with an actual drone operation.**

***Index Terms*—Visual servoing, vision-based navigation, field robots, cooperating robots, object detection, segmentation and categorization, computer vision for automation.**

Received 23 January 2025; accepted 21 April 2025. Date of publication 12 May 2025; date of current version 26 May 2025. This article was recommended for publication by Associate Editor E. Malis and Editor P. Vasseur upon evaluation of the reviewers' comments. This work was supported in part by the Philippine Council for Industry, Energy, and Emerging Technology Research and Development (PCIEERD) and in part by the Philippine Institute of Volcanology and Seismology (PHIVOLCS). *(Corresponding author: Carl John Salaan.)*

Sean Clark Dominguez, Jeanette Pao, Immanuel Paradela, John Mel Bolaybolay, Earl Ryan Aleluya, Francis Jann Alagon, Sherwin Guirnaldo, and Carl John Salaan are with the College of Engineering, Mindanao State University–Iligan Institute of Technology, Iligan City 9200, Philippines (e-mail: seanclark.dominguez@g.msuiit.edu.ph; jeanette.pao@g.msuiit.edu.ph; immanuel.paradela@g.msuiit.edu.ph; johnmel.bolaybolay@g.msuiit.edu.ph; earlryan.aleluya@g.msuiit.edu.ph; francisjann.alagon@g.msuiit.edu.ph; sherwin.guirnaldo@g.msuiit.edu.ph; carljohn.salaan@g.msuiit.edu.ph).

Kazunori Ohno and Yoshito Okada are with the Graduate School of Information Sciences, Tohoku University, Sendai 980-8577, Japan (e-mail: kazunori@rm.is.tohoku.ac.jp; yoshito@rm.is.tohoku.ac.jp).

Argel Bandala is with the Gokongwei College of Engineering, De La Salle University, Manila 1004, Philippines (e-mail: argel.bandala@dlsu.edu.ph).

Abu Ubaidah Bin Shamsudin is with the Faculty of Electric and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat 86400, Malaysia (e-mail: ubaidah@uthm.edu.my).

This article has supplementary downloadable material available at https://doi.org/10.1109/LRA.2025.3569125, provided by the authors.

Digital Object Identifier 10.1109/LRA.2025.3569125

I. Introduction

OVER the years, researchers have been developing different types of robot for harsh and risky environments, such as the volcano environment [1], [2], [3]. These robots observed changes in activities and collected sample data from within risky environments. In some cases, researchers have applied drone-ground robot collaboration to their systems. Researchers improved typical exploration systems by using drones to help the ground robot bypass rugged and dangerous areas. It allowed rapid access to isolated areas for the deployment and exploration of ground robots [4].

The most challenging part of the drone-ground robot system is retrieving the ground robot in harsh and risky areas [5], [6]. Retrieval of specific objects such as ground robot can be made by having a retrieval mechanism attached in the drone [7], [8]. However, the process of retrieving the ground robot should start with aligning the drone and ground robot properly, which is the crucial part of the retrieval operation.

Different researches often involve manually guiding the ground robot to the drone's hovering position for initial alignment. Visual servo control can also be applied for aligning drones and ground robot in facilitating the ground robot recovery operation. However, GPS navigation systems usually produce substantial position errors for drone-ground robot alignment. While aerial visual servo control can implement the alignment using a drone that continuously adjusts its movements, this strategy poses a risk hazard for drone flights, notably with the effects of wind gusts and drone drifting, which produce substantial position errors. Hence, the ground robot can implement visual servo control to compensate for the errors and risks associated with aerial visual servo control.

In [9], a visual servo technique is introduced for the ground robot using an ArUco marker for drone alignment. To further enhance and improve the system (as shown in Fig. 1), this study investigated a visual servo control for the ground robot with drone detection and tilt correction. This particular paper presents the following scientific contributions:

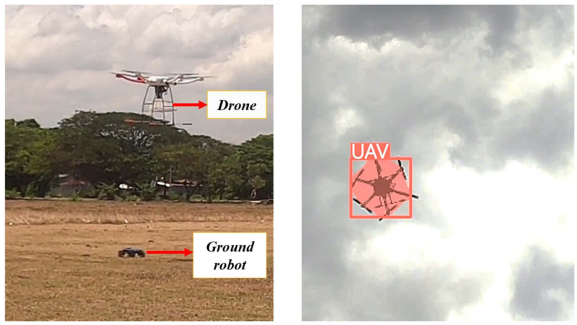

Fig. 1. The drone-ground robot alignment procedure in an uneven surface through ground robot visual servoing (a) with simultaneous tilt correction and (b) drone segmentation and detection in a cloudy weather condition.

1) Integrated YOLO-based drone detection and tilt correction as part of ground robot visual servoing,
2) Contributed application-specific dataset for drone detection under varying weather conditions,
3) Implemented a functional ground robot visual servoing on different ground surface conditions (uneven and rough terrains), and
4) Demonstrated actual ground robot retrieval operation by implementing ground visual servoing technique.

II. Related Literature on Visual Servo Control

In recent years, researchers have applied visual servo control in conditions requiring minimal human supervision or autonomous robot movement. In these scenarios, robots heavily rely on visual data input. Hence, researchers employ a variety of image processing techniques to detect and locate their targets [10].

In [11], an aerial drone is programmed to land on a moving ground vehicle autonomously. A circular-patterned landing deck is mounted and the drone could approach and land above the moving vehicle by detecting the pattern. Since the technique relies heavily on visual data, different methods of image pre-processing are employed to detect and locate visual markers within an image. In [12], a technique for operating a drone that involves tracking and following a marked block is introduced. Although the study accomplished image-based visual servo control, the authors encountered certain obstacles related to the intensity of light and the background of the marker.

To address the problems of poor lighting conditions, the studies [13], [14], [15], [16] implemented deep learning object detection using You Only Look Once (YOLO) network to detect the drone in varying light conditions. Despite the high detection accuracy, the ground robot's actions accumulate errors in the real-world exploration. This challenge is due to natural variables such as the tendency of wheel slipping. This paper presents a tilt correction method with drone detection technology to deal with this problem. In such a way that the Rover can withstand the issue of an uneven terrain.

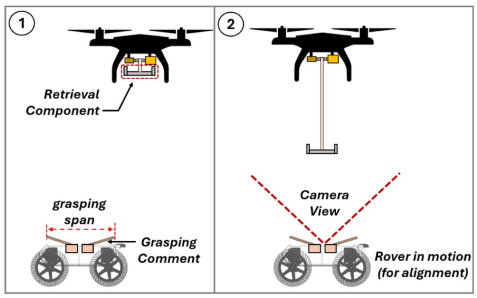

Fig. 2. The general procedure of drone-ground robot alignment: (1) The drone stays at specific height while the ground robot performs visual servoing and aligns to the drone center, (2) After successful alignment, the drone's retrieval component goes down and retrieves the ground robot.

III. Design Considerations for Ground Robot Visual Servo Control

For ground robot operating in harsh and risky environments, the drone will fly to the ground robot's position using GPS coordinates for retrieval. Once in position, the ground robot's onboard camera will confirm that the aerial drone is within its range. The ground robot will then perform visual servo control for drone-ground robot alignment.

The alignment of the drone and ground robot allows the drone's retrieval component and the ground robot's grasping component to function effectively, as depicted in Fig. 2. While the retrieval component is descending, the ground robot will consistently implement visual servo control to adjust its position, considering external factors: drone drifting and wind. It ensures the retrieval component and grasping component are aligned as well, and the grasping component of the ground robot can then grasp the retrieval component effectively.

A. Terrain Conditions

The design requirements of this study considered a predetermined set of conditions for the ground robot retrieval operation. The conditions dictate that the terrain for ground robot retrieval should have the following characteristics:

1) soil should be stable,
2) absence of large obstacles, and
3) absence of elements obscuring ground robot and drone camera vision.

Though these are ideal conditions for ground robot retrieval, this study assumed imperfections in surface terrains. These imperfections include rocks, holes, and gaps that may subtly tilt the ground robot platform, consequently affecting the ground robot's visual servo control.

B. Threshold Error Consideration

The ground robot's grasping component has a maximum grasping span depending on the grasping component size (refer to Fig. 2). For successful docking, the ground robot must align its grasping component with the retrieval mechanism of the drone. Once properly aligned, the grasping component rotates inward to securely latch onto the drone's retrieval system. Therefore, precise alignment between the drone and the ground robot is essential, especially as the retrieval component descends toward

the ground robot. This ensures a smooth and secure engagement between the two systems, minimizing errors in the grasping process. Lastly, the heading angle tolerance is set to ±5-degree error.

1) Considerations on GPS Errors: For retrieving the ground robot, the drone must be positioned at the same location. In this manner, the drone must be set to the same GPS location as the ground robot. However, GPS-based drone waypoint navigation accumulates a margin of error around the destination coordinate. In the imprecise navigation causes alignment errors between the drone and the ground robot.

There are two approaches to correct GPS-induced position errors: aerial visual servo control or ground robot visual servo control. In real-world scenarios, strong winds and drone drift bring a significant deviation from its expected location. Visual servoing performed by the drone can compensate for this deviation. However, the challenge lies in the unpredictable lateral motion and wobbling of the drone.

The visual servo control of the ground robot works effectively when the drone is in its field of view. However, when the drone is outside the field of view or the ground robot moves beyond the drone's circumferential range, the ground visual servo control operates in standby mode until the drone is in the field of view again. In standby mode, the ground robot does not move. Its navigation component is set to standby while the drone detection program still operates. In the case where both the drone and ground robot are away from each other, the operators at the base station manually set the new destination of the drone to the ground robot's GPS location. In this manner, the whole retrieval operation resets, and the visual servoing will re-initialize.

2) Considerations on Drone Recognition: Hence, to compensate the errors and risks from the aerial visual servo control, the ground robot implements visual servo control, which tracks the drone. By using the drone instance segmentation model for drone detection, the drone indicates its position and becomes visible from the ground robot's perspective.

Three scenarios contribute to unstable drone recognition:

1) cloudy and direct sunlight weather
2) an increased drone elevation
3) unique object or subject within the camera frame

It is important to note that the ground robot can still operate under these conditions. Even with a single frame, the ground robot will still move if it recognizes the drone. In the case of unfavorable weather conditions, the drone will still be detected with low confidence score. When a unique object or subject enters the camera frame, the detection program chooses the object with the highest confidence value to ensure that the detected object will always be a drone. Hence, in this setup, the ground robot will consistently detect the drone and implement visual servo control to adjust its position.

3) Drone Health During Operations: It is also important to note that the drone's autopilot system can monitor its overall status during the visual servo control operation. Any unpredicted behavior may trigger the drone to initiate backup emergency flights, such as returning to the base station.

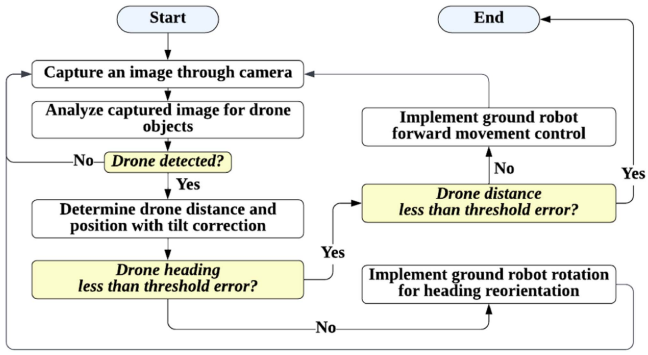

Fig. 3. General flowchart for combined application of object detection and tilt correction through visual servo control to achieve drone-ground robot alignment.

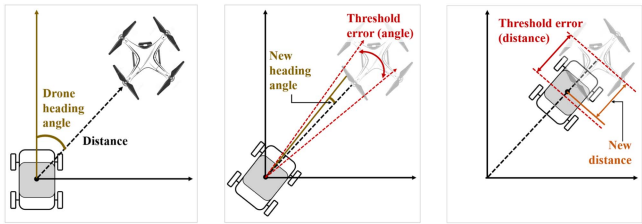

Fig. 4. (a) The ground robot will calculate the drone's position, (b) adjust its heading towards the direction of the drone, and move forward. (c) It will repeat the process until the ground robot is less than threshold error.

In this study, the researchers focused on the ground robot's implementation of visual servo control for alignment. The drone and ground robot do not directly communicate; the only connection between them is the camera attached to the ground robot. However, the operators at the base station can independently communicate the onboard devices at the drone and the ground station.

C. General Procedure for Ground Robot Visual Servo Control

As illustrated in Fig. 3, the study implemented a visual servo control program, which includes instance segmentation, image processing, and motor control. Fig. 4 illustrates the concept of threshold error and ground robot movement for visual servo control.

The process starts from capturing the drone image through the upward-facing camera mounted on the ground robot. Then, the program scans the image to detect the drone using OpenCV. If the drone is detected, then the program calculates the drone's distance and heading angle relative to the ground robot. The Rover uses this information to reorient its position towards the drone. If the drone's distance exceeds the error threshold, the ground robot will move forward at a distance equal to the drone's distance from the ground robot. Afterward, the ground robot recalculates the drone's position and repeats the process until the distance error threshold is reached.

IV. Drone Detection Through Instance Segmentation

Segmentation is the process of dividing an image into distinct regions based on particular criteria or other visual features,

which may be evaluated separately. There are three types of segmentation [17], [18]: semantic segmentation, which divides pixels in an image into broad categories, instance segmentation, which recognizes particular instances of objects inside an image, and panoptic segmentation that simultaneously assigns semantic labels to all pixels in an image while distinguishing between objects.

A. Selected Segmentation Model

Convolutional Neural Networks (CNNs) offer these capabilities, with input and output layers that accept raw data and provide final predictions. CNN model provides lightweight options for implementation on small companion computers and offers a segmentation extension is required for this study. Studies [13] and [19] used CNN models from the YOLO architectural family, specifically YOLOv8, for their light size and segmentation capabilities. Thus, in this study, the segmentation extension and the narrow variation of the YOLOv8n-seg model, the lightest model, were used for drone detection as part of ground robot visual servoing.

B. Gathered Drone Dataset

Using the ground robot's camera, the authors collected images of the drone at a hovering state. To employ data diversity, the authors consider various weather conditions and different heights of the drone with respect to the ground. The final dataset consists of 3,024 annotated photos. The authors split the dataset into training, validation, and test sets at a 7:2:1 ratio. They provided stratified a random sampling strategy to guarantee equal distribution of images from various weather and height conditions. The dataset used for training drone detection under various weather conditions is available in [20].

C. Instance Segmentation Training

The study chose to apply Adam optimizer during the training stage. This optimizer suits the task since it requires low memory, converges quickly, and adjusts learning rates based on past gradients. In addition, it fits the requirements in this study, considering the unique drone features and small-sized dataset.

For training, the Adam optimizer is used with the following hyperparameters: an initial learning rate (lr0) of 0.01, a learning rate factor (lrf) of 0.01, a momentum of 0.937, a weight decay of 0.0005, and a warm-up momentum of 0.8. The batch size is set to 4 to maximize the GPU capacity of the personal computer (NVIDIA GeForce RTX 2070 GPU) with 8 GB memory, The training is set to run in 100 epochs, but a patience is set to 10 to terminate training once overfitting occurs 10 times. Lastly, the input images are forcibly resized into 640 x 640, and three channels for the RGB color space. This image shape is a requirement of the YOLOv8-seg model's input layer.

D. Training Results and Evaluation

The model effectively learned from the dataset which obtained a training loss of 0.01106 and a validation loss of 0.00941 at epoch 59. The Detection Error Tradeoff (DET) curve is shown in Fig. 5. At FPR=0, FNR is approximately 0.0004, indicating that when the classifier does not make false positive errors, it has a false negative rate of 0.04%, indicating that the classifier misses approximately 0.04% of actual positives. The performance of the drone detection under different conditions resulted in low FNR results, which means that the detection model is effective and rarely misses the drone.

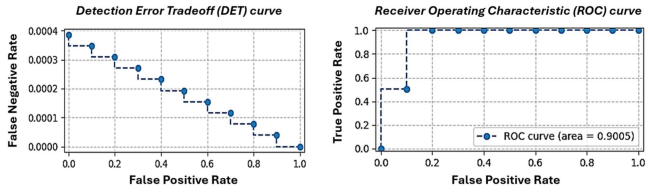

Fig. 5. The Detection Error Tradeoff (DET) curve of the drone detection model under unstable drone recognition scenarios.

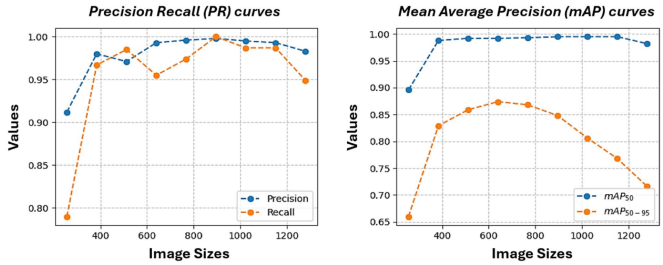

Fig. 6. Result of the trained instance segmentation model from the test set in terms of precision, recall, MAP_{50}, and $MAP_{50\text{-}95}$.

The study evaluated the model's performance several times using various input image sizes, as shown in Fig. 6. Precision, recall, and mAP50 peaked at an image size of 896 and remains stable after an image size of 896 retaining the high level of true positive detections. mAP50-95 increases from an image size of 256 until 640, but decreases after an image size of 640. Considering all the metrics, the best image size for using this model appears to be at around 768. At an image size of 768, precision is almost at its peak with a value of 0.996, a high recall of 0.974 and strong values of mAP50 and mAP50-95 (0.993 and 0.868, respectively).

V. Image-Based Distance and Heading Calculations

For the visual servo control to work, the ground robot must determine the position of the detected drone object; this includes the horizontal distance between the drone and the ground robot and the drone angle based on the ground robot's current heading (where the ground robot is facing). The authors determined the drone's horizontal distance from the ground robot using the concept of pinhole projection. This concept uses a series of formulas based on ratios and proportions to determine the size of an object within an image [21]. In this study, (1) is used to determine the drone's distance from the ground robot.

$$d_o'' = d_i'' \frac{h_o'}{h_i'} \tag{1}$$

where:

- d_o'' - drone center distance from the ground robot body.
- d_i'' - drone center distance from the image center.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE I
GROUND ROBOT DISTANCE AND HEADING CALCULATION RESULTS

Calculation Approach	Parameter	Error
Without Offset	Distance	5.91 ± 3.47 cm
	Heading	5.79 ± 4.15 deg
With Offset	Distance	2.16 ± 1.42 cm
	Heading	2.16 ± 1.94 deg

- h'_o - drone size in cm.
- h'_i - drone size in pixels.

In addition, (2) determines the ground robot's target heading, ϕ, using the Cartesian components of the estimated drone center distance, d''_o.

$$\phi = \arctan\left(\frac{d''_{o^y}}{d''_{o^x}}\right) \tag{2}$$

where:

- ϕ - target heading for ground robot rotation.
- d''_{o^x} - x component of estimated drone center distance.
- d''_{o^y} - y component of estimated drone center distance.

Additionally, to improve the accuracy of the calculations, the study added offsets to consider errors that are inherent to ground robot hardware. The offset values will change whenever the ground robot is altered. As depicted in the distance update rule in (3) and (4), the calculations incorporated the offset values to the Cartesian components d''_{o^x} and d''_{o^y}.

$$d''_{o^x} \leftarrow d''_{o^x} + x_{offset} \tag{3}$$

$$d''_{o^y} \leftarrow d''_{o^y} + y_{offset} \tag{4}$$

where:

- x_{offset} - offset relative to ground robot's x-coordinate.
- y_{offset} - offset relative to ground robot's y-coordinate.

Based on the results summarized in Table I, the researchers determined that the Cartesian offsets from the image-based calculations reduced both the distance and heading errors. Through this initial results, it was found that other than linear position, angular position should be evaluated also for drone-ground robot alignment. Chapter VI discusses the alignment errors due to induced tilts and how to correct it.

VI. VISUAL SERVOING WITH TILT CORRECTION METHOD

In this study, the ground robot must adapt to slightly uneven surfaces containing rocks and potholes, as this causes the ground robot to tilt (Fig. 7) and contribute to distance and heading error. The induced tilt angle distorts drone center distance calculations. Hence, tilt correction to compensate for tilt-induced errors was evaluated considering the image-based distance and heading calculations.

Considering the case illustrated in Fig. 7 where the drone is placed directly above the ground robot at a distance *Height*, the tilt induced by an uneven surface causes the drone center distance calculation to perceive that it is at a distance *Tilt Error* from the ground robot. Using the cosine angles and the magnitude of vector $\overrightarrow{\mathbf{T}}$, the Cartesian components of the tilt error are presented

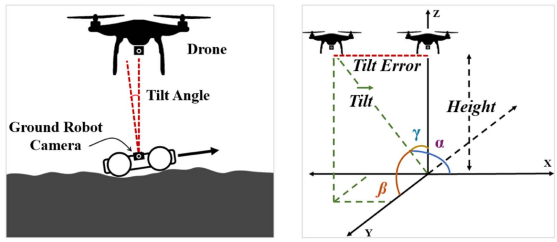

Fig. 7. (a) Uneven surfaces induce tilts to the ground robot, causing it to perceive the drone to be farther than its actual position. (b) The graphical illustration of correcting the tilt error induced.

in (5) and (6).

$$x_{tilt_error} = \left\|\overrightarrow{\boldsymbol{T}}\right\| \cos\alpha \tag{5}$$

$$y_{tilt_error} = \left\|\overrightarrow{\boldsymbol{T}}\right\| \cos\beta \tag{6}$$

where:

- $\overrightarrow{\mathbf{T}}$ - magnitude of vector *Tilt*.
- x_{tilt_error} - tilt error relative to the x-axis.
- y_{tilt_error} - tilt error relative to the y-axis.
- $\cos\alpha$ - cosine angle from the x-axis.
- $\cos\beta$ - cosine angle from the y-axis.

Then, the Cartesian components of $TiltError$ were introduced to (3) and (4) as tilt error-reducing variables, thus producing the distance update rule in (7) and (8).

$$d''_{o^x} \leftarrow d''_{o^x} + x_{offset} + x_{tilt_error} \tag{7}$$

$$d''_{o^y} \leftarrow d''_{o^y} + y_{offset} + y_{tilt_error} \tag{8}$$

In this study, (7) and (8) are used to calculate the position of the drone while also compensating for the tilt-induced errors caused by uneven terrains, thus improving the accuracy of distance and heading calculations. The study evaluated the accuracy of error reduction through tilt correction. To replicate tilts caused by rocks and potholes, researchers inclined the ground robot for about 10 degrees on average.

As for its specifications, the IMU sensor uses a 9-degree-of-freedom sensor fusion system that combines three separate sensors for reading orientation, acceleration, and magnetic fields in three-dimensional space. The sensor uses the built-in gyroscope sensor to provide information about the ground robot's orientation. The researchers determined the offset values for the calibration of the IMU sensor and camera by collecting 32 sample data. Then, the researchers used the resulting mean averages as calculation offsets.

The researchers implemented distance and heading calculations with tilt error correction using the offset values and the measured Euler angles. Here, the study conducted 60 test trials, summarized in Table II. The error convergence speed for both instances (with or without tilt correction) depends on the robot's processing speed; hence, they are similar but differ in the final size of the distance error.

Commonly used strategies for ground robot visual servoing do not contain tilt-error correction. Thus, Table II shows the comparison on ground robot visual servoing with

TABLE II
TEST RESULTS FOR GROUND ROBOT VISUAL SERVOING WITH AND WITHOUT TILT ERROR CORRECTION

Calculation Approach	Parameter	Error
Without Tilt Correction	Distance	18.19 ± 14.66 cm
	Heading	23.49 ± 17.28 deg
With Tilt Correction	Distance	5.51 ± 3.61 cm
	Heading	7.20 ± 5.37 deg

and without tilt-error correction, in which visual servoing with tilt correction resulted in lesser distance and heading errors.

VII. RESULTS AND DISCUSSION

This study used a ground robot with a body of 34 cm in length, 18 cm in width, 7 cm in height, ground clearance of 11 cm, and a wheel diameter of 17 cm. An open-field test was performed with actual drone operation for drone-ground robot alignment.

The ground robot has a grasping component with a maximum grasping component span of 22 cm (see Fig. 2). Thus, the threshold distance error in this ground robot setup will be 22 cm. The threshold distance error can be adjusted depending on the ground robot platform dimension. If the drone's distance exceeds the threshold, the ground robot will not be captured by the drone's retrieval mechanism. To avoid this, the ground robot reorients its position and repeats the iteration until it reaches the drone's center mark.

A. Ground Robot Initial Alignment Test

The study evaluated the ground robot's initial performance in visual servo control through 3 marker alignment tests: a smooth surface test with a 3m-high marker, a rough surface test with a 3m-high marker, and a rough surface test with a 5m-high marker. The average distance errors for the smooth surface test, rough surface test (3m-high marker), and rough surface test (5m-high marker) are 3.39± 1.71cm, 7.93± 2.49cm, and 7.11±3.08 cm, respectively. The researchers noted that all the average errors were within the threshold distance error of 22 cm. The tests affirmed that drone and ground robot alignment is possible in smooth and rocky terrains.

Thus, another alignment tests in an open field with actual drone operation was conducted to evaluate the effectivity of the ground robot visual servoing. This alignment test includes ground robot visual servoing; 1) on an uneven surface with small rocks and potholes, 2) with drone predefined paths, and 3) with actual retrieval operation of the ground robot.

B. Implementation of Drone Instance Segmentation for Drone Detection

The visualization outputs for drone detection in the chosen test images are displayed in Fig. 8. The predicted outputs depict drone detection with high confidence and consistent masking. The evaluation metrics, such as precision, recall, and mAP as previously presented in Fig. 6, support the validation of the predicted outputs. In addition, no external objects were detected beside the drone since the camera is directed upward to the sky. Thus, there is a lower chance of getting false detections. This likelihood is supported by the low false positive rate as previously shown in Fig. 5. The input images were obtained in different weather conditions and at various heights. In Fig. 8, it was observed that all the models successfully captured the polygon shape of the drone. In summary, the instance segmentation models consistently detected the drone object.

(a) (b)

Fig. 8. Drone segmentation outputs under the worst environmental conditions: (a) Direct sunlight exposure affects the visibility of the drone. (b) Cloudy skies add complexity to the environment due to patches that create background irregularities.

C. Implementation of Ground Robot Visual Servo Control With the Actual Drone Operation

The actual drone operation was conducted at an open test area in Iligan City, which has a rocky and uneven landscape. Recorded wind gusts during the drone operations were up to 1.5 m/s, with cloudy weather and direct sunlight exposure. The camera used has a USB connection which allowed faster transmission speed. The ground robot's camera is dedicated to the retrieval and drone detection system, making it unavailable for recording video for the robot itself. The program code for retrieval system operates using multiprocessing, where it processes the camera output as its input.

1) Uneven Surface With Small Rocks and Potholes: The ground robot visual servo control was initially tested on a rough surface with small rocks and potholes (with 12 implementations of drone-ground robot alignment). X-shaped markers of 2 meters apart are marked on the ground. These markers are where the drone was hovered in position of 5-meter height.

The drone on-board camera confirms its hovering position at the markers. The ground robot successfully implemented the visual servo control and aligned below the drone in a rough surface with small rocks present. The ground robot initially at starting position ($t = 0$ s) moved below the drone's first hovering position at point A and aligned with the drone for $t = 20$ s) as shown in Fig. 9. The average alignment error recorded is 17.6 cm that is within the threshold error.

2) With Drone Predetermined Waypoints: The drone-ground robot alignment was implemented with applying drone waypoints or predetermined paths as depicted in Fig. 10 indicated by green markers (1, 2, 3 and 4). At drone hovering at point 2, the ground robot successfully implemented the visual servo control from its initial position near the point 1 to the point 2 ($t = 20$ s). Consecutively, the drone hovering at point 3, the rover

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

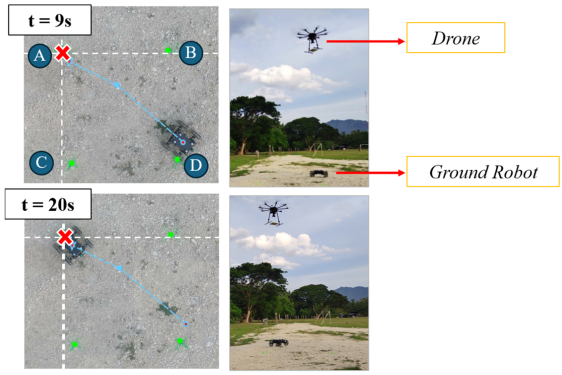

Fig. 9. Implementation of ground robot-drone alignment using visual servo control in a rough uneven terrain with presence of small rocks. The drone hovers at point A with its center indicated by red X-shaped marker while the ground robot performs visual servo control and aligns with the drone's center at t = 20 s. The ground robot's movement is indicated by blue dashed lines.

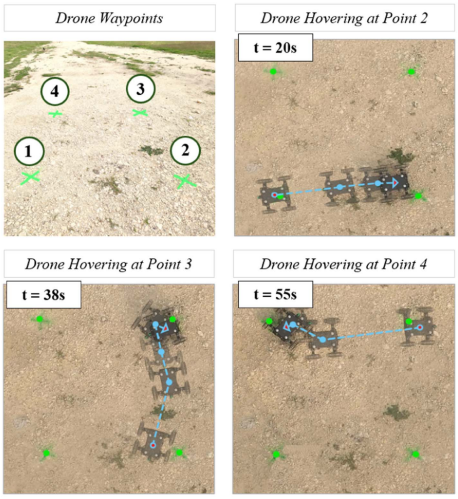

Fig. 10. Implementation of drone-ground robot alignment with drone predetermined waypoints. The ground robot followed the drone using visual servo control as the drone hovered from one point to another.

TABLE III
SUMMARY OF DRONE-GROUND ROBOT ALIGNMENT TESTS FOR DRONE PREDETERMINED PATHS

	Average Alignment Error	
Path 1	19.58 cm	below the
Path 2	20.30 cm	threshold
Path 3	17.80 cm	error
Path 4	6.33 cm	(of 22cm)

implemented again the visual servo control from point 2 to point 3 (t = 38 s).

The researchers recorded the corresponding alignment errors, which are summarized in Table III. The results showed that the maximum average distance error for the alignment is 20.30 cm, and is below the threshold error of 22 cm. Thus, for application of predetermined waypoint paths for the drone, the ground robot effectively implemented visual servo control considering the terrain condition, drone drifting and winds.

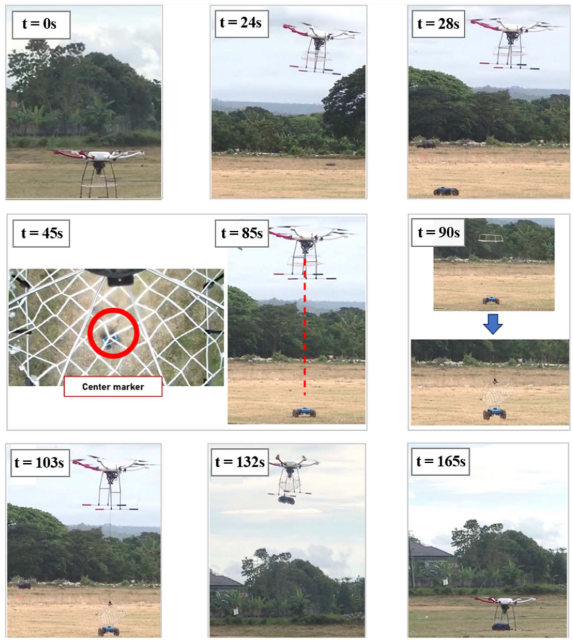

Fig. 11. Implementation of drone-ground robot alignment with actual retrieval operation of the ground robot. The drone hovers up to 3 meters and goes to the retrieval area. Then, the ground robot initially performs visual servoing (t = 28 s). The drone stays at a 3-meter height while the ground robot performs visual servoing and aligns to the drone center, as shown in the onboard camera view of the drone (t = 85 s). After successful alignment, the drone's retrieval component goes down (t = 90 s) and retrieves the ground robot through its grasping component (t = 103 s).

3) With Actual Ground Robot Retrieval Operation: Lastly, the drone-ground robot alignment was implemented with the actual retrieval operation of the ground robot as depicted in Fig. 11. The ground robot is positioned 50 meters away from the drone's hovering area. After the drone goes to the retrieval area where the ground robot is to be retrieved (t = 24 s), the simultaneous application of drone detection and tilt correction for visual servo control of the ground robot initializes (t = 28 s).

Both vehicles are remotely connected to the base station, allowing operators to autonomously control the drone and assign a new position based on the ground robot's GPS location. Since both vehicles utilize GPS sensors with an RTK configuration, this ensures that the drone remains within the field of view of the ground robot's camera. The visual servo control strategy is activated once the ground robot detects the drone within a specified confidence interval.

The average alignment error recorded is 16.58 cm for 6 field tests. After the alignment, the drone successfully retrieved the ground robot through its mechanism.

Furthermore, a stability analysis in Fig. 12 visualizes the alignment error of the ground robot as it approached the aerial drone on another three field tests. The performance curves for three tests verified the functionality of ground robot visual servoing for different test conducted. The results showed that the overall average alignment error is below the threshold value (22 cm). Convergence was achieved within 18 seconds in Test 1, 14 seconds in Test 2, and 13.5 seconds in Test 3.

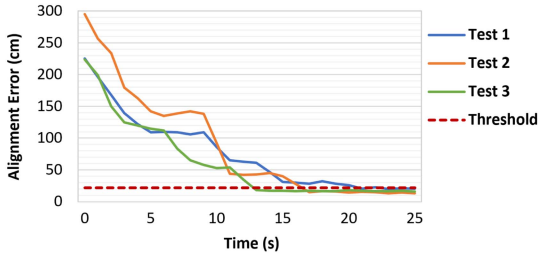

Fig. 12. Stability analysis of ground robot visual servo control for three different actual retrieval tests conducted. Drone-ground robot alignment was made during t = 20 s to t = 25 s.

VIII. Conclusion and Recommendations

For ground robot retrieval applications in harsh and risky environments such as volcanoes, this study implemented visual servo control for drone-ground robot alignment. The visual servo control strategy implemented multiple programs running simultaneously, including drone detection and tilt correction.

In this study, the ground robot used has a grasping span that resulted to threshold error of 22 cm. When the ground robot was subtly tilted in various positions, applying tilt correction reduced the average distance and heading calculation errors to 5.51 cm and 7.20 degrees, respectively. And for drone detection, the instance segmentation models consistently detected the drone object and implemented the visual servo control for alignment.

The authors conducted open field tests to assess the alignment performance with the actual drone operation. The drone-ground robot alignment on uneven surfaces containing rocks and potholes resulted in an average of 17.60 cm alignment error. Consecutively, a 20.30 cm maximum alignment error was found on drone-ground robot alignment implementation with drone predetermine path set-up. Further evaluation was made with the actual retrieval operation of the ground robot. The average alignment error recorded is 16.58 cm, and the drone with its retrieval component successfully retrieved the ground robot through its retrieving mechanism. Overall, the average alignment errors are within the threshold error. And through stability analysis, the functionality of ground robot visual servoing for different test conducted are verified.

The results implied that visual servo control through simultaneous drone detection and tilt correction effectively aligned the drone and ground robot under open outdoor field conditions. To summarize, the ground robot effectively implemented visual servo control and successfully aligned with the drone:

1) on uneven surfaces containing rocks and potholes
2) with actual ground robot retrieval operation, and
3) under cloudy and direct sunlight with drone drifting and winds.

Future work includes adding more tests to further investigate the system's performance, implementing an advanced strategy for visual servo control, and conducting actual operations in complex environments.

References

[1] B. McGuire, C. R. Kilburn, and J. Murray, *Monitoring Active Volcanoes: Strategies, Procedures and Techniques*. New York, NY, USA: Taylor & Francis, 2022.

[2] G. Muscato, F. Bonaccorso, L. Cantelli, D. Longo, and C. D. Melita, "Volcanic environments: Robots for exploration and measurement," *IEEE Robot. Automat. Mag.*, vol. 19, no. 1, pp. 40–49, Mar. 2012.

[3] J. C. Pao, C. A. G. Banglos, C. J. O. Salaan, and S. A. Guirnaldo, "Simulation and strength analysis of flippable mobile ground robot for volcano exploration and monitoring application," in *Proc. IEEE 14th Int. Conf. Humanoid Nanotechnol. Inf. Technol. Commun. Control Env. Manage.*, IEEE, 2022, pp. 1–6.

[4] K. Nagatani et al., "Development and field test of teleoperated mobile robots for active volcano observation," in *Proc. 2014 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, IEEE, 2014, pp. 1932–1937.

[5] M. Laiacker, M. Schwarzbach, and K. Kondak, "Automatic aerial retrieval of a mobile robot using optical target tracking and localization," in *Proc. 2015 IEEE Aerosp. Conf.*, IEEE, 2015, pp. 1–7.

[6] R. Moeller, T. Deemyad, and A. Sebastian, "Autonomous navigation of an agricultural robot using RTK GPS and Pixhawk," in *Proc. 2020 Intermountain Eng. Technol. Comput.*, IEEE, 2020, pp. 1–6.

[7] F. Ruggiero, V. Lippiello, and A. Ollero, "Aerial manipulation: A literature review," *IEEE Robot. Automat. Lett.*, vol. 3, no. 3, pp. 1957–1964, Jul. 2018.

[8] G. A. Yashin, D. Trinitatova, R. T. Agishev, R. Ibrahimov, and D. Tsetserukou, "Aerovr: Virtual reality-based teleoperation with tactile feedback for aerial manipulation," in *Proc. 19th Int. Conf. Adv. Robot.*, IEEE, 2019, pp. 767–772.

[9] S. C. U. Dominguez, J. A. Bolaybolay, E. R. M. Aleluya, and C. J. O. Salaan, "Ground rover visual servo control for drone alignment," in *Proc. IEEE 14th Int. Conf. Humanoid, Nanotechnol., Inf. Technol. Commun. Control Environ. Manage.*, IEEE, 2022, pp. 1–6.

[10] F. Chaumette, "Visual servoing," in *Computer Vision: A Reference Guide*. Berlin, Germany: Springer, 2021, pp. 1367–1374.

[11] A. Keipour et al., "Visual servoing approach to autonomous UAV landing on a moving vehicle," *Sensors*, vol. 22, no. 17, 2022, Art. no. 6549.

[12] T. V. Venna, S. Patel, and T. Sobh, "Application of image-based visual servoing on autonomous drones," in *Proc. 15th IEEE Conf. Ind. Electron. Appl.*, IEEE, 2020, pp. 579–585.

[13] J. M. A. Bolaybolay et al., "YOLO instance segmentation model comparison for drone detection as visual servo control marker," in *Proc. IEEE 15th Int. Conf. Humanoid Nanotechnol. Inf. Technol., Commun. Control Environ. Manage.*, IEEE, 2023, pp. 1–5.

[14] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, "UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios," *Sensors*, vol. 23, no. 16, 2023, Art. no. 7190.

[15] J. M. L. Pepito et al., "Ground robot detection for collaborative UAV-UGV alignment and retrieval operations," in *Proc. IEEE 15th Int. Conf. Humanoid Nanotechnol., Inf. Technol. Commun. Control Environ. Manage.*, IEEE, 2023, pp. 1–6.

[16] K. Sricharan and M. Venkat, "Real-time drone detection using deep learning," in *Proc. 2nd Int. Conf. Emerg. Trends Eng.*, Atlantis Press, 2023, pp. 905–918.

[17] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: State of the art," *Int. J. multimedia Inf. Retrieval*, vol. 9, no. 3, pp. 171–189, 2020.

[18] Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao, "Review the state-of-the-art technologies of semantic segmentation based on deep learning," *Neurocomputing*, vol. 493, pp. 626–646, 2022.

[19] X. Yue, K. Qi, X. Na, Y. Zhang, Y. Liu, and C. Liu, "Improved YOLOv8-seg network for instance segmentation of healthy and diseased tomato plants in the growth stage," *Agriculture*, vol. 13, no. 8, 2023, Art. no. 1643.

[20] S. C. Dominguez et al., "UAV detection dataset," 2025. [Online]. Available: https://doi.org/10.7910/DVN/ENLPOJ

[21] E. Ustunel and E. Masazade, "Vision-based road slope estimation methods using road lines or local features from instant images," *IET Intell. Transport Syst.*, vol. 13, no. 10, pp. 1590–1602, 2019.