國立臺灣大學生物資源暨農學院生物機電工程學系

碩士論文

Department of Biomechatronics Engineering
College of Bioresources and Agriculture
National Taiwan University

Master's Thesis

應用於溫室環境作物監測之自主導航無人機多機協同系統

An Autonomous Multi-UAV Cooperative Navigation System for Crop Monitoring in Greenhouse Environments

徐滋

Tzu Hsu

指導教授：林達德 博士

Advisor: Ta-Te Lin, Ph.D.

中華民國 114 年 07 月

July 2025

doi:10.6342/NTU202503732

誌謝

在大學畢業後準備來碩班就讀的我總是擔心自己有許多的不足之處，擔心自己是否準備好也擔心自己是否能力足夠，在碩班這兩年我想先謝謝我的指導教授林達德老師，由衷的感謝老師收我為學生，我在許多方面得到老師的教導，不管是學術上的知識、研究上的嚴謹以及待人處世上，我都受益良多，謝謝老師總是不厭其煩的給予研究上的指導讓我們突破自己的難關，老師就像是海上的燈塔不讓我們在汪洋大海中迷失，謝謝老師的栽培加入老師的實驗室成為老師的研究生是我最大的福氣。謝謝我的口試委員顏炳郎老師以及楊江益老師，在我的論文上給予精闢的指導，提出許多寶貴的建議以及給予肯定。在 405 實驗室的日子是我這一輩子很難忘的回憶，敬恆是我在研究上的前輩也是好夥伴，每當我遇到困難時總是會在絕境施予援手，在需要於溫室進行實驗也給予我許多的協助和精神上以及技術上的支持，感謝你一直以來的耐心與無私幫助，碩班能遇到你真的很幸運也很快樂！在每天的日常中謝謝姵瑜和柏勳在一路上的互相扶持，每當遇到艱難的時光總是會跟彼此信心喊話，謝謝你們的幫助與包容，謝謝庭瑄、冠辰、稚晴、澔平給予許多協助與關心，跟你們相處總是很有趣，希望你們繼續堅持保持開朗，謝謝 Bob 隋中或在實驗室一起打拼的日子總是很有趣，謝謝黃天洞像個開心果讓生活充滿樂趣，在碩班認識你們讓平凡的日子變得不平凡，謝謝學長晨宇、易霖和喬尹在我們懵懵懂懂碩一時給予我們許多幫助與教導，謝謝軍諺把題目交給了我在碩班初期幫助我進入狀況，謝謝桂芝辛苦地幫我們報帳還會在平時照顧我們，謝謝我的男友在過程中不斷給我精神上的支持，最後謝謝我的家人，這兩年讓你們擔心了，謝謝你們總是給我無限的關愛與支持，最後謝謝堅持下去沒有放棄的自己。

doi:10.6342/NTU202503732

摘要

本研究開發了一套自主多機無人機系統，旨在進行溫室中洋香瓜作物的生長監測。透過多機無人機的協同作業，進行了多角度的影像拍攝，並設計了三種不同的飛行路徑。利用 UWB 定位系統進行無人機自主飛行的精度比較，並利用收集的影像進行三維重建，對洋香瓜植物進行分析。收集的影像經過處理後，用於提取關鍵的植物表型數據，本研究主要分析了植株的高度和展幅，透過高度和展幅進一步對植物的生長進行監測，通過擬合生長曲線，並將其與實際生長數據進行比較。在無人機的飛行精度方面，平行飛行路徑的誤差範圍為 7 至 12 公分，閉環飛行路徑為 5 至 9 公分，而多高度路徑的誤差範圍為 4 至 11 公分，顯示出穩定的飛行精度。接著，我們進一步比較了在相同覆蓋面積下，多機系統與單機系統的效能。結果顯示，多機系統能夠將任務時間縮短 73%，並將電池消耗降低 5%。在作物三維重建方面，我們比較了三種軌跡所收集的三種不同的重建方法，分別是單面、合併以及三個高度。根據評估重建結果的指標，使用 PSNR、SSIM 和 LPIPS 三個指標進行比較，結果顯示三個高度方法在重建質量上表現最佳，PSNR 為 0.37，SSIM 為 9.48，LPIPS 為 0.65。在植物高度的測量上，合併方法達到了最低的 MAE 誤差為 6.6 公分，而在展幅測量方面，單面方法則達到了最低的 MAE 誤差為 5.8 公分。本研究展示了多機無人機系統在溫室作物監測中的應用潛力，還證明了不同重建方法和測量策略在提高農業監測精度和效率方面的有效性。

關鍵詞: 多機無人機系統、Visual SLAM、三維重建、高斯潑濺、表型分析、生長監測

doi:10.6342/NTU202503732

ABSTRACT

This study developed an autonomous multi-drone system for muskmelon crop growth monitoring in a greenhouse. Through collaborative multi-drone operations, multi-angle images were captured and three flight paths were designed. The UWB positioning system was used to compare the accuracy of the UAV autonomous flight, and the collected images were used for 3D reconstruction to analyze muskmelon plants. The images were processed to extract key phenotypic data, focusing on plant height and canopy span, and growth monitoring was performed by fitting growth curves and comparing them with actual growth data. In terms of flight accuracy, the parallel flight path had an error range of 7 to 12 cm, the closed-loop path had an error range of 5 to 9 cm, and the multi-altitude path had an error range of 4 to 11 cm, demonstrating stable flight precision. We also compared the performance of multi-drone and single-drone systems over the same coverage area. The multi-drone system reduced mission time by 73% and battery consumption by 5%. For 3D reconstruction, we compared three methods collected along three different paths: Single-side, Merged, and Three-height. Evaluation metrics showed that the Three-height method provided the best reconstruction quality with PSNR of 0.37, SSIM of 9.48, and LPIPS of 0.65. For height measurement, the Merged method achieved the lowest MAE of 6.6 cm, and for canopy span measurement, the Single-side method achieved the lowest MAE of 5.8 cm. This study demonstrates the potential of multi-UAV systems in greenhouse crop monitoring and proves the effectiveness of different reconstruction methods and measurement strategies in improving monitoring accuracy and efficiency.

Keywords: Multi-UAV system, Visual SLAM, 3D reconstruction, Gaussian Splatting, Phenotyping analysis, Growth monitoring

doi:10.6342/NTU202503732

Table of Contents

doi:10.6342/NTU202503732

doi:10.6342/NTU202503732

doi:10.6342/NTU202503732

List of Figures

doi:10.6342/NTU202503732

doi:10.6342/NTU202503732

doi:10.6342/NTU202503732

doi:10.6342/NTU202503732

List of Tables

doi:10.6342/NTU202503732

CHAPTER 1
Introduction

1.1 General Background Information

Agriculture has long been fundamental to human civilization, yet it now faces unprecedented challenges due to climate change, resource scarcity, and labor shortages (Singh & Singh, 2017). In response, the industry has entered the era of Agriculture 4.0, which integrates advanced technologies such as the Internet of Things (IoT) and artificial intelligence to increase productivity, improve resource allocation, and enable real-time decision-making (Zhai et al., 2020).

As climate variability intensifies, many production systems have shifted from open fields to controlled environments like greenhouses. These environments offer the advantage of precise control over temperature, humidity, and lighting conditions, critical for ensuring crop quality and stability. With the integration of sensors and automation technologies, greenhouses have evolved into smart systems capable of real-time monitoring and adaptive control (Huynh et al., 2023). These advancements are particularly beneficial for cultivating high-value horticultural crops such as tomatoes, strawberries, and muskmelon (*Cucumis melo* L.), which require delicate environmental regulation. Previous studies have shown that greenhouse cultivation can significantly improve muskmelon yield and quality (Pardossi et al., 2000). However, the management of such high-value crops still depends heavily on accurate, timely, and non-invasive monitoring, which presents new technical challenges even in smart greenhouses.

doi:10.6342/NTU202503732

Even within smart greenhouses, effectively monitoring crop health and growth remains a significant challenge. For crops such as muskmelon that require frequent phenotypic assessment, traditional manual inspection is both time-consuming and labor-intensive, and often fails to provide the continuity and resolution necessary for precise crop management. To address this issue, unmanned aerial vehicles (UAVs) have emerged as a promising solution. UAV platforms support remote sensing, high-frequency inspections, and automated data collection, offering new possibilities for large-scale phenotyping and crop health monitoring. Aslan et al. (2022) noted that while UAV applications are well-established in open-field agriculture, their deployment in greenhouse environments remains limited due to constraints such as weak GPS signals and restricted space. Nevertheless, the study highlighted the strong potential of UAV-based monitoring in greenhouses and emphasized the need for solutions such as simultaneous localization and mapping (SLAM) technologies to overcome the challenges of indoor navigation.

While single-UAV systems offer promising solutions for crop monitoring in smart greenhouses, they are inherently constrained by limited flight time, restricted spatial coverage, and vulnerability to single-point failures. These limitations can hinder the ability to perform consistent and large-scale monitoring tasks in complex greenhouse environments. To address these challenges, recent research has increasingly turned to multi-UAV systems, which enable cooperative operations among multiple drones to improve system scalability and resilience. For example, Kim et al. (2019) demonstrated that deploying multiple UAVs in a coordinated manner significantly improves spatial coverage and reduces mission time during agricultural inspections. In a separate study, Ju and Son (2018) proposed a distributed control strategy for cooperative UAVs and showed that such systems enhance energy efficiency and robustness in field operations.

doi:10.6342/NTU202503732

These multi-UAV configurations are particularly well-suited for smart greenhouse applications, where high-frequency, non-invasive monitoring is essential for capturing detailed and dynamic crop traits across large areas.

Building upon the capabilities of cooperative multi-UAV systems in smart greenhouse environments, recent advancements in computer vision and 3D reconstruction have enabled more detailed and non-destructive approaches to plant phenotyping. By capturing images from multiple and complementary viewpoints, UAV-based systems can reconstruct accurate 3D models of plants, allowing for precise estimation of morphological traits such as plant height, canopy width, and leaf area.

Li et al. (2022) demonstrated that multi-view image sequences can effectively reconstruct 3D models of maize seedlings in field conditions, enabling high-resolution measurement of key phenotypic parameters. Expanding on this approach, James et al. (2025) developed a scalable UAV-based pipeline that integrates deep learning for efficient 3D phenotyping of sorghum panicle in greenhouse environments. Their work highlights the feasibility and robustness of applying UAV-based 3D reconstruction technologies in controlled agricultural settings.

Despite these advancements, the application of such phenotyping pipelines in smart greenhouses remains technically challenging due to factors like occlusion from dense foliage, limited viewpoints caused by structural constraints, and the need for seamless coordination among multiple UAVs. In this context, multi-UAV systems offer significant advantages by enabling flexible and comprehensive data acquisition from multiple perspectives, which can substantially improve the completeness and reliability of 3D reconstructions.

doi:10.6342/NTU202503732

1.2 Research Objectives

This study aims to develop an autonomous multi-UAV system for in-greenhouse crop monitoring, leveraging simultaneous localization and mapping (SLAM) to construct maps and estimate UAV poses in GPS-denied environments. The proposed system enables distributed autonomous navigation among multiple UAVs, thereby improving navigation efficiency and scalability within the confined and structurally complex greenhouse environment. By coordinating the UAV fleet for image acquisition from multiple perspectives, the system facilitates 3D reconstruction of plants, which in turn supports accurate phenotypic analysis and continuous growth monitoring. The specific objectives of this research are as follows:

1. To develop an autonomous multi-UAV navigation system adapted for operation in greenhouse environments using SLAM.
2. To reconstruct detailed 3D plant models from multi-view images captured by UAVs inside the greenhouse.
3. To perform plant phenotypic analysis and monitor growth dynamics based on the reconstructed 3D models.

doi:10.6342/NTU202503732

CHAPTER 2
Literature Review

The shift from traditional open-field agriculture to controlled environments such as greenhouses has introduced new challenges for crop monitoring. Among high-value crops cultivated in greenhouses, tomatoes, strawberries, and muskmelons require frequent, accurate, and non-invasive monitoring to ensure optimal growth. While smart greenhouses have adopted fixed sensor networks and automation systems, these solutions often lack the spatial flexibility and adaptability needed for high-resolution, plant-level monitoring across dynamic environments. To address these challenges, unmanned aerial vehicles (UAVs) offer a promising alternative due to their mobility, scalability, and ability to perform high-frequency data collection. Their deployment in greenhouses, however, is not without challenges, particularly in navigation and coordination within GPS-denied, space-constrained environments. The increasing complexity of crop monitoring demands integrated solutions that combine UAV mobility with advanced navigation, mapping, and data analysis capabilities.

2.1 UAV Applications in Greenhouse Environments

The advancement of unmanned aerial vehicle (UAV) technology has become an increasingly important tool in precision agriculture, particularly demonstrating strong potential in greenhouse applications. As global food demand continues to rise, UAVs contribute significantly to improving agricultural productivity, especially in environmental monitoring and crop management. With high versatility and a wide range

doi:10.6342/NTU202503732

of sensing capabilities, UAVs can effectively replace traditional sensors. Their practical applicability and cost-efficiency have gradually established them as essential tools for greenhouse management (Aslan et al., 2022).

2.1.1 Development and Applications

In agriculture, UAVs equipped with various sensors, such as multispectral cameras and thermal imagers, can efficiently collect imagery data and monitor the agricultural environment. Studies have shown that these sensors not only capture high-resolution crop images but also assess plant moisture and growth status, thereby enabling farmers to more precisely monitor crop development, manage irrigation systems, and apply fertilizers (Rejeb et al., 2022). In outdoor agricultural settings, UAVs have been extensively employed to oversee large-scale crops, perform spraying, detect pests and diseases, and estimate yields. As UAV technology has matured, its applications have gradually extended into greenhouse environments. Within greenhouses, farmers can control environmental conditions, and UAVs, through automated systems, can carry out three-dimensional environmental monitoring and data collection, thus improving production efficiency and reducing resource waste (Roldán et al., 2015).

The application of unmanned aerial vehicles in greenhouses is extensive, including monitoring plant growth, detecting diseases, and managing temperature and humidity. For example, UAV systems equipped with multispectral cameras and LiDAR (Light Detection And Ranging) technology can accurately monitor plant health, enabling farmers to detect disease or nutrient deficiencies promptly and apply pesticides or fertilizers in a timely manner. In addition, by combining LiDAR technology with infrared

doi:10.6342/NTU202503732

imaging, it is possible to provide precise temperature distribution data under low-light or night-time conditions, ensuring optimal growth conditions for plants in varying environments (Gu et al., 2020). These technologies assist farmers in rapidly identifying problem areas and making adjustments immediately, greatly reducing the risk of pests and diseases and effectively minimizing both soil moisture evaporation and plant transpiration stress. By integrating these approaches, greenhouse management becomes more precise and efficient.

As technology continues to advance, the capabilities of drones have gradually improved, and depending on the scale, greenhouse operations may require a large amount of labor. Research indicates that through simultaneous localization and mapping (SLAM), drones can navigate autonomously in greenhouses where GPS signals are limited, avoid obstacles and complete tasks effectively. Krul et al. (2021) demonstrated that a small drone with a monocular camera and ORB-SLAM can function reliably in greenhouse corridors despite plant occlusion and low-texture floors, producing sparse point-cloud maps that support real-time navigation and improve inspection efficiency.

Currently, UAV technology has emerged as a critical tool for greenhouse management. Through the integration of multiple sensors and advanced navigation technologies, UAVs can operate effectively under complex greenhouse conditions. Nevertheless, despite ongoing technological advancements, numerous challenges remain in the deployment of UAVs within greenhouses.

doi:10.6342/NTU202503732

2.1.2 Case Studies and Challenges

Utilizing UAVs within greenhouse environments presents numerous challenges, one of which is the stability of communication systems. Pantos et al. (2023) reported that the enclosed environment of greenhouses adversely affects UAV signal propagation, noting that measurements taken near the roof and floor are especially prone to signal attenuation. Structural components within greenhouses induce signal attenuation, compromising reliable data transmission and real time video streaming, both of which are essential for tasks requiring immediate feedback such as environmental monitoring and crop health evaluation. These challenges highlight the need for continued technological innovation in order to achieve optimal UAV performance in greenhouse operations.

In addition, when UAVs are used in greenhouses with low light or rapidly changing lighting conditions, the sensitivity and accuracy of sensors can be adversely affected, resulting in reduced data reliability. For example, infrared sensors or multispectral cameras may not provide clear image data under insufficient lighting, which can compromise the accuracy of plant health assessments or microclimate monitoring (Bagagiolo et al., 2022). These challenges indicate that UAV deployment still requires further improvement of sensing technologies to adapt to the variable conditions found in greenhouses. The case studies demonstrate that although UAV applications in greenhouses can enhance automation efficiency and reduce reliance on human labor, technical limitations remain. For instance, UAV autonomy and sensor accuracy face technical bottlenecks that must be addressed through technology development. To meet these challenges, future research should focus on further improving UAV navigation precision in order to enable effective operation in more complex and dynamic greenhouse environments.

doi:10.6342/NTU202503732

2.2 Single-UAV and Multi-UAV Navigation

UAV technology plays a critical role in modern agriculture, particularly in greenhouse environments. Through autonomous navigation, UAVs can enhance farm management efficiency and reduce labor costs. The precise monitoring requirements in greenhouses, such as temperature, humidity, and crop health, have driven continuous advances in UAV technology. In this setting, UAV deployment not only reduces labor demands but also improves operational accuracy and timeliness (Park et al., 2025). Both single-UAV and multi-UAV systems demonstrate significant advantages in agricultural automation. Single-UAV platforms are generally simple to operate and cost-effective, making them well suited to tasks such as indoor greenhouse inspection or point-specific monitoring (Khosiawan & Nielsen, 2016). In contrast, multi-UAV configurations exhibit superior performance in collaborative missions by leveraging distributed control frameworks that allow simultaneous multi-zone monitoring and data collection, thereby offering greater efficiency and flexibility when managing large greenhouses or complex tasks (Chung et al., 2018).

2.2.1 Single-UAV Navigation Applications and Challenges

Single-UAV technology has been widely applied in agricultural monitoring, logistics delivery, and search-and-rescue operations. In terms of autonomous navigation, UAV navigation can be understood as the planning process that allows the aircraft to safely and quickly reach a target location based on its current environmental and positional information. To successfully complete a mission, a UAV must be aware of its

doi:10.6342/NTU202503732

own state, including its current position, flight speed, heading direction, as well as its starting point and target location (Lu et al., 2018).

In greenhouse environments, a single UAV equipped with visible-light, multispectral, and thermal-infrared sensors can independently perform tasks such as crop monitoring and microclimate detection. According to Hunt Jr and and Daughtry (2018) shown that by flying at different altitudes, a UAV can gather temperature, humidity, and light intensity data and combine this information with multispectral imagery to assess plant growth status. This approach helps analyze how environmental changes inside the greenhouse affect crop development.

However, greenhouses often present severe obstructions and limited GPS signals. Chang et al. (2023) review autonomous UAV navigation in GPS-denied environments and note that these methods fall into map-based and mapless strategies. By using techniques such as vision–inertial fusion and SLAM, UAVs can significantly improve their localization and stability, ensuring they can navigate narrow, highly obstructed passages or dense vegetation areas (Chang et al., 2023). In small-scale greenhouses, single-UAV applications focus on regular patrols, such as inspecting plant health or collecting light-intensity and humidity data. Because of their ease of operation and flexibility, single UAVs offer an ideal low-cost solution for autonomous patrol missions. Nevertheless, their main limitation is flight endurance, which makes it difficult to cover larger areas.

2.2.2 Multi-UAV Navigation Applications and Challenges

Multi-UAV systems enable multiple aircraft to cooperate on tasks that are either impractical or inefficient for a single vehicle, such as large-area monitoring, synchronized

doi:10.6342/NTU202503732

data collection, or persistent coverage in dynamic environments. In greenhouse settings, where rapid and accurate assessment of microclimate conditions and crop health over a wide area is essential, coordinated UAV fleets can divide the workload to cover multiple zones simultaneously.Vetrella et al. (2016) describe a navigation framework in which a UAV formation acts like a virtual sensor array. Each UAV exchanges differential GPS data with its peers and combines this information with vision tracking to reduce magnetic and inertial interference. As a result each UAV can use relative position information from others, which improves accuracy and robustness when individual GPS signals are weak. By using cooperative DGPS together with visual features matching the fleet maintains its formation and keeps reliable state estimates for all UAVs, reducing accumulated positioning errors.

Building on the concept of cooperative sensing, Trujillo et al. (2018) develop a vision-based SLAM method for multi-UAV formations that follow a moving lead agent. When the lead agent, whether it is another UAV, a ground vehicle, or a person, moves unpredictably, follower UAVs must estimate both their own positions and the lead's state. The authors examine three configurations using nonlinear observability analysis, each with different sensor setups. In the first configuration, one UAV (Quad 1) keeps the lead in view with a monocular camera and combines camera-based landmark angles with onboard inertial data. In the second configuration, if the lead is outside the camera's field of view, Quad 1 adds range measurements and altimeter readings to maintain its estimate. In the third configuration, multiple chasing UAVs share visual and inertial data among themselves. In all cases, a high-level control scheme adjusts the formation in real time based on those estimates. Simulations show that this cooperative SLAM approach produces stable position estimates for both the UAV swarm and the lead agent even when GPS is unreliable, enabling reliable formation control and collision avoidance.

doi:10.6342/NTU202503732

Deploying multi-UAV systems in greenhouses still faces major challenges. Greenhouse structures and dense foliage often block the view, making vision-based measurements unreliable, and DGPS alone cannot prevent inertial drift. Maintaining reliable communication between UAVs is difficult because metal frames and high humidity interfere with signals. Unpredictable wind and thermal updrafts inside greenhouses also demand control algorithms that can adjust flight paths on the fly. Finally, small UAVs have limited battery life, so they must plan routes that balance area coverage and energy use. Addressing these challenges requires adding lightweight LiDAR or ultra-wideband radios for backup ranging, designing formations that can tolerate brief communication losses, and creating mission plans that optimize both coverage and power.

In summary, integrating differential GPS with vision based tracking and advanced visual SLAM architectures can greatly improve multi-UAV navigation in complex settings. Yet in greenhouse operations, further work is needed on sensing that resists occlusion, more robust communication strategies, and energy aware coordination algorithms to achieve fully autonomous, large scale aerial monitoring. Among various sensing technologies, vision-based navigation, with its passive, lightweight, and low-power characteristics, serves as a crucial complement for UAV localization and guidance in greenhouse environments.

2.2.3 Vision-based Navigation

Vision-based navigation provides UAVs with the ability to interpret and respond to their surroundings using only camera input. By extracting distinctive visual features and constructing or updating maps in real time, this approach enables both single and multiple

doi:10.6342/NTU202503732

UAVs to maintain accurate positioning and trajectory planning without relying on external signals.

Early SLAM techniques were primarily based on geometric methods, using traditional tools from geometry and probability theory for localization and map construction. These approaches relied on a robot's motion model and observation model to estimate both the robot's position in space and the locations of environmental feature points, thereby building a 3D map. For example, EKF-SLAM (Extended Kalman Filter–SLAM) is a typical geometry-based SLAM method that uses an extended Kalman filter to track and update the robot's pose and the positions of feature points (Bailey et al., 2006). However, as the number of observed points increases, its computational complexity grows quickly, creating a performance bottleneck (Li et al., 2018).

Subsequently, SLAM methods began using image-based approaches. MonoSLAM was the first to use a single camera for both motion estimation and map building (Davison et al., 2007). This image-based SLAM relies on a continuous sequence of camera images and estimates the camera's pose by extracting and tracking feature points. ORB-SLAM relies on ORB features (Oriented FAST and Rotated BRIEF) for localization and mapping, making it fast and efficient for real-time use (Mur-Artal et al., 2015). In contrast, LSD-SLAM (Large-Scale Direct Monocular SLAM) skips discrete feature extraction and instead optimizes pixel intensities directly, producing a semi-dense depth map (Engel et al., 2014). This allows LSD-SLAM to keep tracking in areas with few clear corners or in low-texture scenes, but it requires more computation and can be slower in very large environments. Compared to ORB-SLAM's sparse maps and lower processing demands, LSD-SLAM gives richer detail at the cost of speed. Overall, ORB-SLAM's lower

doi:10.6342/NTU202503732

computational requirements and consistently reliable tracking make it a more practical choice than LSD-SLAM for most real-time applications.

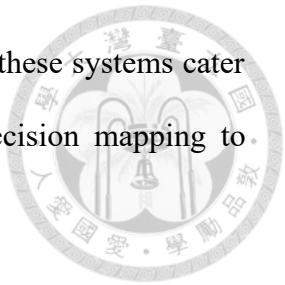
Building on these developments in visual SLAM, subsequent frameworks expanded the capabilities of camera-based localization and mapping by introducing new features and supporting a wider range of sensors and platforms. ORB-SLAM2 is a feature-based SLAM system for monocular, stereo, and RGB-D cameras. It offers reliable loop closure, map reuse, and relocalization, delivering high accuracy in various environments (Mur-Artal & Tardós, 2017). ORB-SLAM3 builds on this by adding visual-inertial fusion and multi-map support. Its joint optimization of camera and IMU data improves tracking during fast motion or brief visual loss, and its Atlas module manages multiple submaps, making it more robust for large-scale mapping (Campos et al., 2021).

While ORB-SLAM2 and ORB-SLAM3 emphasize highly optimized feature extraction and sensor fusion for accuracy, other frameworks emerged with a focus on flexibility and modularity. OpenVSLAM, for example, prioritizes support for diverse camera models and ease of integration into various development workflows. It uses ORB features but supports many camera types, such as fisheye cameras, omnidirectional cameras, and stereo setups (Sumikura et al., 2019). A lightweight map-fusion plugin allows easy switching between sparse, semi-dense or custom map formats without rebuilding the system. Because it does not rely on ROS, it can run more easily on embedded or resource-limited platforms.

In summary, ORB-SLAM2 is praised for its reliable loop closure and relocalization capabilities, and ORB-SLAM3 builds on this foundation by incorporating visual-inertial fusion and multi-map management to enhance motion handling and enable large-scale revisits. Meanwhile, OpenVSLAM stands out for its support of diverse camera types,

doi:10.6342/NTU202503732

modular map formats, and rapid development workflows. Together, these systems cater to a wide range of SLAM requirements, from achieving high-precision mapping to offering flexible deployment with unconventional camera setups.



2.3 Multi-UAV Control and Communication

Multi UAV control and communication architectures can be broadly categorized into centralized, distributed, and hybrid designs. Centralized architectures rely on a primary controller that issues commands and gathers data from each vehicle. This approach simplifies coordination but can create a single point of failure and may struggle with scalability. Distributed architectures share decision making among vehicles so that each unmanned aerial vehicle collaborates based on local information and peer communication. Such systems are more resilient and flexible but require robust algorithms to reach consensus and handle dynamic changes. Hybrid architectures combine elements of centralized and distributed designs by maintaining a global mission plan while allowing individual vehicles to operate autonomously within local constraints. Across all three strategies, effective communication and collaboration mechanisms are essential to exchange state information, coordinate tasks, and ensure overall system performance.

2.3.1 Centralized Architecture

In multi-UAV collaborative control systems, a centralized control architecture is generally considered to provide higher task execution efficiency and decision-making quality. A centralized control system manages the status and task allocation of the entire

doi:10.6342/NTU202503732

UAV swarm through a cloud server or ground control station. Research by Loayza et al. (2017) points out that when centralized control is applied to collaborative UAV missions such as virtual seeding in precision agriculture, it can effectively govern each UAV's movement and prevent collisions, thereby enhancing task stability and effectiveness.

According to Jamshidpey et al. (2024), centralized control demonstrates better speed and efficiency in coverage tasks in multi UAV systems, especially when precise control and efficient task completion are required. However, a drawback of centralized control is its high dependence on a single control point, which leads to poor scalability and makes the system prone to collapse if the control point fails. In simulated tests this centralized control method showed significant advantages in task completion speed and coverage uniformity, but when applied on a large scale it is easily limited by communication bottlenecks and single point failures.

Furthermore, Hu et al. (2018) compared centralized and distributed control and pointed out that although centralized control usually outperforms distributed control in decision quality, it faces a significant increase in delays as the swarm size grows. In centralized control all decisions and computations are concentrated on a cloud server. This enables the system to make decisions based on global data but also makes the system vulnerable to performance degradation due to communication delays or resource constraints.

2.3.2 Distributed Architecture

Compared to centralized control, distributed control architectures place greater emphasis on collaboration among individual UAVs and enable them to make decisions without reliance on a central control unit. According to Asaamoning et al. (2021)

doi:10.6342/NTU202503732

distributed control can effectively address the challenges that UAV swarms face when performing autonomous missions in dynamic and uncertain environments. These control systems have self-organizing capabilities and do not depend on a single control point, thereby improving system scalability and fault tolerance.

To further leverage the benefits of distributed architecture, Zhang et al. (2023) proposed a distributed cooperative search method called DCS-UC that incorporates ant colony optimization to handle unstable communications among UAVs. Their approach enables each UAV to adjust its search path based on local pheromone information and neighbor broadcasts, resulting in improved coverage efficiency and robustness when communication links are intermittent. Simulation results demonstrate that DCS-UC outperforms traditional approaches by achieving higher search success rates and reducing mission completion time under network instability.

However, the disadvantage of distributed control is that each UAV can only make decisions based on its own information and cannot know the positions or battery levels of other UAVs, which may result in lower quality task allocation compared to centralized control. Although distributed control offers a significant advantage in decision latency, when complex missions are involved centralized control still provides superior decision making.

2.3.3 Communication and Collaboration Mechanisms

In cooperative control of UAV swarms, communication and collaboration constitute the core mechanisms for achieving collective objectives. UAV swarms exchange information through diverse communication architectures to ensure smooth execution of missions and coordinated operation among all vehicles.

doi:10.6342/NTU202503732

Azzam et al. (2023) propose a multi-agent reinforcement learning framework based on centralized training and decentralized execution (CTDE). In this framework, all UAV behaviors are managed centrally during the training phase to ensure they acquire cooperative navigation capabilities. During the execution phase, decentralized control is employed so that each UAV can make decisions based on local sensing and peer-to-peer messages, thereby improving system flexibility and scalability. Communication stability is critical during cooperative flight. Zhang et al. (2023) address the problem of unstable communications under radio frequency interference by designing a dual-threshold detection algorithm. This algorithm monitors the collaborative process, prevents collisions, and maintains connectivity, thus enhancing the reliability of the UAV swarm in dynamic environments. To reduce communication latency while maintaining efficient collaboration, Chen et al. (2024) develop a rapid cooperation method for large-scale UAV swarms. Their method relies on local interactions and predictive mechanisms, allowing the swarm to reach consensus quickly even under constrained channel resources. This approach preserves collaboration efficiency when multiple UAV groups merge, enhancing the swarm's adaptability in dynamic conditions.

These research findings demonstrate that, in large-scale applications, careful design and selection of communication mechanisms enable UAV swarms to maintain efficient collaboration and successfully complete complex tasks in dynamic environments.

2.4 Plant Phenotyping

Plant phenotyping quantifies traits such as leaf morphology, leaf count, canopy spread, height, biomass and color to reveal genetic and environmental interactions. By utilizing noninvasive imaging, spectral measurements and other sensor-based techniques,

doi:10.6342/NTU202503732

researchers can capture repeated measurements throughout development to identify trends and detect stress or disease at early stages. These approaches inform precision agriculture practices, such as irrigation, fertilization and pest management, and support breeding programs by pinpointing superior genotypes. Ultimately, plant phenotyping enables high-throughput, non-destructive evaluation of performance across diverse environments.

2.4.1 Plant Phenotyping Techniques

Plant phenotyping aims to quantify plant morphological, physiological and biochemical traits that develop because of interactions between genotype and environment. Methods can be divided into invasive approaches that require destructive sampling of tissues to measure traits such as biomass or root architecture and non-invasive approaches that employ imaging, spectral or other sensor technologies to collect continuous data without disturbing plant growth. While invasive methods deliver precise measurements, they do not allow longitudinal monitoring of the same individual. In contrast, non-invasive techniques combine RGB and multispectral imaging, LiDAR point clouds, thermal imaging and other sensors to achieve high throughput repeated measurements and capture growth and physiological dynamics over time and space (Pieruschka & Schurr, 2019).

Non-invasive phenotyping uses a variety of imaging and sensor technologies to monitor plants without causing damage. Visible-light cameras capture RGB images to measure leaf area and canopy cover, and multispectral and hyperspectral sensors record reflectance at different wavelengths to assess pigment composition, nutrient status and

doi:10.6342/NTU202503732

stress indicators. Thermal infrared cameras map canopy temperature to infer transpiration rates and water stress, and chlorophyll fluorescence imaging evaluates photosynthetic efficiency while also revealing early signs of abiotic stress. Three-dimensional measurements rely on stereo vision and LiDAR scanners, which generate point clouds to reconstruct plant structure and calculate height, volume and branching architecture. Furthermore, tomographic methods such as magnetic resonance imaging, computed tomography and positron emission tomography reveal internal features such as root networks and vascular tissues at high resolution without excavation (Fiorani & Schurr, 2013). In greenhouse environments these sensors underpin systems like PlantEye and FieldScanalyzer. PlantEye combines high-resolution three-dimensional scanning with multispectral imaging to capture detailed morphological and physiological data and FieldScanalyzer enables rapid multispectral imaging of large plant populations (Gao et al., 2024). These non-invasive methods provide a comprehensive toolkit for high-throughput trait measurement.

2.4.2 Two-Dimensional Image-Based Approaches

As high throughput imaging technologies become more widespread, methods for plant phenotyping using two dimensional images have evolved from early traditional image processing toward machine learning and deep learning techniques. According to Chandra et al. (2020), the number of publications employing deep learning in the plant phenotyping field has increased rapidly in recent years. Architectures based on convolutional neural networks are especially common and have been successfully applied to tasks such as leaf counting, species classification, and disease detection. This trend

doi:10.6342/NTU202503732

shows that deep learning methods provide higher accuracy and greater automation compared to traditional machine learning approaches.

Early two-dimensional image-based phenotyping often relied on manually engineered features such as color histograms, texture parameters, and shape descriptors. For example, Zhao et al. (2016) extracted 36 root morphological traits in pea and used a random forest and support vector machine approach to distinguish among 16 European cultivars. Using only the five most informative traits yielded up to 86% accuracy and outperformed models built on more extensive or random feature sets. By contrast, Teshome et al. (2023) employed UAV-derived crop surface models (UAVH) from multispectral imagery collected one to two weeks before harvest to predict sweet corn height, biomass, and yield. A simple linear regression on UAVH achieved a concordance index of 0.99, an R^2 of 0.99, and a mean absolute error of five centimeters for height estimation. Nonlinear models such as random forest, support vector machine and k-nearest neighbors predicted biomass with concordance values between 0.88 and 0.99, demonstrating that machine learning applied to two-dimensional UAV imagery can reach near-field measurement accuracy just weeks before harvest. These studies confirm the feasibility and precision of combining two-dimensional imagery with machine learning for plant phenotyping. However, machine learning methods depend heavily on manually designed features, which become a performance bottleneck when faced with high-dimensional, nonlinear images containing complex backgrounds. In contrast, deep learning can automatically learn hierarchical representations from raw images, reducing reliance on manual feature engineering and achieving superior predictive accuracy and generalization in complex phenotyping scenarios.

doi:10.6342/NTU202503732

Deep learning's automatic feature learning further enhances accuracy and automation in phenotyping. Ampatzidis and Partel (2019) used a UAV equipped with a multispectral camera to capture citrus orchard images. They applied the YOLOv3 model to achieve plant detection accuracy of 99.8 percent with precision of 99.9% and recall of 99.7%. They also estimated canopy area with accuracy of 85.5% and identified tree gaps with accuracy of 94.2 %. Williams et al. (2024) proposed a zero-shot leaf segmentation method based on the Segment Anything Model (SAM) that operates without annotated data. In potato leaf segmentation their method achieved average precision of 60.3% and average recall of 63.2%, while a fine-tuned Mask R-CNN model achieved precision of 74.7% and recall of 78.7%. This demonstrates that deep learning methods are feasible when data are limited. These studies show that deep learning automatically extracts multilevel features from two-dimensional images and providing higher generalization in complex scenarios.

2.4.3 Three-Dimensional Point Cloud-Based Approaches

Two-dimensional image-based plant phenotyping techniques can extract information about organs such as leaves and stems from single or multiple 2D images, but they remain heavily affected by occlusions, lighting conditions, and limited viewpoints, often failing to fully reconstruct the plant's three-dimensional structure. To overcome these limitations, researchers have gradually shifted their attention to three-dimensional point cloud data, aiming to use depth information to rebuild a more realistic representation of plant morphology and thereby improve the accuracy and robustness of segmentation and phenotypic feature extraction.

doi:10.6342/NTU202503732

Common methods for obtaining three-dimensional point clouds include laser scanning, photogrammetry, and depth cameras. Laser scanning can directly produce high-precision, high-density point clouds, but the equipment is costly and operation is complex. In contrast, photogrammetry uses Structure-from-Motion to estimate camera parameters and then applies Multi-View Stereo to generate dense point clouds. By capturing multiple overlapping images with a standard camera or drone, one can automatically reconstruct a three-dimensional model. Tools such as COLMAP perform camera calibration, sparse reconstruction, and dense reconstruction (Schönberger & Frahm, 2016), while GLOMAP further optimizes distributed computing and deep feature matching efficiency (Pan et al., 2024). In addition, RGB-D cameras and stereo cameras can capture depth information in real time. Although their range and accuracy in low-texture regions are limited, they offer a distinct advantage in capturing depth for close-range or dynamic scenes (Wang et al., 2020).

Utilizing these point cloud acquisition methods, subsequent algorithms are able to perform precise segmentation and reconstruction of plant structures within three-dimensional space. Shi et al. (2019) first captured plant images simultaneously from multiple cameras at different angles and then applied a convolutional neural network to each two-dimensional image for pixel-wise semantic and instance segmentation to identify parts such as stems and leaves. They subsequently used Structure-from-Motion (SfM) and Multi-View Stereo (MVS) techniques to reconstruct a dense point cloud, projecting the 2D segmentation results back into three-dimensional space to complete the final 3D point cloud segmentation. In their results, the precision for pixel-wise stem segmentation using only 2D images was approximately 0.77, whereas it increased to around 0.97 when analyzed with 3D point clouds. For leaf segmentation, the 2D method achieved a precision of roughly 0.93 to 0.95, while the 3D approach reached 1.00,

doi:10.6342/NTU202503732

demonstrating that 3D point clouds significantly improve segmentation accuracy. Building on Shi et al.'s demonstration that 3D point clouds outperform 2D-only methods under controlled conditions, Li et al. (2022) proposed a stem-leaf point cloud dataset and applied a 3D Edge-Preserving Sampling (3DEPS) strategy to preprocess the input point clouds. They then designed PlantNet, a dual-function network that performs both stem-leaf semantic segmentation and individual leaf instance segmentation by using a Local Feature Extraction Operation (LFEO) module based on dynamic graph convolutions and a Semantic-Instance Feature Fusion Module (FFM). When tested on tobacco, tomato, and sorghum point clouds, PlantNet achieved an average precision of 92.49% and an intersection-over-union (IoU) of 85.86% in semantic segmentation tasks. In instance segmentation, PlantNet achieved a mean precision (mPrec) of 83.30% and a mean coverage (mCov) of 78.62%. These results highlight PlantNet's effectiveness in single-plant or controlled-environment point clouds, where manually annotated samples ensure high-fidelity training data.

In contrast, Zarei et al. (2024) addressed the challenge of large-scale, highly overlapping outdoor field conditions by introducing a digital-twin approach to generate a synthetic sorghum field point cloud dataset. They trained PlantSegNet, a graph convolution network designed for instance segmentation in complex field scenarios. On real sorghum point cloud data, PlantSegNet achieved an mCov of 0.53 and an AP of 0.69, outperforming TreePartNet, which was trained on the same synthetic and real data. While PlantNet attains higher precision and IoU under controlled conditions, PlantSegNet's use of large-scale synthetic training makes it more robust to field-scale occlusions and leaf overlaps, demonstrating superior accuracy when segmenting geometrically similar plant organs in outdoor environments.

doi:10.6342/NTU202503732

CHAPTER 3

Materials and Methods

3.1 Architecture of Autonomous Multi-UAV System

3.1.1 UAV and Edge Computing Hardware Architecture

The hardware architecture consists of a centralized ground control station and multiple distributed edge nodes, each composed of a computing unit, a wireless communication module, and a drone platform, as illustrated in Fig. 3-1.

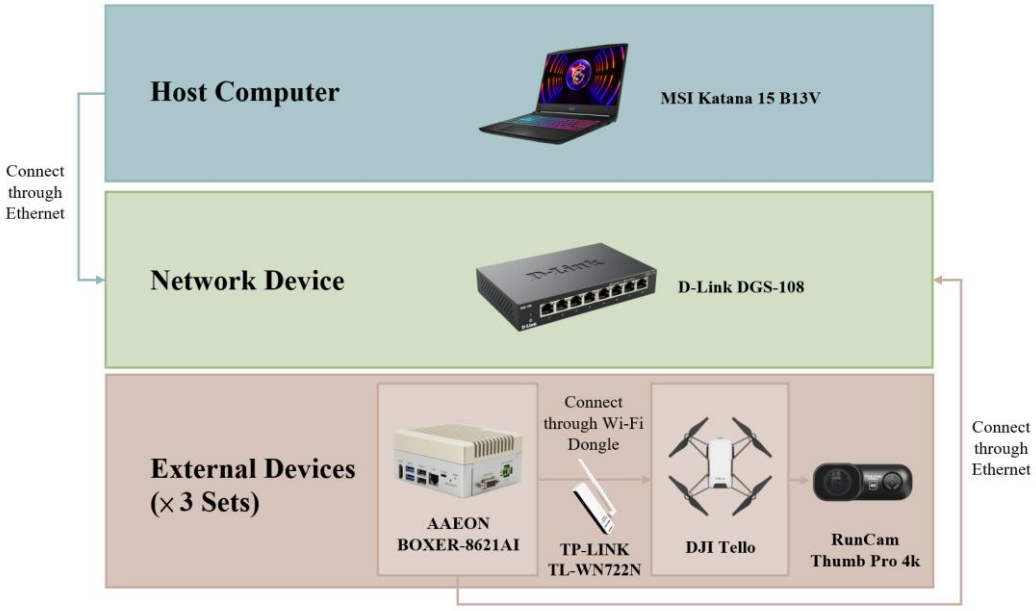


Fig. 3-1. Hardware architecture of the multi-UAV system

The ground control station is an MSI Katana 15 B13V laptop, featuring a 13th-generation Intel Core i7 processor, an NVIDIA GeForce RTX 4060 Laptop GPU, 16 GB of DDR5 RAM, and a 1 TB PCIe SSD. It runs Ubuntu 22.04 with ROS 2 Humble, and

doi:10.6342/NTU202503732

handles central mission planning, data management, and orchestration of all ROS 2 nodes. For connectivity with the edge nodes, the station is linked via Gigabit Ethernet to a D-Link DGS-108 8-port unmanaged switch, providing a 1 Gbps-per-port LAN backbone with high stability and low latency.

We employ the AAEON BOXER-8621AI embedded computer as our edge node platform. The specifications are shown in Table 3-1 NVIDIA Jetson Orin Nano Super Mode provides the computation performance needed for UAV cruising. The six-core Arm Cortex-A78AE CPU together with 4 GB of LPDDR5 memory supports image transmission and autonomous navigation algorithms. A 128 GB M.2 2242 SSD ensures that flight data can be recorded quickly and reliably. With dimensions of just 10.5 cm × 9 cm × 5.2 cm, the device is compact while retaining powerful performance. It is therefore ideal as an edge node for multi-UAV operations. Its extended operating temperature range guarantees stable performance even in the high temperatures of a greenhouse environment.

Table 3-1. AAEON BOXER-8621AI Edge Node Specifications

Category	Specification
AI Accelerator	NVIDIA® Jetson Orin Nano™ with Super Mode
CPU	6-core Arm® Cortex®-A78AE ARMv8.2 64-bit CPU
System Memory	4GB LPDDR5
Storage Device	128GB M.2 2242 B+M Key SSD
Dimension	10.5 cm × 9 cm × 5.2 cm
Gross Weight	1.1kg
Operating Temperature	-15˚C ~ 60˚C

doi:10.6342/NTU202503732

To connect the embedded computer to our UAV, we use the TP-Link TL-WN722N adapter, which supports 802.11n wireless communication at speeds up to 150 Mbps and is equipped with a detachable 4 dBi antenna to maintain a dedicated 2.4 GHz point-to-point link with the UAV.

In this study, we selected the DJI Tello as our UAV platform. The detailed specifications are shown in Table 3-2. The Tello is an affordable commercial drone, retailing on DJI's official website at USD $99 including one battery. It is equipped with an HD 720p 30 fps RGB camera. The UAV weighs only 80g and its compact dimensions make it well suited for flight in narrow greenhouse aisles.

Table 3-2. DJI Tello Specifications

Category	Specification
Weight	80g
Dimensions	9.8 cm × 9.25 cm × 4.1 cm
Video Quality	HD 720 p @ 30 fps
Max Flight Distance	100 m
Max Speed	8 m/s
Max Flight Time	13 min
Max Flight Height	30 m
Field of View	82.6°

To obtain higher-quality video and images, each DJI Tello is equipped with a RunCam Thumb Pro W camera, as illustrated in Fig. 3-2. According to the official DJI documentation, the maximum payload capacity of the Tello is 20 g. To meet this weight limitation, the outer casing of the RunCam Thumb Pro W camera was removed, and a

doi:10.6342/NTU202503732

lightweight lithium battery along with a voltage boost module was integrated. This configuration allows stable power supply to the camera while keeping the total added weight within the allowable limit. The camera captures 4K RGB footage at 30 fps with a 155° field of view and records video in synchronization with the UAV's flight operations.

Fig. 3-2. DJI Tello Equipped with a RunCam Thumb Pro W Camera for Enhanced Imaging

Table 3-3. RunCam Thumb Pro W Specifications

Category	Specification
Weight	16g
Dimensions	5.4 cm × 2.55 cm × 2.1 cm
Resolution	4K@30fps\2.7K@60fps\1440P@60fps\1080P@120fps\1080P@60fps
Field of View	155°

doi:10.6342/NTU202503732

3.1.2 Software Architecture

The UAV ground station runs Ubuntu 22.04 LTS with the corresponding Robot Operating System 2 (ROS 2) Humble release. ROS 2 is the next-generation open-source robotics framework built on the Data Distribution Service (DDS) middleware, providing a native distributed architecture and quality-of-service controls that allow modules to exchange data reliably, flexibly and in real time. Compared with ROS 1, it offers significant enhancements in security and real-time performance and supports a wide range of operating systems and embedded platforms while retaining familiar packages for perception, motion control, navigation and task planning. Its modular design and cross-language support for C++ and Python simplify system integration and debugging and have established ROS 2 as a leading choice for both robotics research and industrial applications.

Fig. 3-3 illustrates the overall software architecture for multi-UAV operations. The Tello Driver can use the Tello SDK to send commands for takeoff, landing and manual flight operations and can also publish topics to collect the UAV's RGB video, odometry and IMU data. During a navigation mission each UAV first transmits real-time H.264 compressed video to the ground station over Wi-Fi. Within the ROS 2 framework the Tello Driver decompresses each frame and forwards it to Stella VSLAM to obtain localization data. To enable autonomous flights a map must be created first. Once the map is generated it is transferred via Secure Copy Protocol (SCP) to each embedded computer. Using the received map each UAV defines its waypoint positions in Stella VSLAM and establishes its waypoint mission. Finally the multi-UAV autonomous flight is executed. Each UAV computes the error between its current position and the target waypoint based on the Stella VSLAM output. A PID controller then calculates the

doi:10.6342/NTU202503732

velocities required to reach each waypoint. Those velocity commands are sent back to the Tello Driver to control the UAV's motion. Through this process the multi-UAV autonomous navigation is completed.

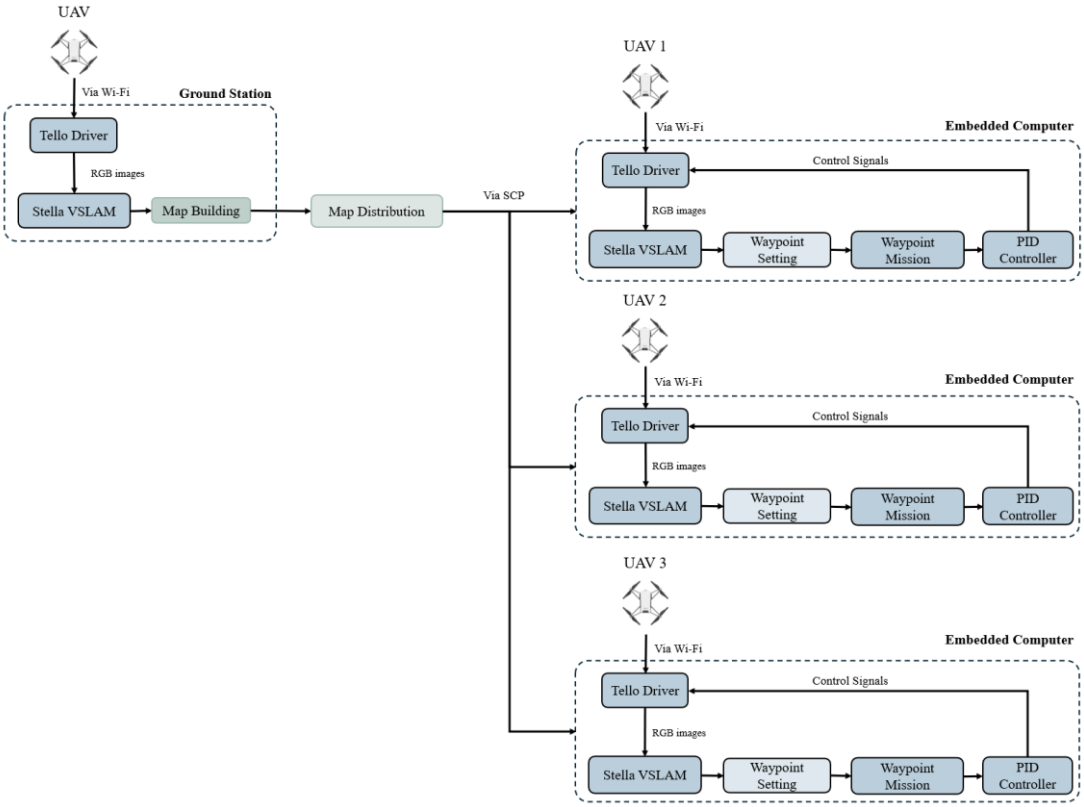


Fig. 3-3. Overview of the software architecture of multi-UAV system

3.1.3 Multi-UAV Communication Architecture

The multi-UAV communication architecture consists of a host computer, computing nodes and UAV units as shown in Fig. 3-4. The ground control station acts as the central command node. It orchestrates mission coordination, aggregates data from distributed edge computing units and makes high-level decisions. All devices in the system communicate over a dedicated local area network using a Gigabit Ethernet switch and each node has a static IP address to ensure consistent and deterministic network

doi:10.6342/NTU202503732

addressing. Each edge computing node has its own unique static IP address. These nodes perform intensive computational tasks in a Docker container, such as real-time visual SLAM processing, PID based control and waypoint mission management. This containerization simplifies deployment, ensures consistency across nodes and streamlines updates and maintenance. In operation, each embedded computer communicates wirelessly with its assigned DJI Tello UAV using a dedicated point-to-point 2.4 GHz channel. This setup delivers low latency and high reliability for continuous video streaming and real-time flight command feedback. The overall architecture leverages ROS 2's DDS based middleware. We configured namespaces for each UAV avoid conflicts and quality of service policies ensure robust message exchange between distributed nodes and the host system.

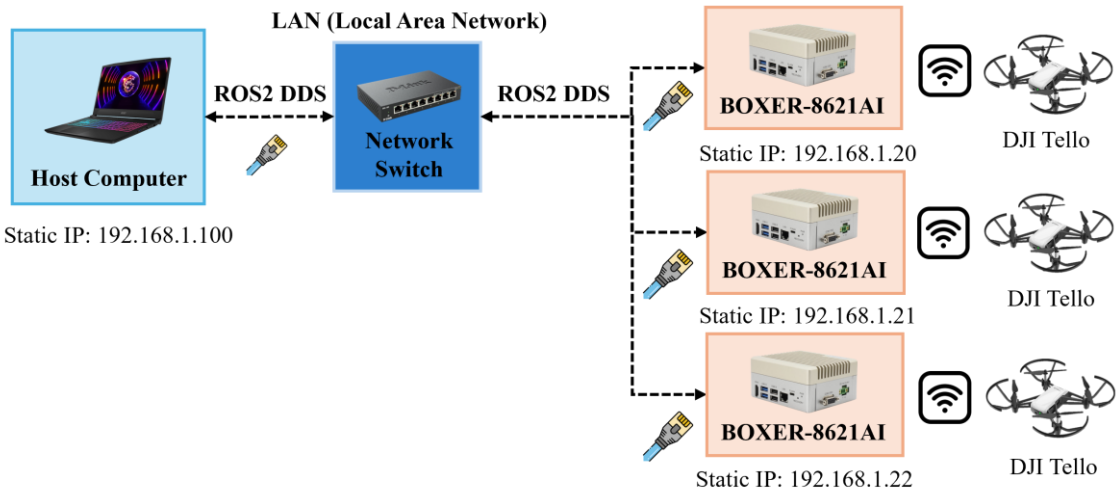


Fig. 3-4. Multi-UAV communication Architecture

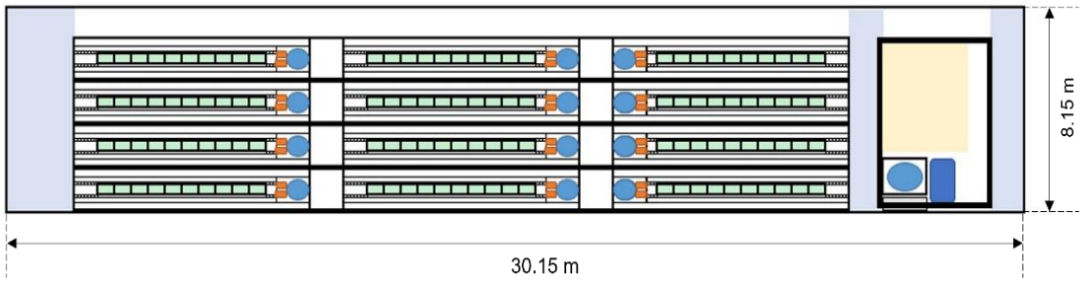
3.2 Autonomous Multi-UAV Flight in Greenhouse

3.2.1 Experimental Setup in Greenhouse

Experiments for this study were conducted in the Intelligent Energy-saving Greenhouse at the National Taiwan University Agricultural Experiment Station. The

doi:10.6342/NTU202503732

greenhouse measures 30.15 m in length and 8.15 m in width and contains twelve sets of angle steel frames. Each frame measures 5.4 m in length, 0.75 m in width and 2.1 m in height. Our experiments took place in the rear section of the greenhouse, which is outlined by the red box in Fig. 3-5(a). The greenhouse is planted with muskmelon of the Summer No. 2 Japanese Arus variety, as shown in Fig. 3-5(b). Each steel frame supports five plants and yields up to sixty fruits per quarter. Fig. 3-5(c) presents a photograph of the site, where the aisle width is 1.16 m, providing a suitable environment for UAV flight experiments.



(a)

(b)

(c)

Fig. 3-5. Experimental area in the experimental greenhouse: (a) schematic layout highlighting the rear test section; (b) Summer No. 2 Japanese Arus muskmelon;(c) on-site view of the 1.16 m-wide aisle used for UAV flights.

doi:10.6342/NTU202503732

3.2.2 Visual SLAM for Localization and Mapping

In this study we use Stella VSLAM, which is a branch created from OpenVSLAM (Sumikura et al., 2019) to continue its development. This VSLAM framework supports monocular, stereo and RGBD camera types and, like ORB-SLAM (Mur-Artal et al., 2015) and ProSLAM (Schlegel et al., 2018), employs an indirect SLAM algorithm with sparse features. Its core architecture builds upon ORB-SLAM and ORB-SLAM2 and the entire codebase has been redesigned to improve scalability, readability and performance. Enhancements include map storage and loading capabilities and a modular system design that encapsulates several functions into independent components.

Each UAV performs intrinsic and distortion calibration of its monocular camera using the ROS 2 camera_calibration package's cameracalibrator tool before using Stella VSLAM, and we then add the resulting parameters to Stella VSLAM's configuration file. Each UAV streams its calibrated monocular RGB video to its onboard SLAM module, which performs visual odometry and map generation in real time. Stella VSLAM automatically selects keyframes, triangulates sparse landmarks and carries out local bundle adjustment to refine both pose and structure estimates. After loop closures are detected using a bag-of-words model, the system performs global pose graph optimization to ensure mapping consistency over time. The resulting keyframe map is lightweight, reusable and well suited for multi-UAV localization. Once the map is generated, the ground station distributes the map files to each UAV via the Secure Copy Protocol (SCP). During autonomous flight, the onboard Stella VSLAM module loads the prebuilt map and matches live RGB descriptors against the stored landmarks to achieve real-time pose estimation in all six degrees of freedom, including translation along the x, y and z axes and rotation in roll, pitch and yaw.

doi:10.6342/NTU202503732

3.2.3 PID Controller for Waypoint Navigation

To execute precise flight paths, the system leverages UAV position and yaw error data derived from Stella VSLAM as inputs to a cascade PID controller. This controller operates separate proportional-integral-derivative (PID) loops for translation and rotation, generating the velocity and yaw-rate commands transmitted to the flight controller. This control architecture was pioneered by Bouabdallah et al. (2004) in their influential work comparing PID and LQ control techniques on an indoor micro-quadrotor. Their research demonstrated that despite the theoretical advantages of optimal control methods, a well-tuned PID controller could achieve centimeter-level waypoint tracking precision.

In our implementation, continuous closed-loop feedback is maintained between the UAV's pose, which is estimated through Stella VSLAM, and the current waypoint, effectively minimizing position and heading errors. Like Bouabdallah's approach, we adapt their fundamental control structure to support our autonomous navigation. Control commands are unified into geometry_msgs/Twist messages before transmission to the UAV's flight controller. A key advantage of this control strategy lies in the real-time tunability of PID gains and constraints, enabling adaptation to dynamic environmental conditions and external disturbances. Once each target position is reached and its prescribed dwell time completed, the mission autonomously proceeds through the sequence of waypoints.

3.2.4 UWB-based Ground Truth and SLAM Error Evaluation

In order to evaluate UAV localization accuracy via Stella VSLAM, we used ultra-wideband (UWB) technology for ground truth measurements. UWB devices transmit information using extremely short pulse sequences that last less than one nanosecond and

doi:10.6342/NTU202503732

create a signal bandwidth exceeding 500 MHz. This wide bandwidth gives UWB signals high temporal resolution and greatly reduces the effects of non-line-of-sight conditions on measurements (Zafari et al., 2019). As a result, UWB systems can accurately measure Time of Flight (ToF) and enable precise navigation and positioning based on time-of-flight ranging (Zhang et al., 2025).

In this study we used the Nooploop LinkTrack UWB ranging and positioning system as our experimental ground truth. Furthermore, according to the official Nooploop LinkTrack documentation, the system achieves typical one- and two-dimensional positioning accuracy of ±10 cm and three-dimensional accuracy of ±30 cm, supports update rates up to 200 Hz with end-to-end latency as low as 5 ms, accommodates up to 120 anchors and 200 tags, and provides a data transmission bandwidth of up to 3 Mbps. The combination of high accuracy and high update rate makes it well suited for UAV flight ground truth. In this study we used five LinkTrack P-A as shown in Fig. 3-6(a). Four of these units served as UWB anchor nodes and one functioned as the UWB ground station. Each UAV was equipped with an LTP-AS2 as its UWB tag, as illustrated in Fig. 3-6(b). The actual installation of the UWB tag on the DJI Tello is shown in Fig. 3-6(c).

(a) (b) (c)

Fig. 3-6. Nooploop LinkTrack UWB ranging and positioning system: (a) LinkTrack P-A; (b) LTP-AS2; (c) DJI Tello UAV equipped with a UWB system tag.

doi:10.6342/NTU202503732

Prior to multilateration, we removed statistical outliers using an interquartile-range filter with a multiplier k tuned for each tag and then applied a zero-phase Gaussian smoothing filter to suppress residual high-frequency noise and multipath artifacts. Filtered ranges were converted to Cartesian positions via a least-squares multilateration algorithm, producing a time-series of UWB-derived poses. Both the UWB and SLAM trajectories were then aligned by estimating a single rigid body transform from the take-off segment to remove constant offset and synchronized at common timestamps. Let $p_{SLAM,i}$ be the UAV position estimated by Stella VSLAM at time i and $p_{UWB,i}$ the ground-truth position from UWB at the same instant. At each matched time point t_i , the instantaneous localization error e_i is computed using Equation 3-1:

$$e_i = \| p_{SLAM,i} - p_{UWB,i} \| \tag{3-1}$$

We aggregated these values into root-mean-square error (RMSE), mean absolute error (MAE), standard deviation (Std Dev), maximum and minimum error, and median error. By also plotting the error time series and its empirical cumulative distribution, we captured temporal drift, transient spikes during aggressive maneuvers, and the proportion of samples within specified accuracy bounds, which provides a comprehensive evaluation of Visual SLAM based flight control under dynamic conditions.

3.2.5 Visual SLAM Map Optimization

This experiment investigates how mapping the same closed-loop trajectory with a handheld UAV using different numbers of passes affects map quality. Only the number of loops performed during the handheld scan varies between trials.

doi:10.6342/NTU202503732

We first acquired imagery of the environment with a handheld UAV carrying a monocular camera, following the rectangular circuit shown in Fig. 3-7. We complete one loop, two loops, and three loops at roughly the same speed and altitude. Stella VSLAM then processes these image sequences using local bundle adjustment to generate an initial sparse point-cloud map for each loop count.

After the map is built, the UAV performs a single loop flight, matching its live monocular images to the preconstructed map to estimate pose and refine it through local bundle adjustment. Meanwhile, an ultra-wideband (UWB) localization system records the UAV's true trajectory with centimeter-level accuracy to serve as ground truth for error analysis.

Finally, the maps created with one, two, and three passes are compared in terms of feature count, keyframe count, and pose error between the VSLAM estimate and the UWB reference trajectory. This comparison reveals the effect of the number of handheld loops on Visual SLAM map optimization.

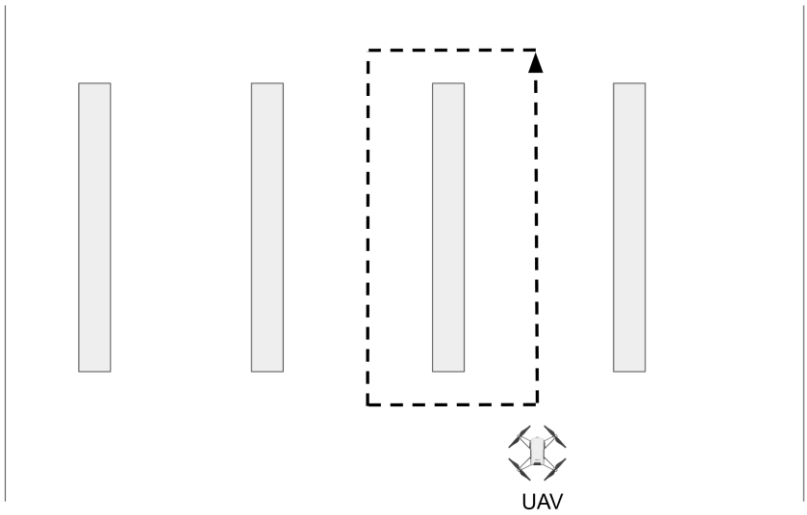


Fig. 3-7. Mapping trajectories used for Visual SLAM map optimization in a greenhouse environment

doi:10.6342/NTU202503732

3.3 Multi-UAV Cooperative Path Planning and Analysis

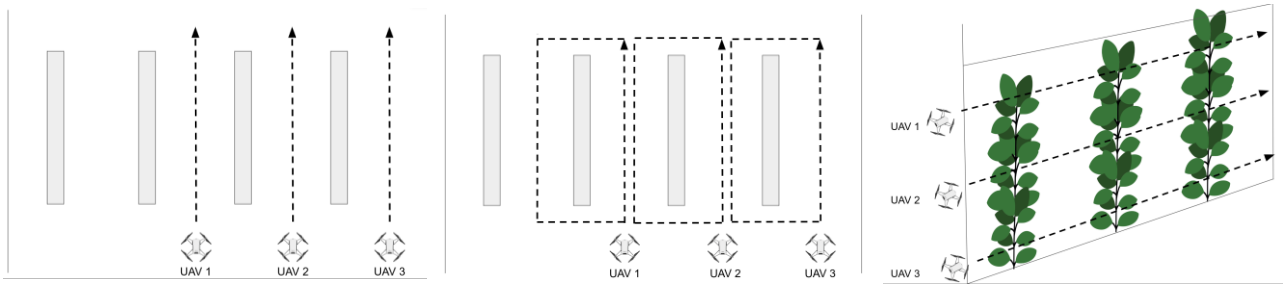
3.3.1 Cooperative Path Planning Design

In this study, we designed three distinct cooperative multi-UAV flight trajectories to capture high-resolution, multi-perspective imagery of muskmelon plants within a greenhouse environment. Our primary objective was to leverage multiple UAVs simultaneously to capture comprehensive visual data for Structure from Motion (SfM) reconstruction, as diverse viewing angles significantly enhance feature correspondence and depth estimation in SfM algorithms. These cooperative approaches explore how different flight patterns affect 3D model quality and completeness.

Fig. 3-8 illustrates three different cooperative multi-UAV flight trajectories. Fig. 3-8(a) shows the parallel-aisle flight pattern, where three UAVs navigate simultaneously along adjacent crop aisles, each equipped with a RunCam Thumb Pro W camera oriented laterally to record the side surfaces of the plants. This configuration maximizes greenhouse coverage efficiency while maintaining consistent imaging distances. Fig. 3-8(b) shows the closed-loop flight trajectory, in which UAVs circle around each planting row to capture both front and rear views of the plants. Capturing imagery from both sides significantly improves surface completeness, reduces occlusions, and enhances feature correspondence, ultimately leading to more accurate and robust 3D reconstruction results. By providing comprehensive dual-side information for each plant, we retain only the images containing both front and back views of the plants and discard the remaining images that do not capture the target plants. Fig. 3-8(c) presents the multi-altitude flight, which employs a height-staggered arrangement where UAVs maintain distinct altitude

doi:10.6342/NTU202503732

levels. This vertical distribution strategy effectively overcomes the blind spots inherent in single-plane imaging and ensures comprehensive coverage of taller specimens from multiple vertical perspectives. The integration of horizontal and vertical diversity in camera positions provides more robust geometric information for the SfM algorithm, resulting in more detailed plant models with improved accuracy in vertical structures.



(a) (b) (c)

Fig. 3-8. Three cooperative multi-UAV flight trajectories used for greenhouse imaging: (a) parallel-aisle flight, (b) closed-loop flight, and (c) multi-altitude flight.

The combination of these cooperative UAV configurations demonstrates how strategic deployment of multi-UAV collaborative approaches can enhance the quality and completeness of plant phenotyping data beyond what could be achieved with single-UAV methods.

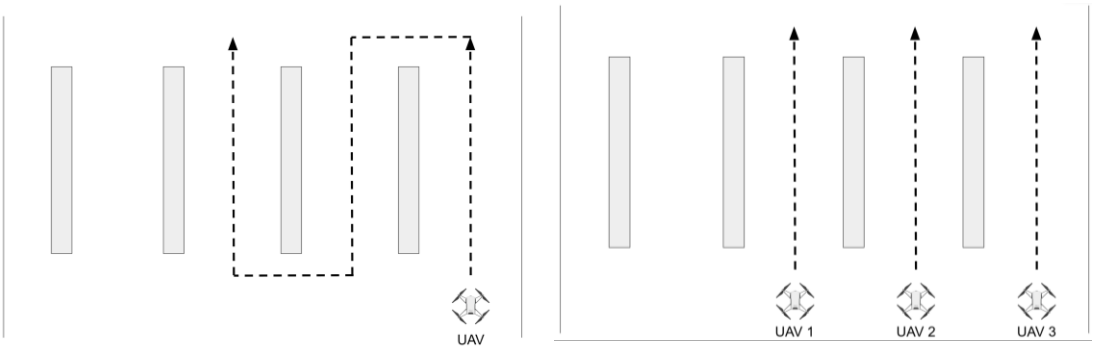
3.3.2 Comparison Between Single and Multi-UAV Flights

To evaluate the advantages of collaborative UAV systems in greenhouse phenotyping, it is essential to compare their performance against that of traditional single-

doi:10.6342/NTU202503732

UAV operations. Such a comparison helps reveal differences in coverage efficiency, flight time, and other performance evaluation metrics under identical task requirements.

As shown in Fig. 3-9(a), the single-UAV configuration involves one UAV following an S-shaped path to sequentially traverse the entire greenhouse area, covering all three crop rows and completing the mission independently. This approach reflects the conventional method commonly used in agricultural remote sensing, where a single UAV is responsible for surveying the entire field. Although this results in longer flight durations, the mission planning process is relatively straightforward. In contrast, as shown in Fig. 3-9(b), the three crop rows are divided into separate paths assigned to individual UAVs. This collaborative strategy significantly reduces the total mission time and enhances overall operational efficiency, demonstrating the advantages of multi-UAV deployment.



(a) (b)

Fig. 3-9. Comparison of flight path designs: (a) Single-UAV approach with sequential field coverage and (b) Multi-UAV distributed approach with parallel field coverage

To quantitatively compare the performance of single- and multi-UAV configurations, we defined three key evaluation metrics. Flight Time Efficiency (FTE) is defined as the total duration from takeoff to landing for a single mission; for the multi-UAV system, it

doi:10.6342/NTU202503732

is defined as the maximum individual flight time among the three UAVs. Area Coverage Rate (ACR) quantifies the field area surveyed per unit time, expressed in m^2/min. The total surveyed area is estimated by assuming each image captures a fixed 8.5 m × 4.1 m ground footprint and multiplying this area by the number of valid frames. Battery Consumption is evaluated using the total battery percentage consumed across all UAVs. In the multi-UAV system, this value is computed as the sum of battery usage from each drone. The derived energy efficiency is defined as the surveyed area per total battery percentage used ($m^2 \cdot \%^{-1}$). These metrics provide a consistent and comparable basis for evaluating operational performance across different deployment scenarios.

3.4 3D Reconstruction Methods of Plants

3.4.1 Structure from Motion

For 3D reconstruction of muskmelon plants, we employed Structure-from-Motion (SfM) techniques to process the multi-perspective imagery collected by our cooperative UAV system. Structure from Motion (SfM) is a vision-based 3D reconstruction technique that recovers both the relative camera poses and the three-dimensional structure of a scene from a collection of two-dimensional images captured from different viewpoints. Rather than relying on external positioning systems, SfM infers spatial information purely from image correspondences and geometric constraints, which makes it especially suitable for environments such as greenhouses where GPS signals are unavailable. The typical SfM pipeline consists of several key stages. It begins with feature detection and descriptor computation using methods such as SIFT or ORB. Corresponding features are then matched across image pairs to estimate the relative camera poses. Triangulation is used

doi:10.6342/NTU202503732

to convert these matched points into 3D coordinates, resulting in a sparse point cloud. The reconstruction is further refined through bundle adjustment, which jointly optimizes camera parameters and 3D point positions. In the final stage, a dense reconstruction process is applied to generate a more complete and detailed 3D model of the scene. SfM is well suited for plant phenotyping and morphological modeling due to its ability to automatically integrate large volumes of multi-angle imagery without requiring specialized hardware. In this study, we adopted SfM as the core method for 3D reconstruction using imagery captured by our cooperative UAV system.

COLMAP is a widely used Structure-from-Motion and Multi-View Stereo (MVS) framework that follows an incremental reconstruction strategy. It begins by selecting an initial image pair and gradually incorporates additional images into the model, repeatedly performing feature matching, triangulation, and bundle adjustment throughout the process (Schönberger & Frahm, 2016). This approach is effective for small to medium-sized image sets with sufficient overlap and sequential capture order, and has been extensively applied in aerial and terrestrial 3D reconstruction tasks.

While COLMAP represents a widely adopted SfM implementation, we adopted GLOMAP (Pan et al., 2024) for our reconstruction pipeline because it offers higher global consistency and better robustness under multi-UAV scenarios with large inter-viewpoint baselines and near-parallel trajectories, which are common in greenhouse environments. GLOMAP's one-step global optimization framework avoids the drift and model fragmentation issues typically observed in incremental pipelines and is particularly effective when combining image sets captured simultaneously from multiple viewpoints.

As illustrated in Fig. 3-10 the GLOMAP pipeline consists of three main stages: correspondence search, global estimation, and output reconstruction. Unlike traditional

doi:10.6342/NTU202503732

global SfM methods that separate translation averaging and triangulation, GLOMAP performs a single global positioning step that jointly optimizes both camera poses and 3D structure, thereby improving robustness and convergence.

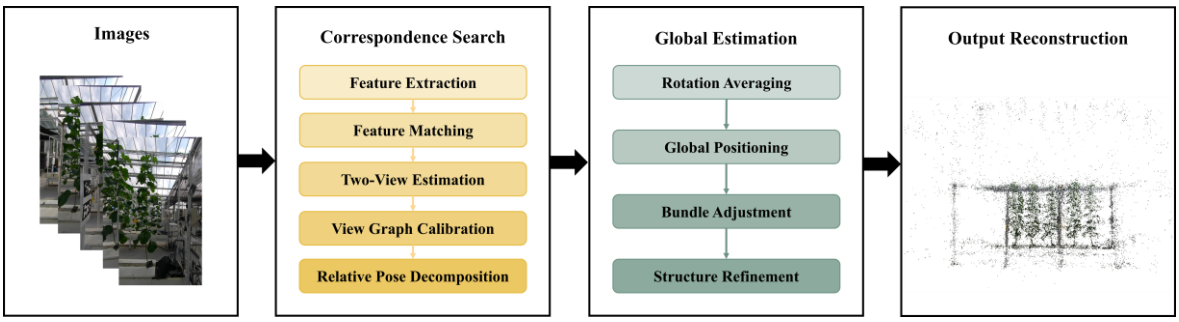


Fig. 3-10. Pipeline of the GLOMAP system. (Pan et al., 2024)

To execute the GLOMAP pipeline, we utilized a containerized workflow based on the jinwj1996/glomap Docker image (https://hub.docker.com/r/jinwj1996/glomap). The reconstruction process included feature extraction, image matching, global mapping, and model conversion stages. Feature extraction was performed using SIFT with GPU acceleration, and the camera model was set to SIMPLE_PINHOLE. Matching was conducted using COLMAP's sequential matcher with an overlap of 4 on single-sided image sets, whereas for merged and three-height image acquisitions we applied the exhaustive matcher. Mapping was executed through the glomap mapper command to generate the sparse reconstruction. The final model was converted into both TXT and PLY formats using the COLMAP model converter module.

3.4.2 Gaussian Splatting

In order to obtain a dense, photo-consistent volumetric model from the sparse point cloud and camera poses produced by our GLOMAP pipeline, we turn to 3D Gaussian

doi:10.6342/NTU202503732

Splatting. Unlike mesh based multi view stereo, Gaussian Splatting represents the scene as millions of tiny anisotropic Gaussians whose means, covariances, colors and opacities are jointly optimized so that rasterized splats match the input images. This approach yields extremely fast novel-view rendering while naturally capturing fine geometry such as thin leaves and stems. The unstructured cloud of Gaussians that results from vanilla optimization does not immediately yield an editable surface mesh because the Gaussians tend to be arbitrarily positioned and overlapping.

To bridge this gap, we adopt the Gaussian Splatting method SuGaR, proposed by Guédon and Lepetit (2023). SuGaR introduces a regularization term during optimization that enforces each Gaussian to lie nearly tangent to the true surface and to adopt a flat shape aligned with local normals. Concretely, SuGaR derives a signed distance function from the Gaussian-induced density and compares it to an ideal signed distance function in which level sets correspond exactly to the underlying surface. By minimizing their discrepancy, SuGaR encourages Gaussians to align and distribute evenly across the scene surface. Once this alignment is achieved, SuGaR extracts the mesh via Poisson reconstruction on a chosen density level set, producing a high-quality watertight triangle mesh within minutes. Optionally, SuGaR performs a brief joint refinement that binds new Gaussians to the mesh triangles, further enhancing photometric fidelity and enabling standard mesh editing and animation workflows. In our implementation, we initialize the Gaussians from the GLOMAP sparse points and camera parameters and then run SuGaR's surface alignment and mesh extraction stages to produce a dense, editable model ideally suited for downstream phenotyping and morphological analysis.

We selected density-normalization-consistency as our regularization method because preliminary experiments showed it best preserved fine surface detail. We set the number of refinement iterations to 15 000 to ensure convergence and enabled the high-

doi:10.6342/NTU202503732

poly preset to capture thin leaves and stems. All other parameters were left at their defaults since they already produced visually faithful results. In our study, all optimization iterations are executed on an NVIDIA RTX A6000 GPU with 48 GB of VRAM, ensuring stable convergence. The method efficiently learns an explicit 3D representation that allows real-time rendering of novel views. The SuGaR optimization settings used in our experiments are summarized in Table 3-4.

Table 3-4. SuGaR Optimization Training Parameters

Parameter	Value
Regularization method	density-normalization consistency
Refinement iterations	15000
Gaussians per triangle	1
Surface iso-level	0.3
Output mesh vertices	1000000
High-poly preset	enabled
Export OBJ	enabled
Export PLY	enabled
Optimizer	Adam
Learning rate	1×10^{-3}

doi:10.6342/NTU202503732

3.4.3 Evaluation Metrics

To quantitatively assess the fidelity of our 3D reconstruction and novel-view rendering, we employ three complementary image-based metrics: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and learned perceptual image patch similarity (LPIPS). The PSNR between a rendered image I_r and its ground-truth counterpart I_g, where L is the maximum possible pixel value and $\mathrm{MSE}(I_r, I_g)$ is computed as the average squared difference between corresponding pixels in the rendered and ground-truth images, is defined as Equation 3-2:

$$\mathrm{PSNR}(I_r, I_g) = 10 \log_{10}\left(\frac{L^2}{\mathrm{MSE}(I_r, I_g)}\right) \tag{3-2}$$

The structural similarity index between a rendered image I_r and its ground-truth counterpart I_g, where μ_r and μ_g are local means, σ_r^2 and σ_g^2 are local variances, σ_{rg} is the local covariance, and C_1, C_2 are stabilizing constants, is defined as Equation 3-3:

$$\mathrm{SSIM}(I_r, I_g) = \frac{(2\mu_r\mu_g + C_1)(2\sigma_{rg} + C_2)}{(\mu_r^2 + \mu_g^2 + C_1)(\sigma_r^2 + \sigma_g^2 + C_2)} \tag{3-3}$$

doi:10.6342/NTU202503732

We adopt Learned Perceptual Image Patch Similarity (LPIPS) as our third evaluation metric to capture perceptual differences that go beyond pixel-wise error. LPIPS operates by passing both the rendered image I_r and the ground-truth image I_g through a pretrained convolutional network and extracting intermediate feature maps at several layers. At each selected layer l, the feature maps $\phi_l(I_r)$ and $\phi_l(I_g)$ are first spatially normalized and then compared via an l_2 distance computed channel-wise. These distances are weighted by learned per-channel scaling factors w_l, and the final LPIPS score is obtained by averaging the weighted distances across all spatial locations and summing over layers. Zhang et al. (2018) demonstrated that this learned, deep-feature based measure aligns closely with human judgments of image similarity, making LPIPS a powerful complement to PSNR and SSIM for assessing novel-view rendering quality.

In our study, we evaluate the reconstruction results obtained from three different flight trajectories using the PSNR, SSIM, and LPIPS metrics.

3.5 Phenotyping for Muskmelon Plant

3.5.1 Extraction of the Single Plant

To measure each plant's height and canopy spread more precisely, we need to segment the Gaussian splatting reconstruction, which represents an entire row of plants, into individual plant point clouds for subsequent phenotypic analysis. In Fig. 3-11 shows the whole workflow of extraction of single plant.

In this study, we first leverage the Segment Anything Model 2 (SAM 2) with the base model checkpoint to generate 2D masks of the target plant in each RGB image. By

doi:10.6342/NTU202503732

specifying a rectangular bounding box around the plant region, SAM 2 produces one binary mask per view, which serve as soft annotations indicating where the plant appears in the scene. These masks encode, in image space, the rough silhouette of the individual plant and greatly reduce reliance on manual point-cloud cleaning. Next we map each two dimensional mask into the three dimensional scene using the camera intrinsics and extrinsics obtained from COLMAP. For each mask pixel we trace a ray from the camera center and define a pyramidal viewing volume that extends through the reconstruction. By gathering all such volumes we obtain a set of frusta that cover every masked region. This frustum-based masking step then retains only those points that lie inside at least one frustum and removes the rest of the point cloud, effectively carving out the plant of interest and discarding points belonging to neighboring vegetation or background structures (Qi et al., 2017).

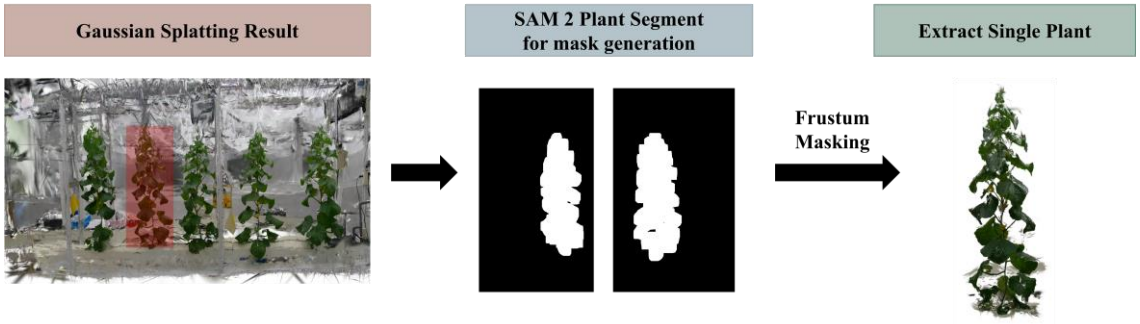


Fig. 3-11. Extraction of the Single Plant workflow

3.5.2 Plant Height and Canopy Span Measurement

In this section we present the detailed steps of our algorithm for extracting plant height and canopy width from a reconstructed three-dimensional point cloud. Fig. 3-12 illustrates the complete processing pipeline.

doi:10.6342/NTU202503732

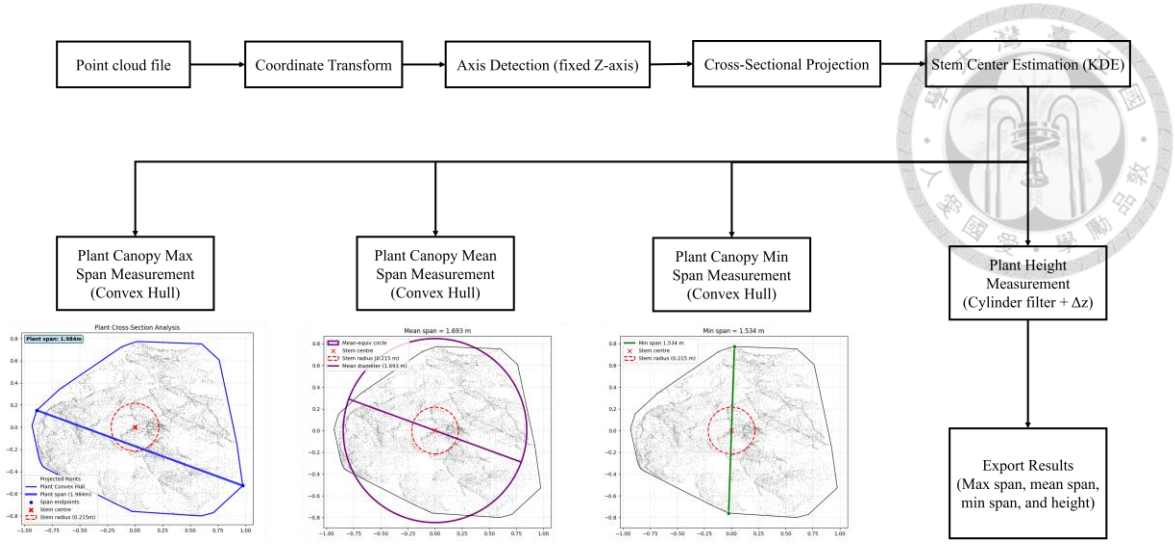


Fig. 3-12. Plant Height and Canopy Span Measurement Pipeline

The single-plant point cloud is first loaded into the Open3D framework, where the original COLMAP Y and Z axes are swapped and the depth axis is inverted to align the global Z direction with the plant's vertical orientation. The cloud is then translated so that its centroid coincides with the origin. To generate cross sections, two orthonormal basis vectors are defined on the plane perpendicular to the Z axis, and each 3D point is projected onto this plane while preserving its Z coordinate, resulting in two-dimensional coordinates for each elevation level. A two-dimensional kernel density estimator with a bandwidth of 0.15 is applied on a 200 × 200 grid in the projected plane. The highest density peak nearest the origin is found and used to establish the stem center.

For canopy span measurement, the convex hull of the projected cross section is first computed. The maximum span is obtained as the largest distance between any two hull vertices. A virtual line is then rotated through the hull while anchored at the stem center to identify the shortest chord passing through the center, representing the minimum central span. The area enclosed by the convex hull is subsequently calculated, and the diameter of an equivalent circle is derived from this area as the average span, providing a global width estimate that balances local indentations and protrusions.

doi:10.6342/NTU202503732

To measure height, we compute radial distances from each projected point to the stem center and define the stem radius as the 20th percentile of these distances. We then filter points within this radius, determine the minimum and maximum Z values, and calculate their difference as the plant height. Finally, all span and height values are scaled by 1000 to convert meters to millimeters for phenotypic analysis.

Using the above method, estimated plant height and canopy span can be obtained. In our validation experiments, the span ground truth is defined as the manually measured maximum canopy width, which is the distance between the two most distant leaf tips on the plant. To compare these estimates with real-world measurements, the COLMAP lengths must be converted to true scale. We use the measurement tools in CloudCompare to determine distances within the COLMAP point cloud and then apply a scale factor based on real-world reference lengths. The tool interface is shown in Fig. 3-13.

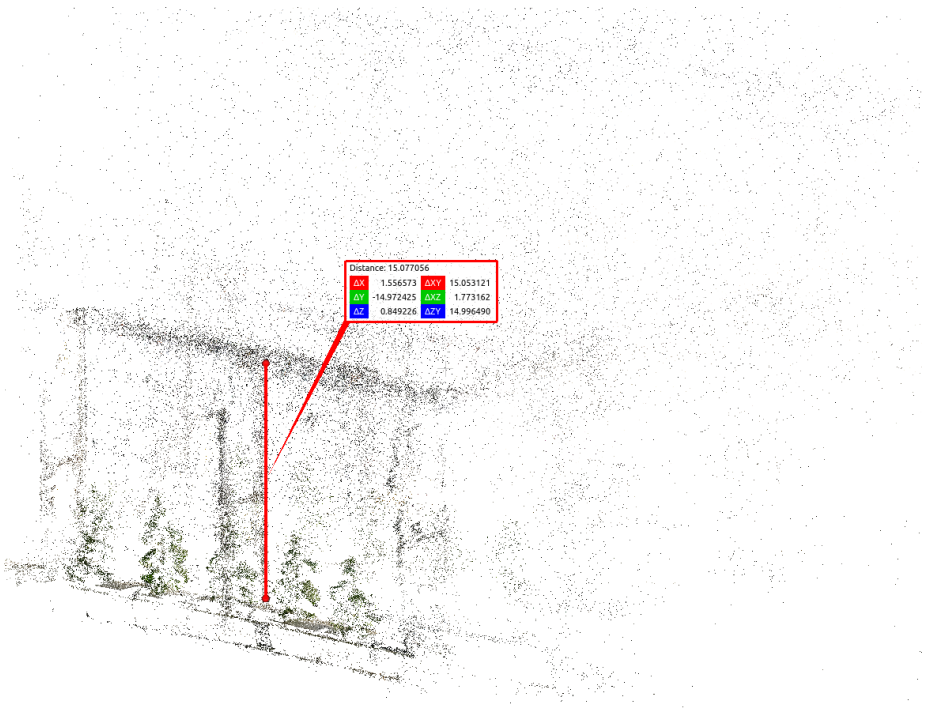


Fig. 3-13. Measurement of distances on a COLMAP point cloud using CloudCompare

doi:10.6342/NTU202503732

3.5.3 Muskmelon Growth Monitoring

Throughout the cultivation cycle, we sampled the same set of plants at regular intervals to monitor muskmelon growth. The experiment was conducted in a controlled greenhouse environment with 15 muskmelon plants arranged in three rows of five plants each. Data collection spanned from April 9 to May 11, 2025, with measurements taken at 2-day intervals. At each sampling point, our cooperative UAV system captured multi-angle images within the greenhouse, and we extracted individual plant height and canopy span measurements from the 3D point cloud reconstructions.

Prior to growth curve fitting, the extracted plant height and canopy span measurements from 3D point cloud data underwent quality control preprocessing, including outlier detection and removal. Identified outliers were replaced using linear interpolation to maintain temporal continuity, ensuring accurate modeling. We then organized each plant's height and canopy span data into a table indexed by sampling date and plant identifier. For each plant, we fitted its height and span trajectories with nonlinear growth models, including the Gompertz function (Gompertz, 1825), by minimizing the sum of squared residuals. Model performance was assessed using the coefficient of determination (R^2) and the root mean square error (RMSE). R^2 indicates the proportion of variance in the data explained by the model, while RMSE reflects the average magnitude of the residuals. A high R^2 and low RMSE indicate a good fit of the growth model to the data.

Under controlled greenhouse conditions, this integrated approach provides high-resolution monitoring of muskmelon development by combining UAV-based imaging, three-dimensional reconstruction, and quantitative growth modeling.

doi:10.6342/NTU202503732

CHAPTER 4

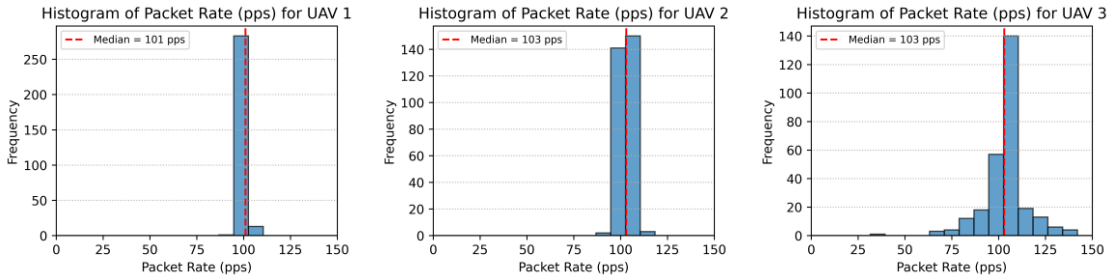
Results and Discussion

4.1 Evaluation of Multi-UAV Communication Performance

The performance of the UAV communication system was evaluated by analyzing packet transmission rates and stability during greenhouse operations. Network traffic was captured using pcap files and analyzed to assess communication reliability, which is critical for ensuring consistent control commands and real-time data acquisition in precision agriculture applications.

To evaluate the communication reliability between the edge computing node and each UAV, packet transmission rates were analyzed over a 4-5-minute operation in the greenhouse. Packet capture (pcap) files were collected and processed using a custom analysis pipeline, with rates computed in 1-second bins. Each UAV communicated via a dedicated 2.4 GHz Wi-Fi channel. Fig. 4-1 presents the histogram distributions of the packet rates for all three UAVs. UAV 1 demonstrated the most stable performance, with a median rate of 101 pps and 99.66% of values falling within the 95-105 pps range. UAV 2 had a slightly higher median and more variability, with 98.31% of values within the 95-110 pps range. UAV 3 also had a median of 103 pps but showed the greatest variability, with only 88.09% of values falling within the broader 85-120 pps range.

doi:10.6342/NTU202503732



(a) (b) (c)

Fig. 4-1. Histogram distributions of packet rates (pps) for three UAVs operating in greenhouse environment: (a) UAV 1 (median = 101 pps), (b) UAV 2 (median = 103 pps), and (c) UAV 3 (median = 103 pps).

The broader spread observed for UAV 3 is likely attributable to its location near greenhouse walls and glass windows. These structural elements can reflect or attenuate 2.4 GHz Wi-Fi signals and thereby cause unstable transmission. Despite these challenges, all three UAVs maintained consistent median packet rates, which underscores the robustness of the communication architecture. The system's wireless configuration uses detachable 4 dBi antennas to support reliable transmission under the difficult radio frequency conditions of a greenhouse. In such an environment, metallic frames, water-rich crops and electronic devices introduce multipath effects that complicate signal propagation. Occasional reductions in packet rate occurred when values fell below 95 percent of the median. These reductions were infrequent, affecting UAV 3 at a rate of 16.61 %, UAV 2 at 1.69 %and UAV 1 at just 0.34 %.

These results demonstrate that UAV 1 and UAV 2 benefited from favorable placements within the greenhouse, while UAV 3's location introduced more variability. Overall, the architecture successfully ensured stable communication for multi-UAV coordination, even under complex environmental interference.

doi:10.6342/NTU202503732

4.2 Multi-UAV Trajectory and Mapping Accuracy

4.2.1 Visual SLAM Map Building Result

For a visual navigation system, the map is critical, we present the results of the map-building process for three different flight scenarios using Visual SLAM. Each method involves loop closure to ensure that the entire path is accurately mapped.

In Fig. 4-2 shows the map for parallel-aisle flight, generated 229 keyframes and 11,141 landmarks. This approach uses parallel paths to capture keyframes at significant points during the flight. These keyframes represent crucial moments in the visual input, helping to maintain a consistent map as the drone progresses. The landmarks, which are distinctive points identified throughout the environment, provide spatial references and enable tracking across frames. The combination of keyframes and landmarks is vital for ensuring an accurate representation of the environment and minimizing any drift in the drone's path. Fig. 4-3 presents the map for closed-loop flight, resulting in 331 keyframes and 17,557 landmarks. This approach incorporates loop closure, a process that corrects errors in the path accumulated over time. When the drone revisits previously explored areas, loop closure ensures the map remains globally consistent by aligning overlapping segments. This method requires more keyframes and landmarks due to the additional data needed for accurate map stitching and path correction, providing a more reliable and robust reconstruction. As shown in Fig. 4-4, the map for multi-altitude flight generated 129 keyframes and 5,888 landmarks. This approach involves varying the drone's flight height during the mission. As the drone moves through different altitudes, it captures keyframes and landmarks that help create a reliable map across vertical layers of the environment. Although fewer keyframes and landmarks were generated compared to the

doi:10.6342/NTU202503732

other methods, they still contribute to the overall map, highlighting the adaptability of the SLAM system to different flight conditions.

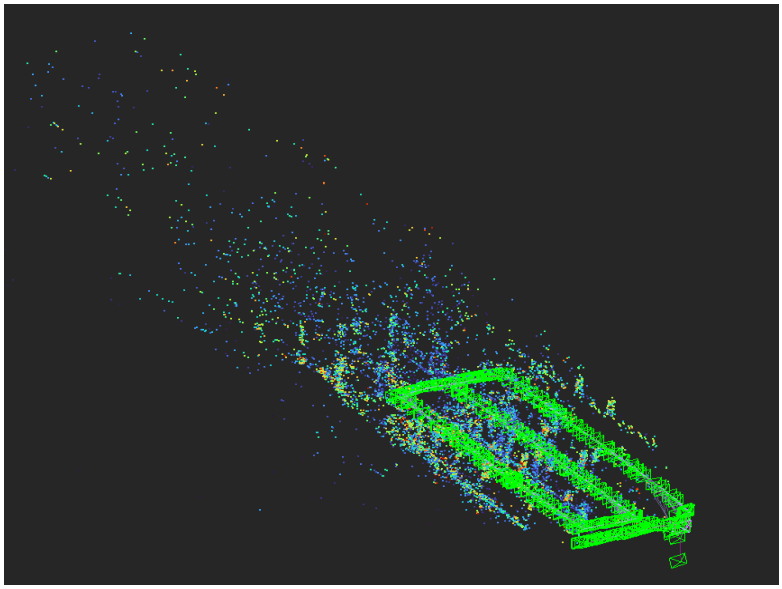


Fig. 4-2. Map Built Using Visual SLAM for Parallel-aisle Flight

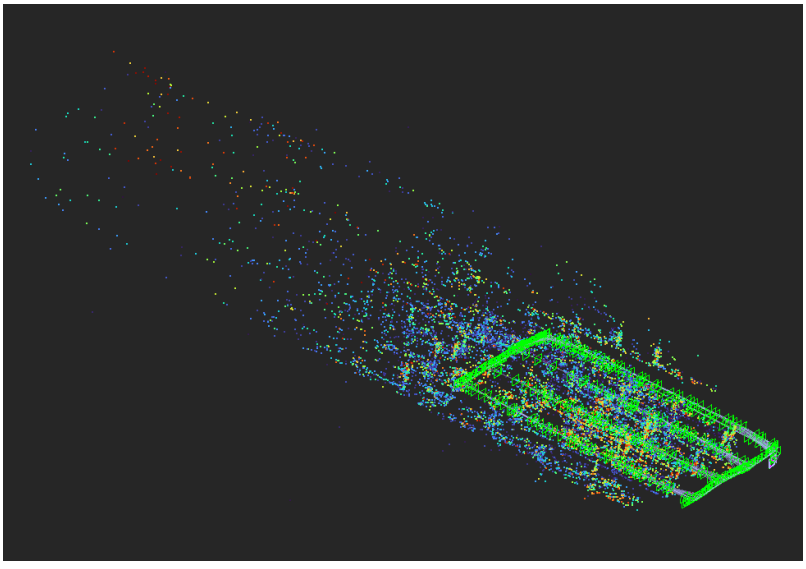


Fig. 4-3. Map Built Using Visual SLAM for Closed-loop Flight

doi:10.6342/NTU202503732

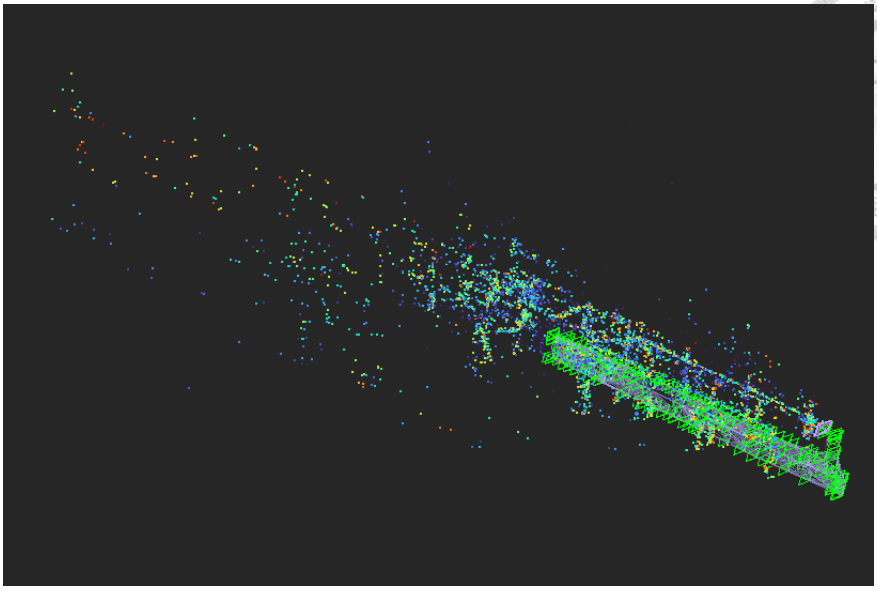


Fig. 4-4. Map Built Using Visual SLAM for Multi-altitude Flight

Keyframes and landmarks play a crucial role in the effectiveness of Stella VSLAM. Keyframes are selected to maintain the map's consistency, while landmarks provide spatial references necessary for localization. The number and quality of these components directly impact the accuracy of the generated map and the stability of the drone's trajectory throughout its journey.

4.2.2 Visual SLAM Map Optimization Result

Fig. 4-5 presents the sparse point cloud maps generated by Stella VSLAM after one, two and three handheld loops around the same rectangular circuit. In these renderings, blue points denote features that the system deems stable across multiple keyframes. After a single loop the stable features are concentrated mainly at the corners and along the nearest walls, resulting in modest overall coverage. With two loops the blue regions expand noticeably along the entire circuit, reflecting additional observations that

doi:10.6342/NTU202503732

reinforce feature persistence. The third loop yields the richest distribution of stable features, with walls, corners and mid-section surfaces all densely populated in blue.

We extracted the number of keyframes and landmarks for each mapping trial. The one-loop run produced 120 keyframes and 6175 landmarks. The two-loop experiment increased the keyframes to 189 and recorded 6057 landmarks. Only the three-loop pass drove both metrics upward, reaching 259 keyframes and 9849 landmarks. This trend in landmark count highlights that map quality cannot be judged by landmark quantity alone, since the second loop reduces weak or spurious features and trades landmark count for greater feature stability. The third loop delivers the highest combination of keyframe redundancy and landmark richness, while also emphasizing that map optimization must balance landmark abundance with feature reliability rather than pursue maximum point count.

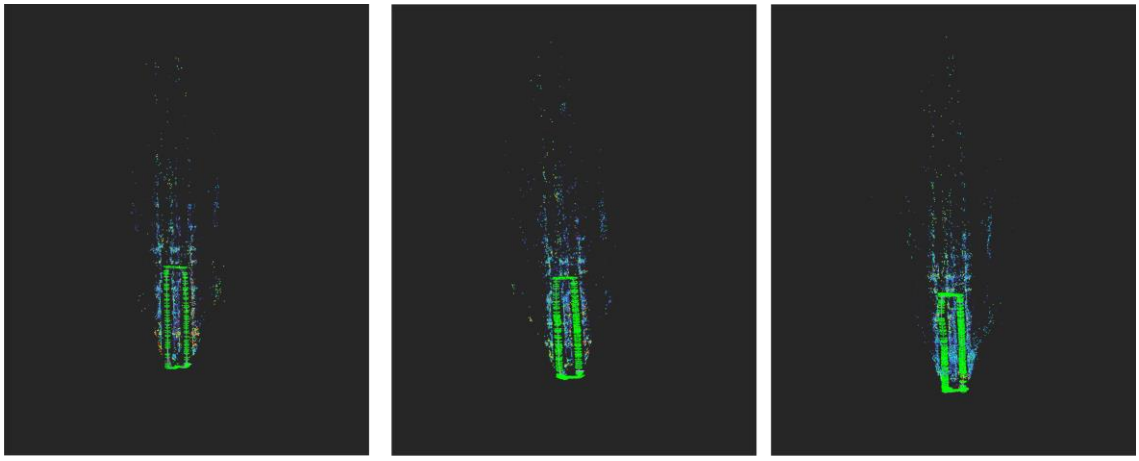


Fig. 4-5. Stella VSLAM map results: (a) one loop, (b) two loops, and (c) three loops

In order to further evaluate how the number of mapping loops affects optimization, we use UWB as ground truth to compare the accuracy of the SLAM trajectories. In Table 4-1 shows error metrics improve markedly from one loop to two loops. The RMSE

doi:10.6342/NTU202503732

decreases from 14 cm m to 8.5 cm and the MAE from 11.5 cm to 7 cm. Adding a third loop reverses these gains, with RMSE rising to 11.1 cm and MAE to 9.2 cm. These results indicate that two loops deliver the best trade-off between additional observations and error reduction. The extra pass strengthens stable feature matches and loop closures without accumulating excessive drift or adding redundant low quality points. By contrast, a third loop yields diminishing returns because the longer mapping time allows small pose errors to compound, spurious correspondences to appear and computational load to grow. All of these factors combine to reduce net accuracy relative to the two loop case.

Table 4-1. Error metrics for trajectories from different map building loops

Loop	RMSE (cm)	MAE (cm)	Std Dev (cm)	Max Error (cm)	Min Error (cm)	Median Error (cm)
1	14.0	11.5	8.0	48.3	0.1	9.9
2	8.5	7.0	4.7	22.6	0.1	6.0
3	11.1	9.2	6.3	34.9	0.1	7.8

We used UWB as the ground truth trajectory and compared it with the trajectory estimated by Stella VSLAM to further analyze the flight error of each mapping loop. We then presented the distribution of trajectory errors following one, two, and three handheld mapping loops.

In Fig. 4-6(a), which corresponds to a single loop, the MAE is approximately 11.5 cm. Red and yellow error points cluster at the circuit corners and along one side of the

doi:10.6342/NTU202503732

longitudinal segments, indicating that a single pass lacks sufficient viewpoint overlap and suffers from pronounced local drift. Green error points appear intermittently along the entire path, while blue stable features remain sparse. This pattern reflects how a single observation can miss many features.

Fig. 4-6(b) shows the result after two loops, with the MAE reduced to about 7 cm. Error points between 2.2 cm and 7 cm dominate, red and yellow points nearly disappear, and blue stable features increase and spread more uniformly, especially around the corners. These changes demonstrate that the second loop provides critical redundant observations that reinforce multi-view feature matches and loop closures, sharply reducing local drift and improving global consistency.

In Fig. 4-6(c) the MAE rises slightly to 9.2 cm. Although blue stable features still cover much of the path and green error points remain common, new red and yellow points appear at mid-section and corner locations. This behavior shows that the third loop, while adding feature redundancy, also introduces additional computational load and drift risk. In some areas those extra observations generate incorrect correspondences or small pose errors that offset the benefits seen with two loops. These three plots confirm that two loops achieve the most concentrated error distribution, the highest ratio of stable features, and the lowest MAE. The maximum error decreased from 48.3 cm with one loop to 22.6 cm with two loops, then rose to 34.9 cm with three loops. Standard deviation followed a similar trend. Stable features peaked at 55 % with two loops. Overall, two loops reduced MAE by 39 %, while a third loop increased it by 31 %, showing diminishing returns. A single loop suffers from insufficient overlap, and a third loop yields only marginal gains with the potential for degraded accuracy, making two loops the ideal strategy for high-precision map optimization.

doi:10.6342/NTU202503732

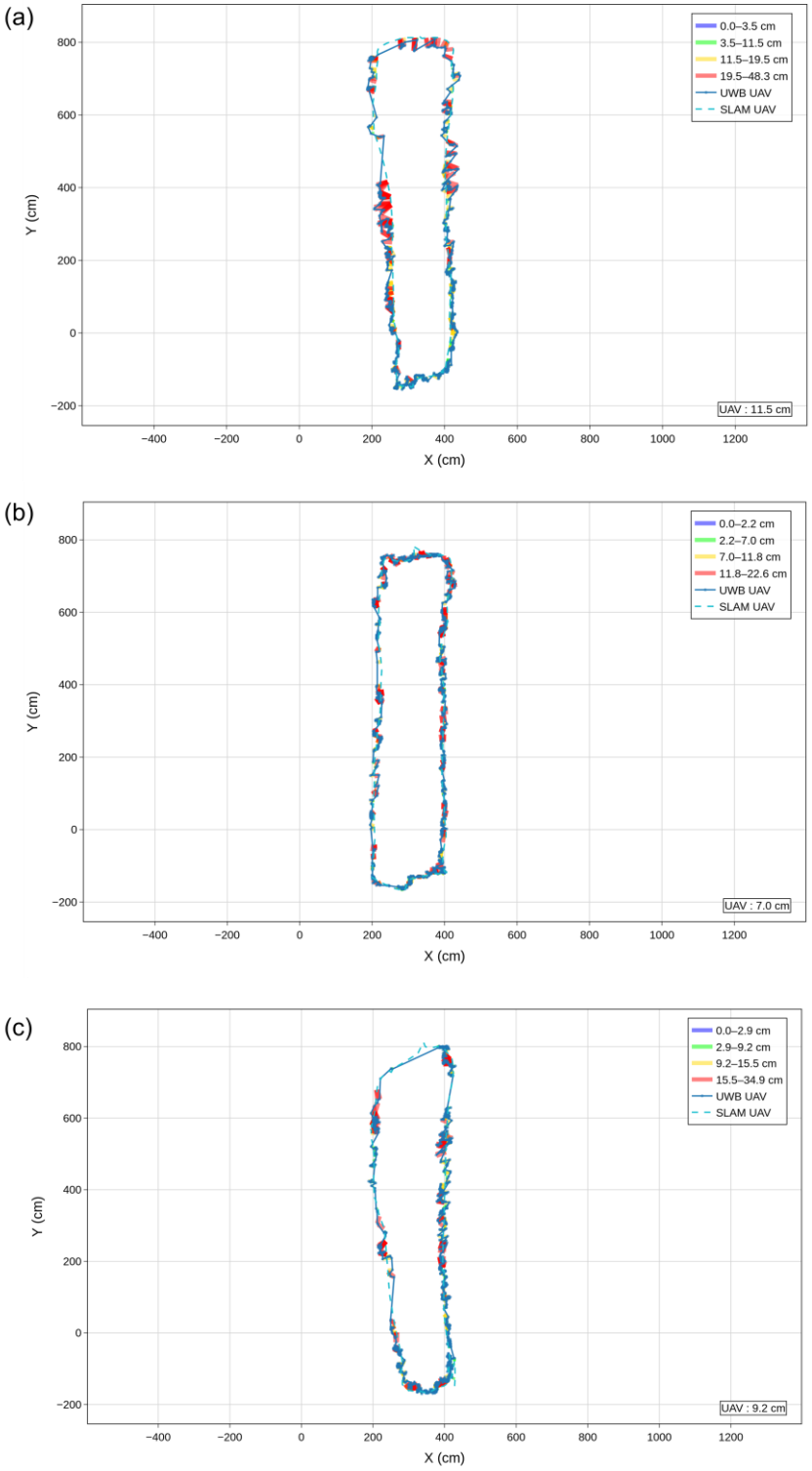


Fig. 4-6. Trajectory comparison of different loop map building, showing UWB positioning (solid lines) versus SLAM-based localization (dashed lines) with error magnitude indicated by color: (a) one loop, (b) two loops, (c) three loops.

doi:10.6342/NTU202503732

4.2.3 Parallel-aisle Flight Evaluation

Based on the path planning presented in Fig. 3-8, we tested each path three times, using UWB as the ground truth and comparing it with the positions estimated by Stella VSLAM to evaluate flight accuracy.

Each UAV executed the parallel aisle trajectory in three independent trials. For each run, we computed the root mean square error, MAE, standard deviation, maximum error, minimum error and median error by comparing the positions estimated by Stella VSLAM against UWB ground truth. The full set of results appears in Table 4-2, and all values are expressed in centimeters. UAV 1 yielded RMSE values of 10.5 cm, 12.7 cm and 11.8 cm across the first, second and third trials, and in the second trial it recorded its largest maximum deviation of 61.8 cm alongside its smallest error of 1 cm and its highest standard deviation of 9.9 cm. UAV 2 demonstrated the greatest consistency, achieving RMSEs of 7.4 m, 8.4 m and 8.8 cm in successive trials with maximum errors below 32.2 m, standard deviations never exceeding 5.2 cm and median errors remaining under 6.5 cm in every run. UAV 3 recorded RMSEs of 11.3 cm, 14.1 cm and 10.7 cm for trials one through three, with its highest single-trial deviation of 41.7 cm in the second trial, minimum errors of 0.1 cm to 0.3 cm in both the second and third trials and a peak standard deviation of 7.6 cm in the second run.

Overall, these findings indicate that the centrally located UAV 2 achieved superior SLAM based tracking fidelity. In contrast, flights by UAVs 1 and 3 along the greenhouse edges were more susceptible to multipath reflections and boundary effects and thus exhibited larger positional deviations.

doi:10.6342/NTU202503732

Table 4-2. Error metrics comparison across three UAVs for parallel-aisle trajectories

UAV ID	Trial ID	RMSE (cm)	MAE (cm)	Std Dev (cm)	Max Error (cm)	Min Error (cm)	Median Error (cm)
1		10.5	9.2	5.0	23.9	1.2	9.2
2	1	7.4	6.2	3.9	15.1	0.2	5.9
3		11.3	10.0	5.0	23.6	0.7	9.5
1		12.7	7.9	9.9	61.8	0.1	5.2
2	2	8.4	7.1	4.5	24.1	0.1	6.5
3		14.1	12.0	7.6	41.7	0.3	11.5
1		11.8	9.2	7.4	42.3	0.1	7.9
2	3	8.8	7.2	5.2	32.2	0.2	6.5
3		10.7	8.7	6.2	23.5	0.2	7.3

Based on Fig. 4-7, all three trials exhibit a consistent spatial error pattern. The central UAV's trajectory remains mostly blue and green, indicating low, uniform errors. In contrast, the two edge UAVs show yellow and red bands during takeoff and later stages, marking localized deviations. These high-error regions occur near greenhouse walls, where visual features are sparse and reflections stronger. Even with loop closure, edge trajectories retain these error pockets, while the middle UAV maintains stable accuracy. This repeated pattern highlights the importance of symmetric, feature-rich observations for SLAM accuracy and the vulnerability of edge flights to boundary effects.

doi:10.6342/NTU202503732

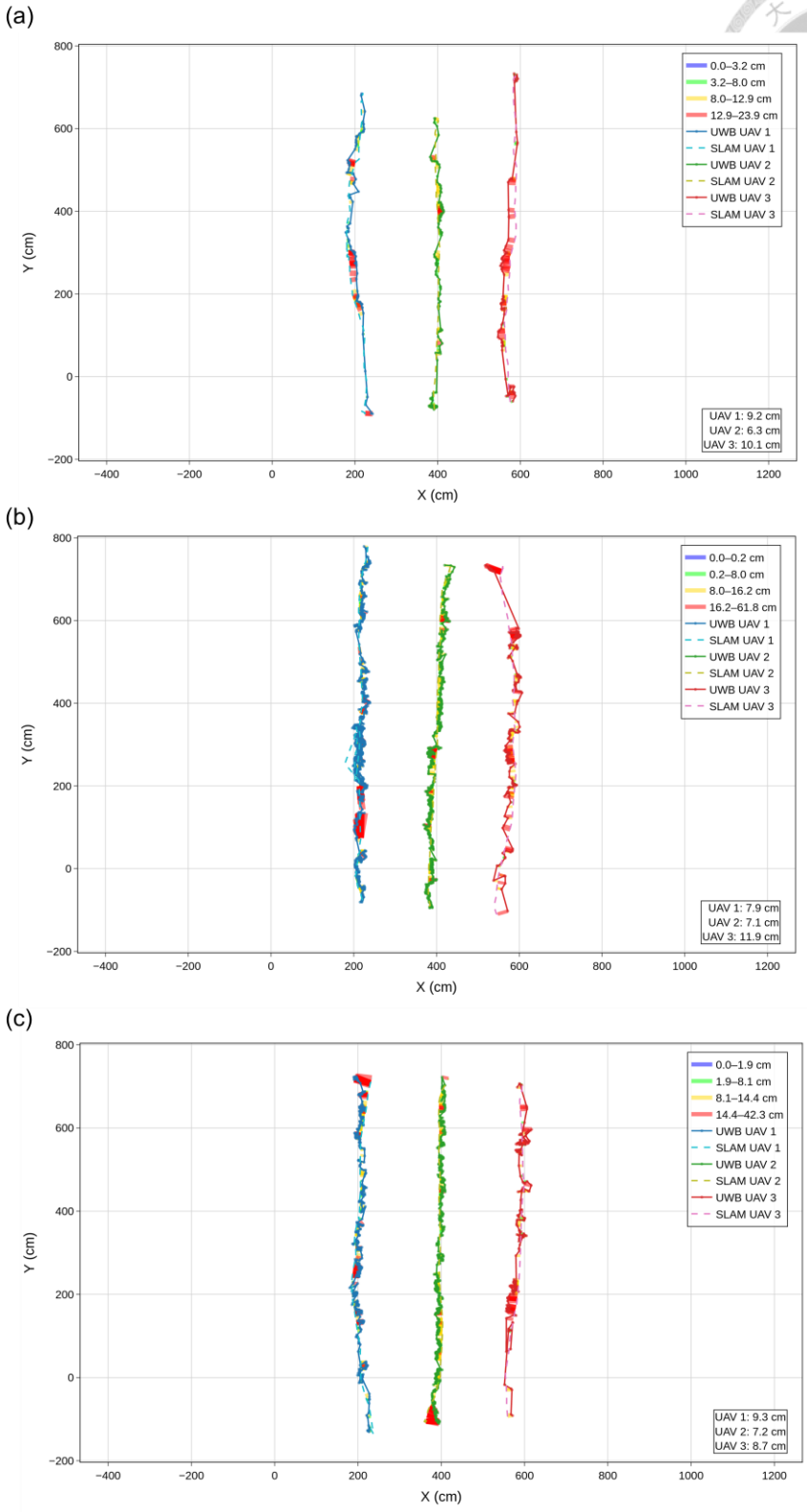


Fig. 4-7. Trajectory comparison of parallel-aisle flights with three UAVs, showing UWB positioning (solid lines) versus SLAM-based localization (dashed lines) with error magnitude indicated by color: (a) Trial 1, (b) Trial 2, (c) Trial 3.

doi:10.6342/NTU202503732

4.2.4 Closed-loop Flight Evaluation

Table 4-3 offers a comparative view of error metrics for three UAVs following closed-loop trajectories. UAV 1 exhibited a root mean square error between 6.3 and 8.6 cm and a mean absolute error below 6.9 cm. The standard deviation ranged from 3.8 to 5.2 cm and the median error remained between 4.3 and 6.0 cm, demonstrating consistently stable localization performance across all trials. UAV 2 recorded a root mean square error ranging from 7.8 to 10.9 cm and a mean absolute error from 6.2 to 8.8 cm. Its standard deviation rose to 6.5 cm in the second trial, and the maximum error reached 37.1 cm. The median error fell between 5.0 and 7.4 cm, indicating moderate consistency with occasional higher deviations. UAV 3 showed a root mean square error between 8.4 and 10.6 cm and a mean absolute error from 6.8 to 8.0 cm. The standard deviation peaked at 6.9 cm in the final trial, and the maximum error reached 55.5 cm, the highest among all UAVs. The median error varied from 5.9 to 6.1 cm, suggesting relatively stable central tendency despite larger outliers.

From a comparative perspective, UAV 1 consistently achieved the lowest errors and narrowest variability, suggesting better overall stability. UAV 2 exhibited moderate errors with occasional spikes, while UAV 3 showed the highest variability and maximum errors, particularly in the third trial. This pattern indicates that UAVs operating near greenhouse boundaries or under less feature-rich conditions may be more prone to larger deviations, highlighting the influence of environmental factors and positional asymmetry on SLAM localization accuracy.

These observations highlight that the first UAV maintained the most stable performance under closed loop conditions while the third UAV experienced the greatest variability and highest single trial deviation.

doi:10.6342/NTU202503732

Table 4-3. Error metrics comparison across three UAVs for closed-loop trajectories

UAV ID	Trial ID	RMSE (cm)	MAE (cm)	Std Dev (cm)	Max Error (cm)	Min Error (cm)	Median Error (cm)
1		6.3	5.1	3.8	19.7	0.1	4.3
2	1	7.8	6.3	4.7	24.2	0.1	5.0
3		8.4	6.8	4.8	26.7	0.1	5.9
1		7.7	6.2	4.6	23.6	0.1	5.2
2	2	10.9	8.8	6.5	37.1	0.1	7.4
3		9.4	7.6	5.6	29.5	0.1	6.5
1		8.6	6.9	5.2	36.8	0.1	6.0
2	3	8.9	7.3	5.1	27.1	0.1	6.4
3		10.6	8.0	6.9	55.5	0.1	6.1

Based on Fig. 4-8, all three closed-loop trials exhibit a stable spatial error pattern. UAV 2's path is predominantly blue and light green, indicating tightly constrained deviation from UWB ground truth. In contrast, the edge UAVs consistently show yellow and red patches along outbound and return paths, reflecting localized error spikes caused by sparse landmarks and reflective surfaces. These high-error regions persist and intensify across trials, while UAV 2 maintains its narrow low-error band. The consistent color map for UAV 2 highlights the advantage of balanced, feature-rich observations, whereas persistent yellow and red patches for edge UAVs emphasize boundary effects and landmark scarcity despite loop closure optimization.

doi:10.6342/NTU202503732

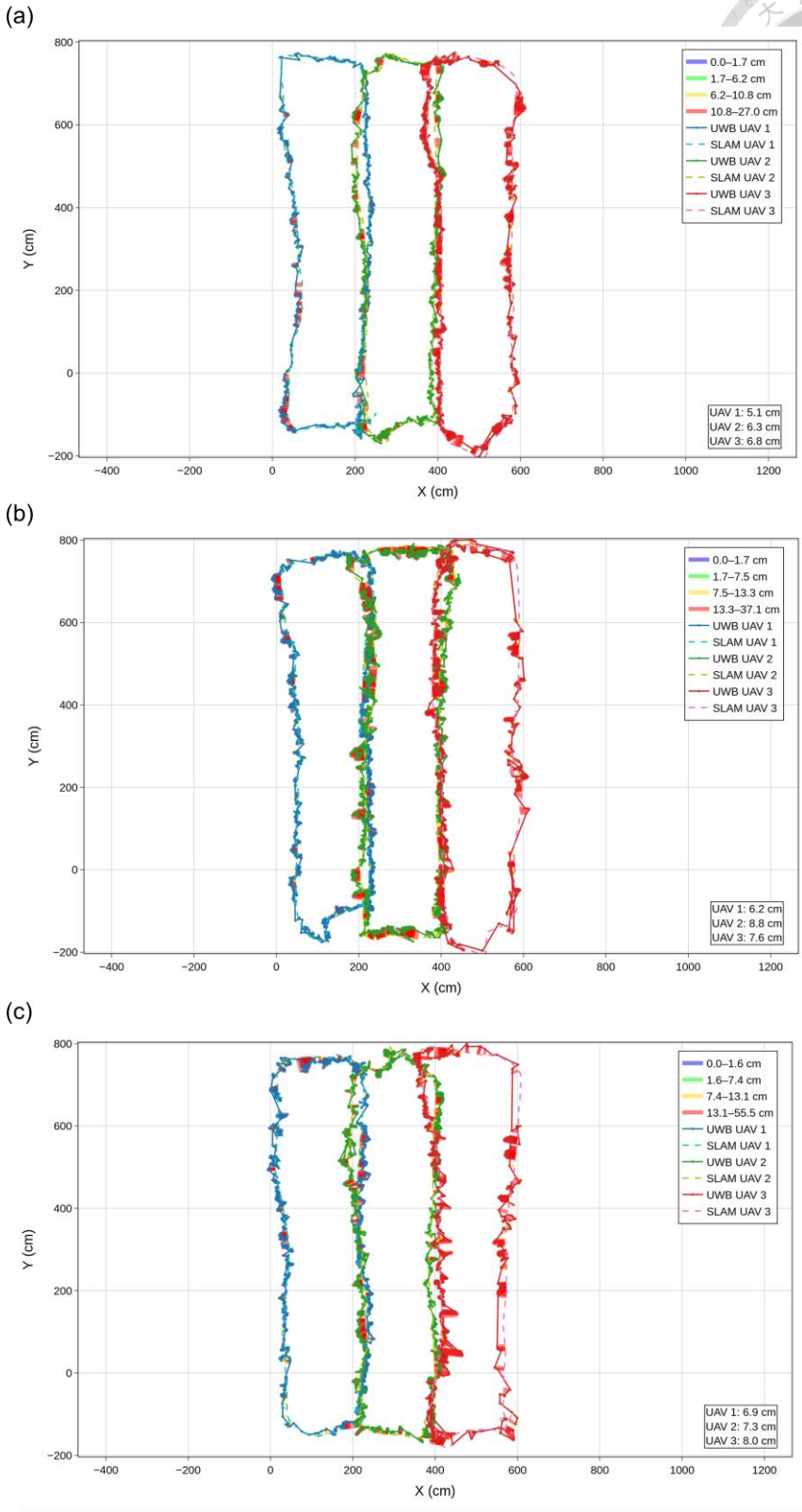


Fig. 4-8. Trajectory comparison of closed-loop flights with three UAVs, showing UWB positioning (solid lines) versus SLAM-based localization (dashed lines) with error magnitude indicated by color: (a) Trial 1, (b) Trial 2, (c) Trial 3.

doi:10.6342/NTU202503732

4.2.5 Multi-altitude Flight Evaluation

In Table 4-4 presents a comparative analysis of error metrics for three UAVs following multi altitude trajectories over three trials. The UAV 1 showed root mean square error values of 12.1 cm in the first trial and then improved to 9.6 cm and 10.4 cm in the second and third trials respectively. Its mean absolute error likewise dropped from 10 cm to 7.4 cm before rising slightly to 8.7 cm. The standard deviation contracted from 6.7 cm to 5.6 cm by the third run. These trends indicate that the first UAV achieved its best consistency during the second trial but experienced modest variations thereafter. The second UAV began with a root mean square error of 7.8 cm which increased to 13.4 cm in the second trial and then receded to 9.2 cm in the third trial. Its mean absolute error followed a similar pattern rising from 6.3 cm to 11.2 cm before improving to 7.6 cm. The standard deviation peaked at 7.4 cm in the second trial and then declined to 5.2 cm. These fluctuations suggest that the UAV 2 encountered greater challenges in maintaining stable localization when altitude changes were more pronounced. The UAV 3 recorded root mean square error values of 13.1, 5.1 cm and 11.8 cm across trials one two and three respectively. Its mean absolute error fell from 11.4 cm to 4.1 cm before rising again to 10.4 cm. The standard deviation reached a minimum of 2.8 cm in the second trial and then increased back to 5.6 cm. These results show that the third UAV achieved its best accuracy in the second trial when altitude variations were moderate but saw its performance degrade when altitude transitions became more extreme.

UAV 1 maintained stable performance with slight improvements in the second trial. UAV 2 showed larger fluctuations, struggling during greater altitude changes. UAV 3 achieved its best accuracy in the second trial but degraded with more extreme altitude variations. Moderate altitude transitions generally resulted in more stable localization.

doi:10.6342/NTU202503732

Table 4-4. Error metrics comparison across three UAVs for multi-altitude trajectories

UAV ID	Trial ID	RMSE (cm)	MAE (cm)	Std Dev (cm)	Max Error (cm)	Min Error (cm)	Median Error (cm)
1		12.1	10.0	6.7	24.7	0.4	9.7
2	1	7.8	6.3	4.6	21.8	0.1	5.8
3		13.1	11.4	6.4	30.1	0.1	10.7
1		9.6	7.4	6.1	31.5	0.1	5.7
2	2	13.4	11.2	7.4	33.6	0.1	10.1
3		5.1	4.1	2.8	13.2	0.1	4.0
1		10.4	8.7	5.6	25.8	0.3	7.6
2	3	9.2	7.6	5.2	18.7	0.1	6.8
3		11.8	10.4	5.6	26.6	0.1	9.7

Based on Fig. 4-9, the multi-altitude flight trials reveal localized error clusters despite each UAV maintaining a fixed altitude. Red and yellow segments appear intermittently along portions of the trajectories, particularly during the early and middle stages of flight, before returning to stable green and blue regions. These deviations are likely caused by insufficient visual features and unstable repeated feature matching at different altitudes. Additionally, airflow disturbances inside the greenhouse may introduce UAV vibrations, further amplifying SLAM errors. Overall, while altitude control remains stable, these factors contribute to noticeable vertical localization inaccuracies.

doi:10.6342/NTU202503732

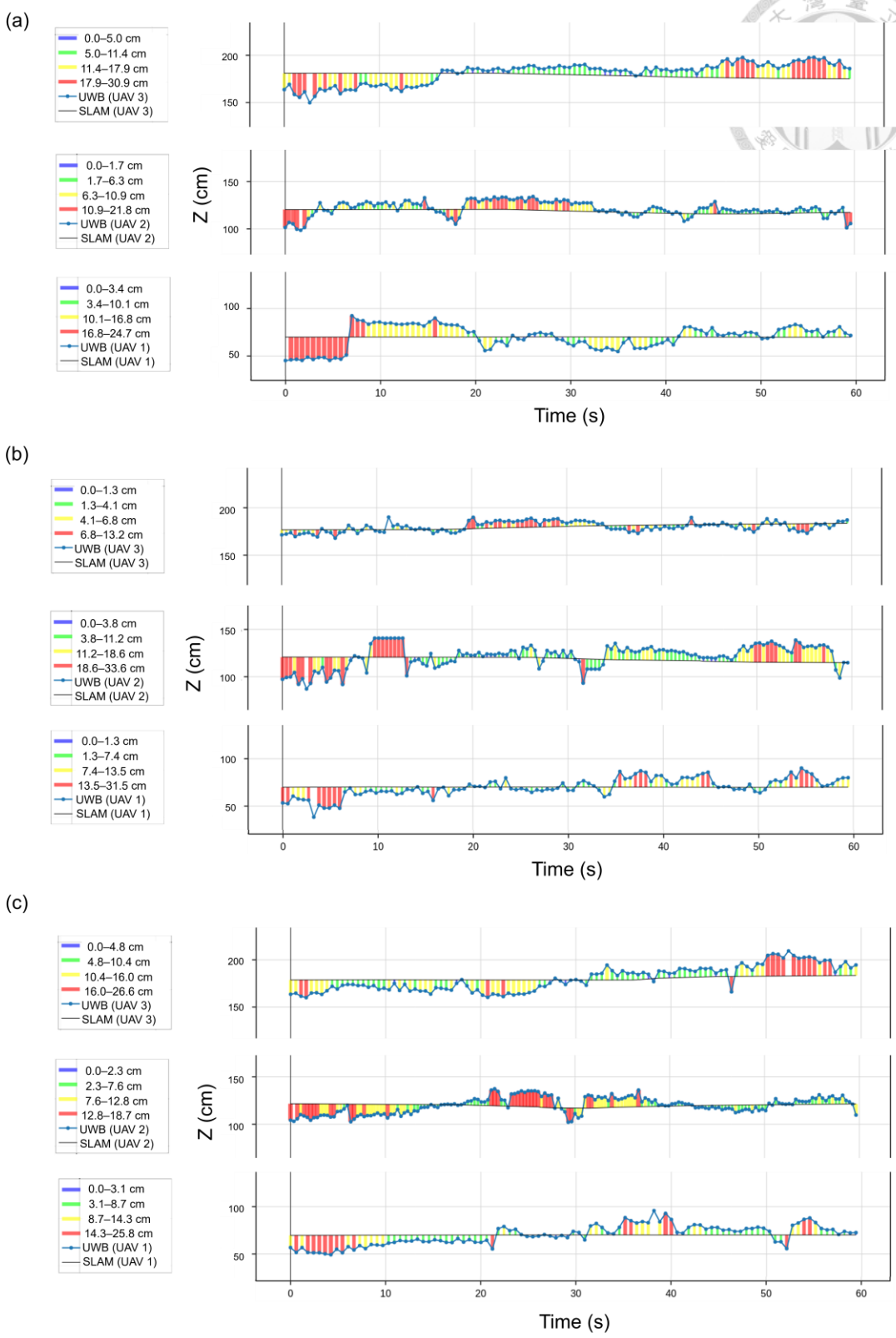


Fig. 4-9. Trajectory comparison of multi-altitude flights with three UAVs, showing UWB positioning (solid lines) versus SLAM-based localization (dashed lines) with error magnitude indicated by color: (a) Trial 1, (b) Trial 2, (c) Trial 3.

doi:10.6342/NTU202503732

4.2.6 Multi-UAV Flight-Path Performance Comparison

Table 4-5 provides a detailed comparison of mean absolute error (mAE), mission duration, and area coverage rate (ACR) for the three coordinated flight strategies: Parallel-aisle, Closed-loop, and Multi-altitude. Among them, the Closed-loop strategy achieves the lowest overall mAE, with UAV 1 reaching as low as 6.1 cm, demonstrating the highest localization precision. However, this accuracy comes at the cost of efficiency, as the mission requires the longest completion time (average 4.1 minutes) and results in a moderate ACR of 17.0 m^2/min.

In contrast, the Parallel-aisle strategy offers the most balanced performance. It maintains relatively low errors across UAVs, with a minimum mAE of 6.8 cm for UAV 2 and an average coverage rate of 21.5 m^2/min, the highest among all strategies. Mission time is significantly shorter at 1.6 minutes, indicating superior operational efficiency while still providing acceptable accuracy. The Multi-altitude strategy prioritizes speed, completing the mission in just 1.1 minutes. However, its overall precision is moderate, with mAE values ranging from 8.4 to 8.7 cm, and it achieves the lowest ACR of 10.8 m^2/min. The variability in errors, particularly for UAV 3 with a standard deviation of 4.0 cm, suggests less stable localization performance under rapid altitude adjustments.

Overall, when measurement accuracy is the primary goal, the Closed-loop approach is recommended due to its superior precision. For applications requiring both precision and high coverage efficiency, the Parallel-aisle strategy offers the most advantageous trade-off. Meanwhile, the Multi-altitude approach is best suited for time-sensitive missions where reduced coverage and moderate accuracy are acceptable. These findings highlight the trade-offs between accuracy, efficiency, and coverage, providing guidance for selecting optimal flight strategies under varying operational priorities.

doi:10.6342/NTU202503732

Table 4-5. Statistical Summary and Comparison of Flight Paths

Flight Path	UAV ID	Mean MAE (cm)	Std Dev (cm)	mAE (cm)	Mean Mission time (min)	ACR (m²/min)
Parallel-aisle	1	8.8	0.8	8.8±0.8	1.6	21.5
	2	6.8	0.6	6.8±0.6		
	3	10.2	1.7	10.2±1.7		
Closed-Loop	1	6.1	0.9	6.1±0.9	4.1	17.0
	2	7.5	1.3	7.5±1.3		
	3	7.5	0.6	7.5±0.6		
Multi-altitude	1	8.7	1.3	8.7±1.3	1.1	10.8
	2	8.4	2.5	8.4±2.5		
	3	8.6	4.0	8.6±4.0		

4.3 Evaluation of Multi-UAV Cooperative

Building on the flight path design defined in the Fig. 3-9, we executed both single-UAV and cooperative multi-UAV missions to quantitatively compare the two deployment strategies in terms of area coverage efficiency and energy utilization.

As summarized in Table 4-6, the cooperative multi-UAV system completed the full-coverage mission over a 4.1 m × 8.5 m survey area in just 1.25 minutes, representing a

doi:10.6342/NTU202503732

73.2 % reduction compared to the 4.67 min required by a single UAV. Parallel operation increased the area coverage rate to 27.88 m²/min, which is 3.73 times higher than the 7.47 m²/min achieved by the single-UAV flight, thereby substantially accelerating the overall survey process.

Table 4-6. Performance metrics for single-UAV vs. cooperative multi-UAV missions over field

Metric	single-UAV	multi-UAV
Flight Time Efficiency (FTE, min)	4.67	1.25
Area Coverage Rate (ACR, m² / min)	7.47	27.88
Total Battery Consumption (%)	42	40
Energy Efficiency (m² / % battery)	0.83	0.87

With respect to energy consumption, the cooperative multi-UAV deployment consumed a total of 40% battery across all UAVs, slightly less than the 42% used by the single-UAV system. Despite involving three separate agents, the parallel operation maintained comparable overall energy usage while still improving efficiency. Specifically, the energy efficiency reached 0.87 m² per percent battery, slightly higher than the 0.83 m² / % battery attained by the single-UAV mission.

Overall, these findings demonstrate that, under identical path planning and hardware conditions, distributed multi-UAV cooperation not only markedly reduces mission

doi:10.6342/NTU202503732

duration and increases coverage throughput, but also greatly enhances energy utilization, thereby outperforming the conventional single-UAV approach across all key performance metrics.

4.4 3D Reconstruction Analysis for Plant

4.4.1 Sparse 3D Reconstruction with GLOMAP

The GLOMAP reconstruction process produced a sparse point cloud that accurately captures the structural layout of muskmelon plants, including their spatial arrangement, stem architecture, and leaf distribution, as shown in Fig. 4-10. Despite the relatively low density, this point cloud provides essential initialization for Gaussian Splatting, preserving the spatial relationships among plants and distinguishing them clearly from the surrounding infrastructure.

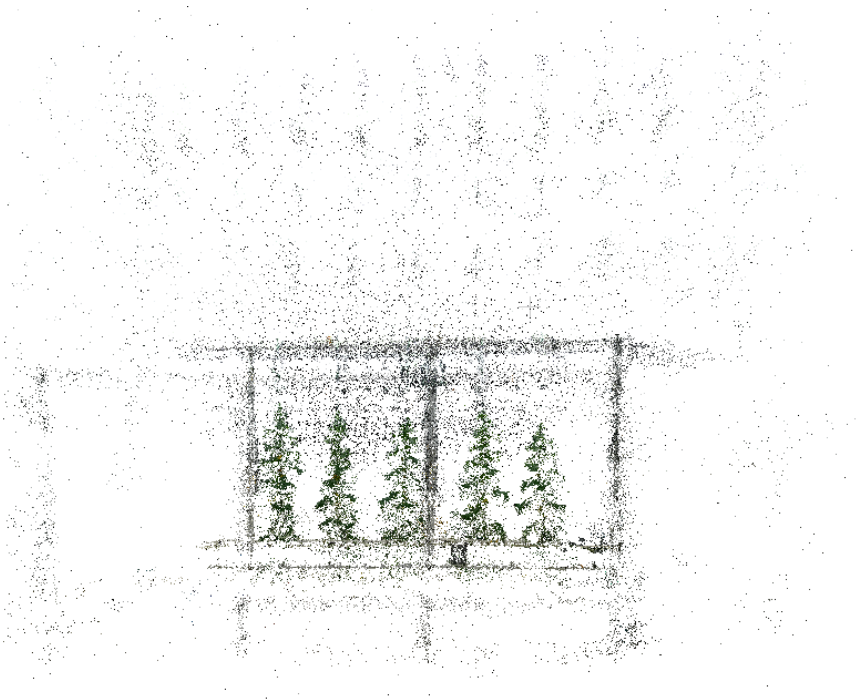


Fig. 4-10. Sparse point cloud reconstruction of muskmelon plants generated by GLOMAP.

doi:10.6342/NTU202503732

According to our flight trajectories we divided the acquired images into three categories. The single-side category contains single-view images, the merged category contains two-view images of each plant and the three-height category contains single-view images captured at three different altitudes. Image acquisition for the Single-Side and merged datasets took place between April 8 and May 11, 2025, while the three-height dataset was collected from April 21 to May 11, 2025. During reconstruction some datasets failed to yield complete point clouds. We believe this was caused by insufficient image overlap, occlusions from dense foliage or inconsistent lighting. Out of 102 reconstruction attempts in the single-side category 85 succeeded. In the merged category 43 of 102 attempts succeeded and in the three-height category 38 of 63 attempts succeeded. The lower success rates in the merged and three-height sets may be due to the greater difficulty of matching features across multiple views and the challenges of maintaining consistent imaging conditions at different altitudes. We will use the successfully reconstructed point clouds for Gaussian Splatting.

4.4.2 Evaluation of Gaussian splatting for 3D plant reconstruction

In this section, we present the results of our 3D reconstruction using Gaussian Splatting based on the GLOMAP output. We analyzed a total of 85 Single-Side reconstructions, 43 Merged reconstructions, and 38 Three-Height reconstructions to compare the effectiveness of different capture methodologies. Table 4-7 summarizes the detailed characteristics of these reconstruction datasets, including acquisition periods,

doi:10.6342/NTU202503732

total attempts, and successful reconstructions. For each flight video, the average image overlap was 60%, ensuring sufficient feature matching for reliable reconstruction.

Table 4-7. Detailed Characteristics of 3D Reconstruction Datasets

Scene Category	Acquisition Period	Total Attempts	Successful Reconstructions
Single-Side	Apr 9 – May 11, 2025	102	85
Merged	Apr 9 – May 11, 2025	102	43
Three-Height	Apr 21 – May 11, 2025	63	38

To provide a direct visual comparison between the reconstructed models and the real plants, Fig. 4-11 illustrates the Single-Side reconstruction. In this figure, (a) shows the reconstructed plant, (b) and (c) display detailed views of the upper and lower leaves, and (d) presents the corresponding original plant image. This figure highlights that the upper leaf structure reconstructed by the Single-Side method appears blurry and lacks detailed definition compared to the real plant.

Fig. 4-12 depicts the Merged reconstruction. Similar to the Single-Side figure, (a) presents the reconstructed plant, (b) and (c) show upper and lower leaf details, and (d) is the original plant image. The Merged method introduces noticeable noise and misalignment artifacts, resulting in a less clean reconstruction compared to the original plant.

doi:10.6342/NTU202503732

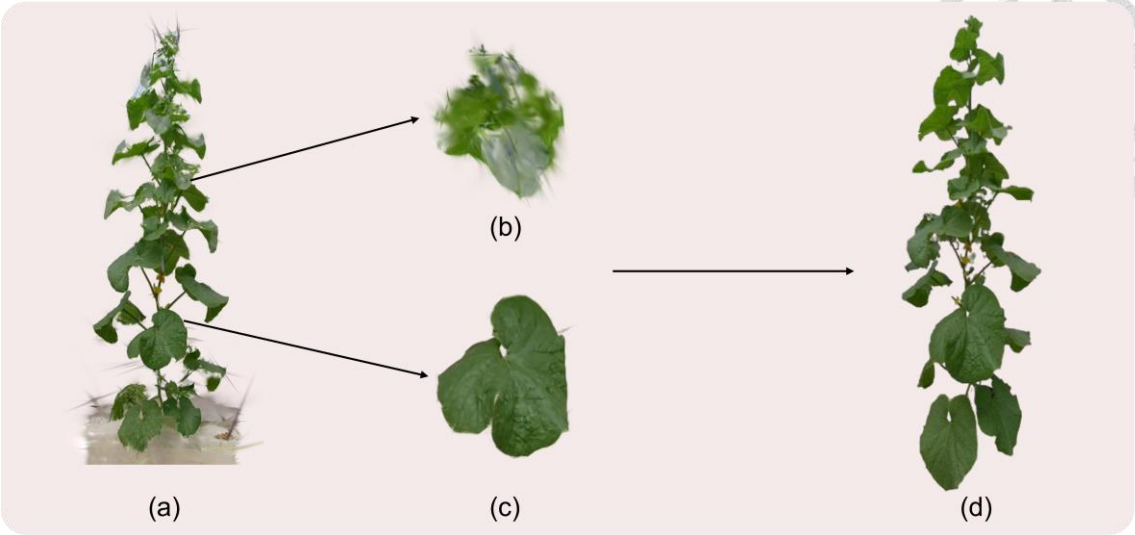


Fig. 4-11. Single-Side reconstruction compared with the original plant image: (a) reconstructed plant, (b) upper leaf detail, (c) lower leaf detail, and (d) corresponding original plant image.

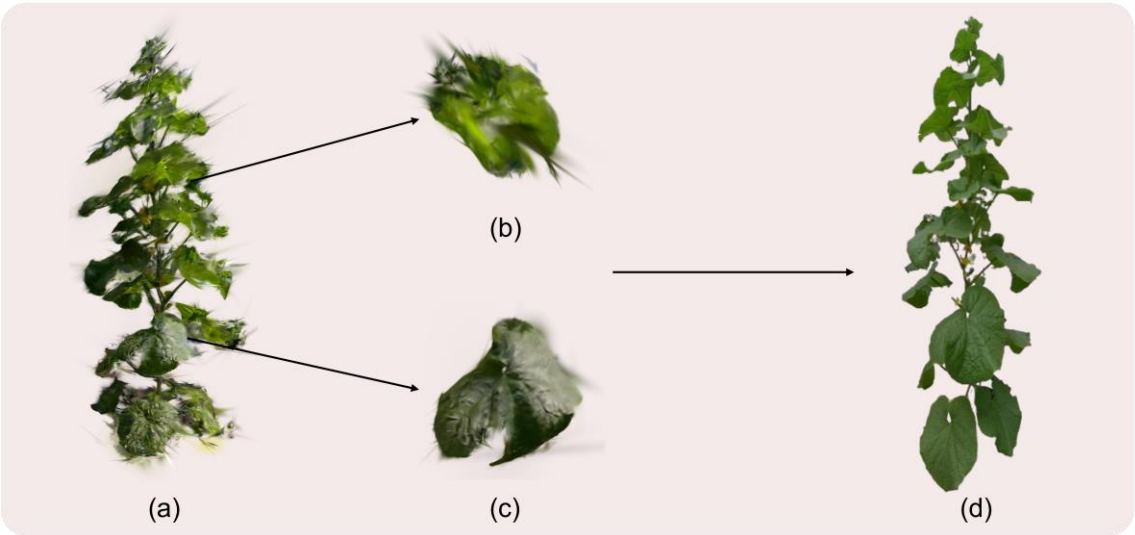


Fig. 4-12. Merged reconstruction compared with the original plant image: (a) reconstructed plant, (b) upper leaf detail, (c) lower leaf detail, and (d) corresponding original plant image.

Fig. 4-13 shows the Three-Height reconstruction. As with the other figures, (a) displays the reconstructed plant, (b) and (c) highlight detailed views of the leaves, and (d)

doi:10.6342/NTU202503732

is the original image. This method achieves better clarity and leaf definition, particularly in the upper plant regions, offering a reconstruction more faithful to the real plant.

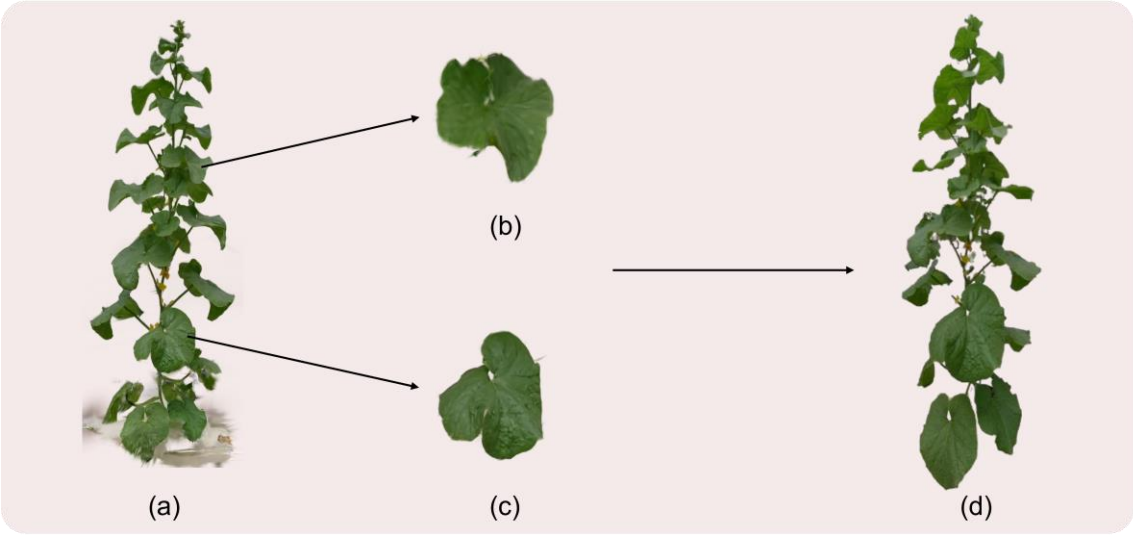


Fig. 4-13. Three-Height reconstruction compared with the original plant image: (a) reconstructed plant, (b) upper leaf detail, (c) lower leaf detail, and (d) corresponding original plant image.

Fig. 4-14 shows the reconstruction results of the same row of plants using three different capture methodologies. For consistent comparison, we removed the background from all reconstructions. Each methodology is presented with three components: the camera arrangement diagram on the left side, the complete plant reconstruction in the center, and detailed views of upper and lower leaves from a selected plant on the right side.

In the Single-Side reconstruction shown in Fig. 4-14(a), we observe that the upper portions of the plants appear blurry with less defined leaf structures. This limitation is

doi:10.6342/NTU202503732

particularly evident in the upper leaf detail shown in the top right image, where the leaf lacks clear definition and shows reduced detail fidelity. The Merged method shown in Fig. 4-14(b) exhibits significantly more noise compared to the other techniques. This noise is likely due to the challenges in accurately aligning and merging images captured from two different sides of the plants. The misalignment creates artifacts during point cloud generation, resulting in the scattered noise patterns visible throughout the reconstruction. The Three-height reconstruction method shown in Fig. 4-14(c) effectively addresses these limitations by capturing images at different heights. By incorporating multiple vertical perspectives, this method successfully reconstructs the upper portions of the plants with much greater clarity. The improvement is clearly visible when comparing the upper leaf details across all three methods. The Three-Height method provides significantly better leaf definition and structural integrity in the upper regions of the plants. The lower portions of the plants, as shown in the bottom right images of each row, show relatively consistent quality across all three methods, indicating that the primary differences in reconstruction quality are most pronounced in the upper regions of the plants. This suggests that the Three-Height method offers a more complete and accurate reconstruction by addressing the limitations of single-perspective captures, particularly for taller plants with complex canopy structures.

Further examination supports these observations. The Single-Side method lacks multi-view coverage, resulting in sparse upper-canopy points, incomplete geometry, and less reliable trait measurements. The Merged method introduces spatial misalignments that cause artifacts and fragmented leaf structures, limiting analysis accuracy. The Three-Height method overcomes these issues by improving vertical feature detection, enabling more complete 3D reconstructions that enhance trait measurement accuracy.

doi:10.6342/NTU202503732

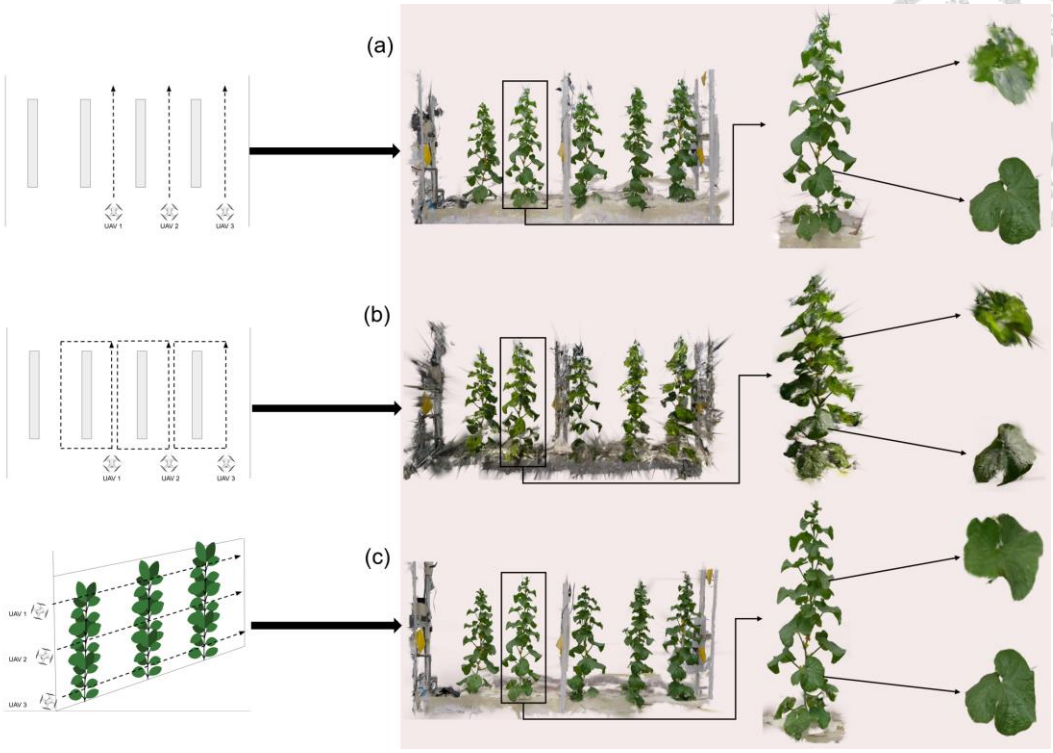


Fig. 4-14. Comparison of three different plant capture methodologies: (a) Single-Side, (b) Merged, and (c) Three-Height.

Despite the increased noise observed in Fig. 4-14, the Merged method demonstrates superior performance in capturing the complete plant structure. As shown in Fig. 4-15(b), the Merged approach successfully reconstructs back-side information of the plant that is not available in the other methods. This is a significant advantage over Single-Side in Fig. 4-15(a) and Three-height in Fig. 4-15(c) approaches, which primarily capture information from only one viewing angle, resulting in incomplete data about the back surfaces of the plants. For reference, Fig. 4-15(d) presents the original backside photo of the plant, enabling a direct visual comparison between the reconstructed models and the true plant morphology. This back-side representation is particularly valuable for comprehensive plant phenotyping and structural analysis, as it reveals features that would otherwise be completely missing in single-angle captures. The ability to reconstruct both front and back surfaces provides a more complete morphological dataset for plant assessment.

doi:10.6342/NTU202503732

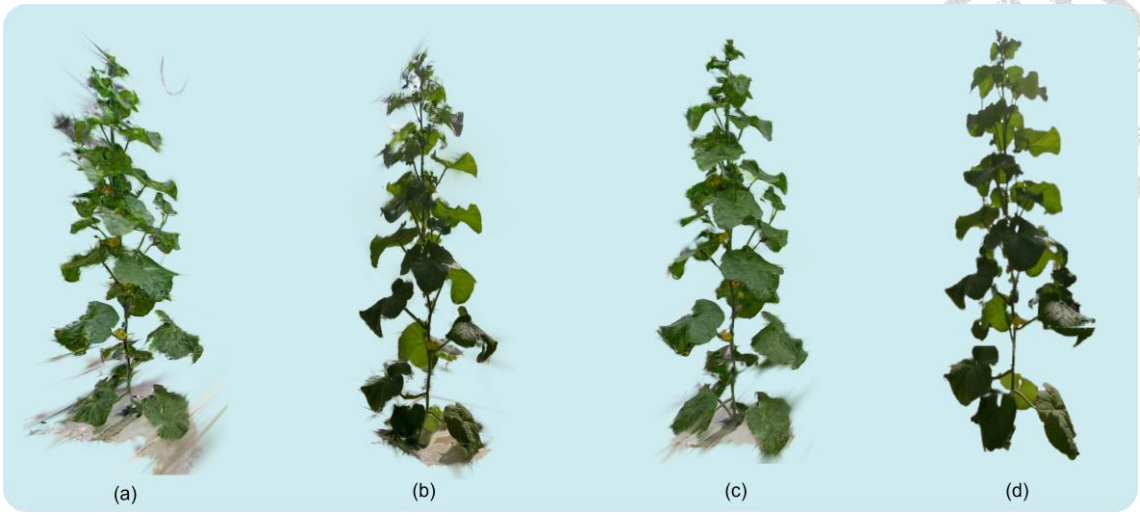


Fig. 4-15. Comparison of plant reconstructions from different angles. (a) Single-Side reconstruction showing backside view. (b) Merged reconstruction showing backside view. (c) Three-Height reconstruction showing backside view. (d) Original photo showing backside view of the plant.

After evaluating the performance of different reconstruction methods, we further present the reconstructed models of a plant across its growth stages. Fig. 4-16 shows the dense three-dimensional models of a single plant at five successive stages from seedling to maturity. Data collected on and after April 21 were processed using the three-height reconstruction method to capture comprehensive multi-view structural information while earlier time points were handled with the more stable single-side workflow. The series illustrates that plant height increases continuously over time as the small form seen during the seedling stage develops into the tall structure of the mature plant. At the same time leaf number and canopy spread exhibit significant expansion evolving from sparse young leaves into a dense arrangement. Moreover outlines of the main stem and lateral branches become increasingly distinct with branch nodes and stem thickness variations accurately captured at each stage. This set of models demonstrates that the multi-UAV system combined with the GLOMAP and SuGaR reconstruction pipeline can reliably capture

doi:10.6342/NTU202503732

detailed plant morphology throughout the entire growth cycle and provides a clear visual basis for subsequent quantitative measurements of plant height and canopy spread.

Fig. 4-16. Three-Dimensional Reconstruction Models of Muskmelon Plants at Multiple Growth Stages

To complement our visual comparison of the Single-Side, Three-Height, and Merged reconstruction methods, we conducted a quantitative evaluation using three standard image quality metrics: Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Learned Perceptual Image Patch Similarity (LPIPS). Fig. 4-17 presents the average values of these metrics across the three reconstruction categories, allowing for an objective assessment of structural fidelity, noise resilience, and perceptual

doi:10.6342/NTU202503732

similarity. The results highlight noticeable differences in reconstruction quality, particularly in how each method preserves fine details and overall image consistency.

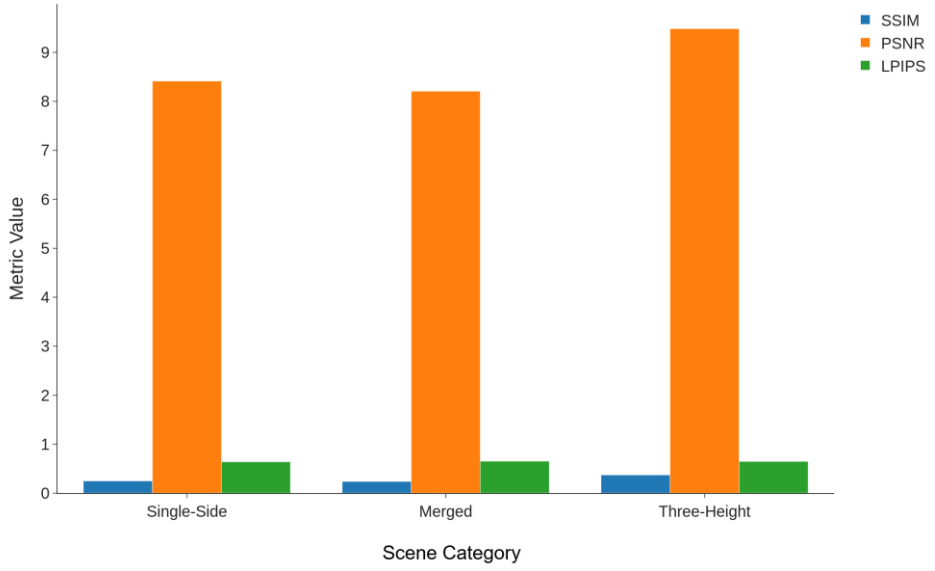


Fig. 4-17. Average SSIM, PSNR, and LPIPS values across different reconstruction methods: Merged, Single-side, and Three-Height scenes.

As shown in Table 4-8, the Single-side method, with the largest sample size, offers moderate performance and balanced visual quality, though with some limitations in capturing upper canopy details. The Merged method yields the lowest PSNR and SSIM due to alignment noise, but maintains a comparable LPIPS, highlighting its perceptual similarity and unique ability to reconstruct back-side structures. LPIPS values across all methods remain consistent between 0.64 and 0.65, suggesting comparable perceptual quality despite variations in structural fidelity. While Three-Height excels in vertical detail with low noise, and Merged captures full structural geometry, Single-side offers reliable performance with robust sampling.

doi:10.6342/NTU202503732

Table 4-8. Image Quality Metrics for Different Reconstruction Methods

Scene Category	PSNR ↑	SSIM ↑	LPIPS ↓
Single-Side	0.25	8.41	0.64
Merged	0.24	8.21	0.65
Three-Height	0.37	9.48	0.65

4.5 Phenotyping Result of Muskmelon

4.5.1 Plant Measurement Analysis

Using the plant reconstruction results, we further employed the reconstructed point clouds to estimate each plant's height and canopy span. For each of the three reconstruction methods, we computed the estimated height and span and calculated their errors relative to the ground-truth measurements.

In Table 4-9 showing the plant measurements of the height and canopy span, which evaluated by comparing each method's estimated height and canopy span to the corresponding ground-truth values. The Single-side reconstruction exhibits a height MAE of 7.3 cm and a span MAE of 5.8 cm. The Merged method exhibits a height MAE of 6.6 cm and a span MAE of 7.3 cm. RMSE values from both methods range from 9.9 cm to 10.3 cm, indicating close agreement with the true dimensions. By contrast, the Three-Height approach shows substantially larger deviations and greater variability. Height estimates remain under 8% for all methods. Span error for Three-Height reaches 17.4%, while Single-side and Merged record 10.8% and 13.4%, respectively. These findings

doi:10.6342/NTU202503732

indicate that the Single-side and Merged techniques provide more consistent and reliable phenotypic measurements than the Three-Height strategy.

Table 4-9. Key Phenotyping Metrics by Reconstruction Method

Reconstruction Method	Height MAE (cm)	Height RMSE (cm)	Height MAPE (%)	Span MAE (cm)	Span RMSE (cm)	Span MAPE (%)
Single-side	7.3	9.9	6.5	5.8	9.2	10.8
Merged	6.6	10.3	5.9	7.3	10.7	13.4
Three-Height	11.4	15.2	7.3	10.7	14.8	17.4

We further analyze the results through scatter plots and error distributions. In Fig. 4-18(a), the height scatter plot shows that calculated values closely follow the ground-truth line with only minor deviations. The corresponding error histogram in Fig. 4-18(c) is tightly clustered between 0 and 10 centimeters and shows a slight tail toward negative values. This indicates that the Single-side approach produces consistently accurate height estimates with only occasional underestimation.

Fig. 4-18(b) presents the span scatter, which displays a similarly tight pattern, and Fig. 4-18(d) shows its error distribution centered near 0 cm with most errors below 10 cm. These results demonstrate that Single-side reconstruction yields a compact and reliable error distribution.

doi:10.6342/NTU202503732

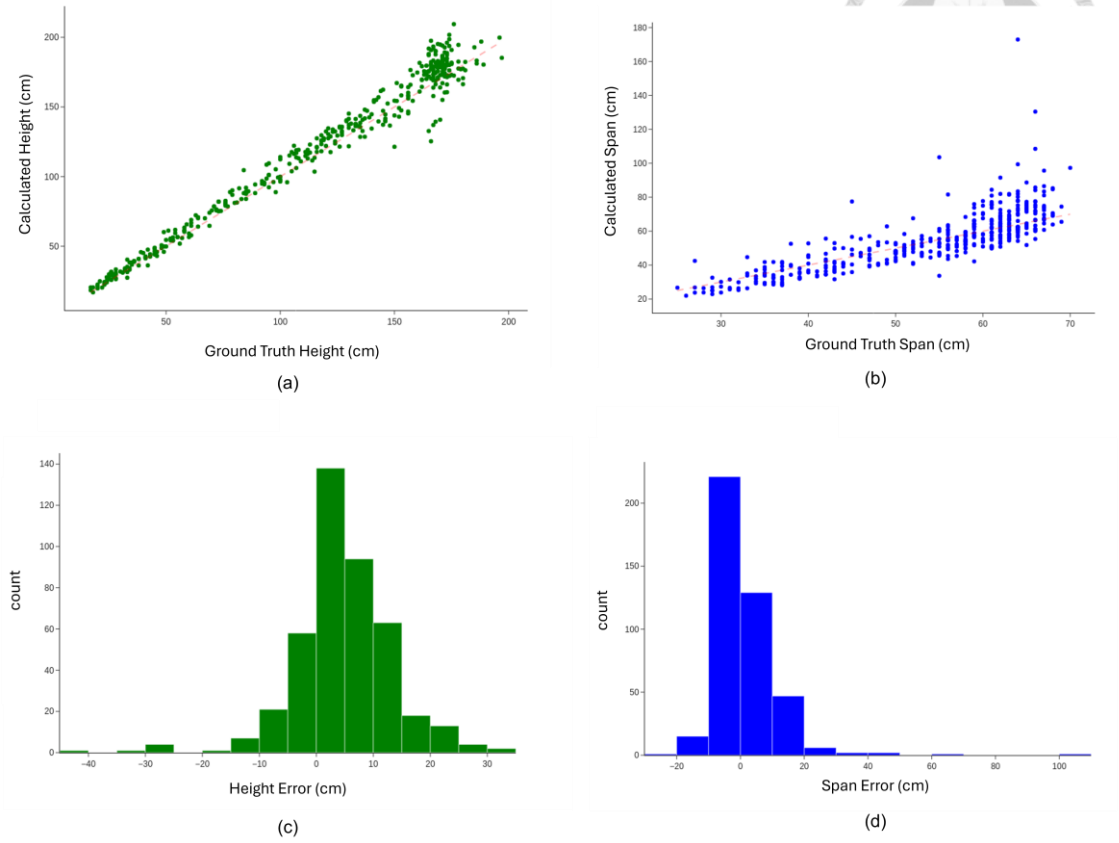


Fig. 4-18. Plant Height and Span Measurement Errors for Single-Side Reconstruction: (a) scatter plot of calculated height versus ground-truth height, (b) scatter plot of calculated span versus ground-truth span, (c)error distribution of calculated height, and (d) error distribution of calculated span

Fig. 4-19 illustrates the Merged method. In Fig. 4-19(a), the height scatter remains well aligned with the identity line but with slightly greater spread than Single-side, especially at higher values. The height error histogram in Fig. 4-19(c) shows that most errors are concentrated between -15 to 20 centimeters, with a very small number of samples exhibiting errors below -80 cm. In Fig. 4-19(b) the span scatter exhibits broader dispersion, and Fig. 4-19(d) shows its error distribution extending further into higher positive values than in the height case. These patterns suggest that the Merged technique

doi:10.6342/NTU202503732

tends to mildly overestimate both height and span and exhibits greater variability when integrating multiple viewpoints.

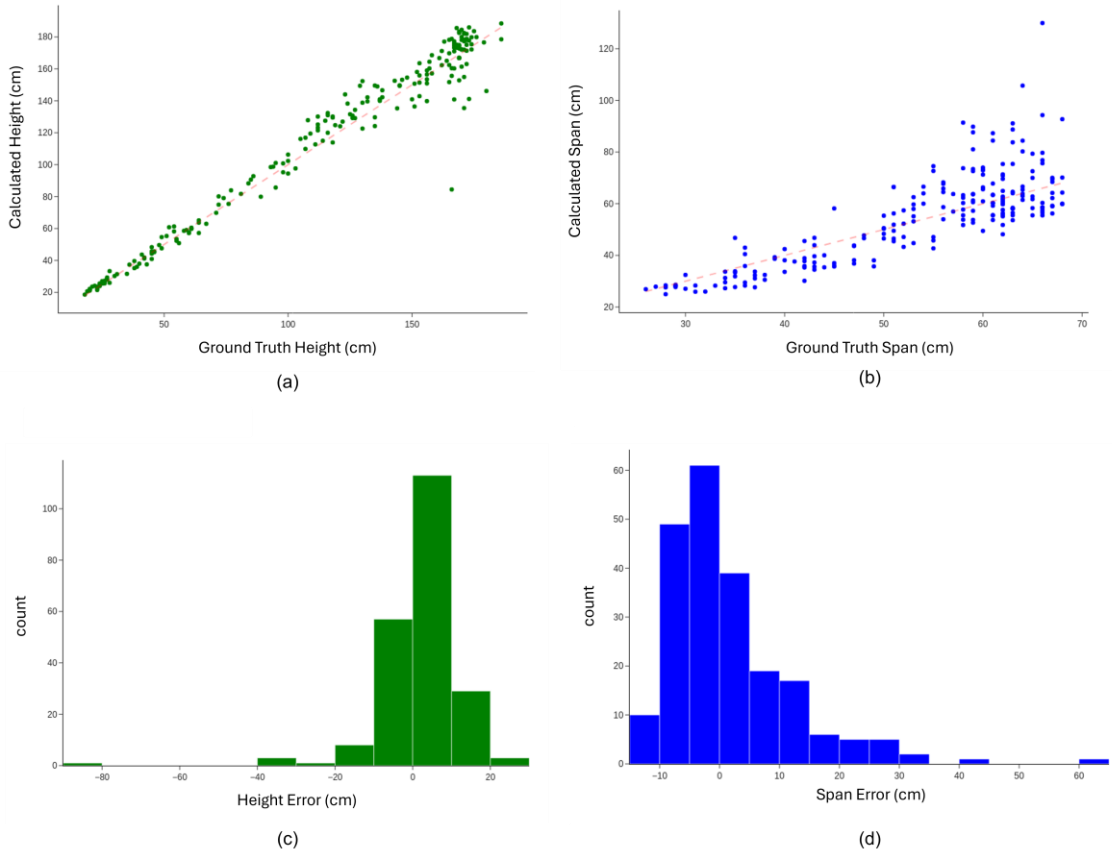


Fig. 4-19. Plant Height and Span Measurement Errors for Merged Reconstruction: (a) scatter plot of calculated height versus ground-truth height, (b) scatter plot of calculated span versus ground-truth span, (c) error distribution of calculated height, and (d) error distribution of calculated span

Fig. 4-20 illustrates the Three-Height method. In Fig. 4-20(a), the height scatter plot displays substantial scatter around the identity line and notable outliers at the upper range. The height error histogram which shows in Fig. 4-20(c) covers a wide band from -60 to 25 centimeters indicating large and uneven deviation. In Fig. 4-20(b) presents the span

doi:10.6342/NTU202503732

scatter likewise shows wide variability and in Fig. 4-20(d) shows the span error histogram extends from -20 to 60 centimeters. These wide and irregular distributions reveal that Three-Height reconstruction often introduces alignment noise that leads to large and unpredictable measurement errors.

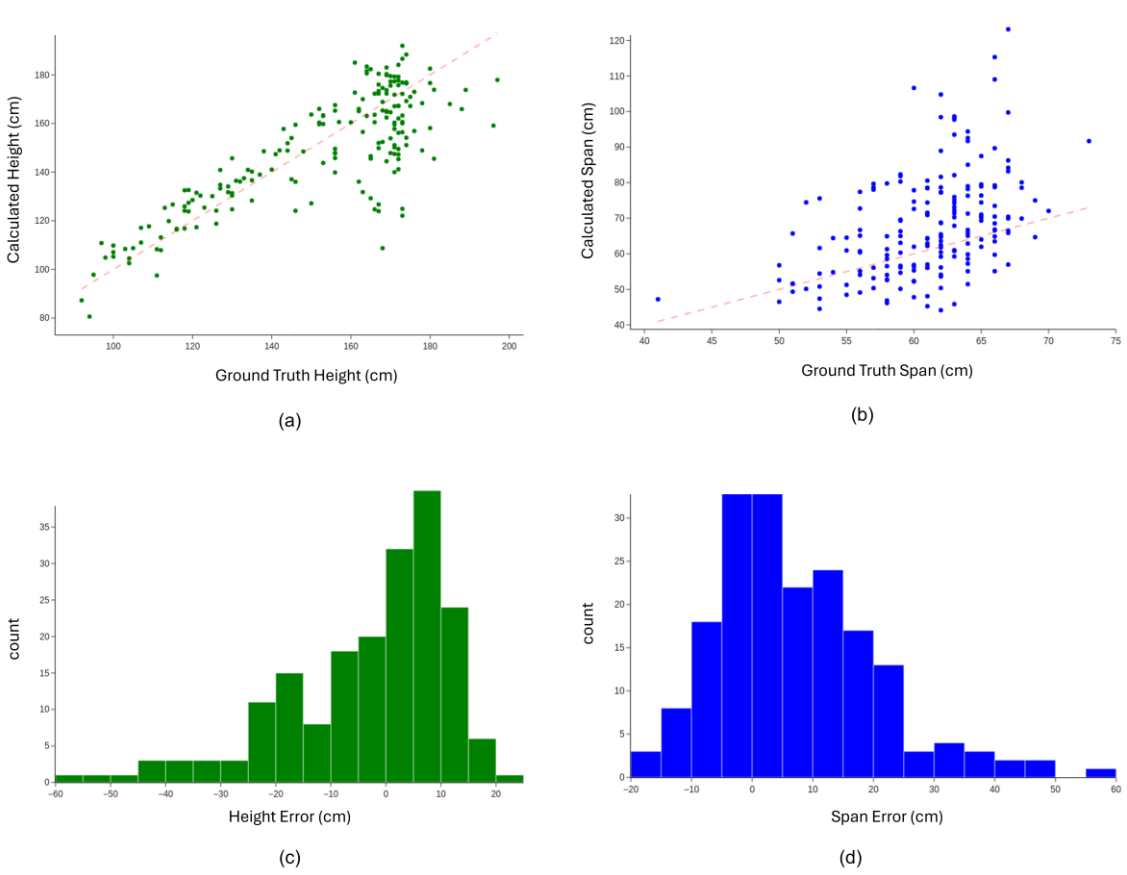


Fig. 4-20. Plant Height and Span Measurement Errors for Three-Height Reconstruction: (a) scatter plot of calculated height versus ground-truth height, (b) scatter plot of calculated span versus ground-truth span, (c)error distribution of calculated height, and (d) error distribution of calculated span

Comprehensive error analysis shows that the Single-Side reconstruction produces the smallest and most concentrated deviations in height and canopy span measurements and delivers the best performance. The Merged method yields slightly larger errors but

doi:10.6342/NTU202503732

still outperforms the Three-Height strategy. The Three-Height approach exhibits the most scattered and largest errors and is not recommended for applications requiring stable measurements.

4.5.2 Muskmelon Growth Monitoring Analysis

Based on our previous analysis of reconstruction errors in plant height and canopy span measurements, we used the Single-Side method to monitor growth. We measured height and span for each of fifteen plants, plotted individual growth curves, and calculated R^2 and RMSE against the ground-truth values.

Table 4-10 shows the height and span results for each plant using the Single-Side reconstruction. Height tracking was very reliable. The R^2 values ranged from 0.92 to 0.99 for all fifteen plants. The RMSE values did not exceed 15 centimeters. Row C had the best height performance with an average R^2 of 0.99 and an average RMSE of 5.9 centimeters. Row B followed with an average R^2 of 0.97 and an average RMSE of 7.9 centimeters. Row A showed the largest variation with an average R^2 of 0.95 and an average RMSE of 10.4 centimeters. Within that row Plant A5 had the lowest R^2 at 0.92 and the highest RMSE at 14.5 centimeters. Span estimation proved more challenging. The R^2 value fell between 0.77 and 0.98. The RMSE values ranged from 1.8 to 5.9 centimeters. The weakest span estimates occurred for Plant C2 with an R^2 of 0.77 and an RMSE of 5.9 centimeters. Plant A2 had an R squared of 0.84 and an RMSE of 5.2 centimeters. On average Row B led all rows for span accuracy with an average R^2 of 0.90 and an average RMSE of 3.3 centimeters. Row A followed with an average R^2 of 0.90 and an average RMSE of 4.1 centimeters. Row C had an average R^2 of 0.88 and an average RMSE of 4.2 centimeters.

doi:10.6342/NTU202503732

Table 4-10. Individual Plant Growth Monitoring Metrics

Row	Plant ID	Height R^2	Height RMSE (cm)	Span R^2	Span RMSE (cm)
A	1	0.99	5.1	0.95	2.8
	2	0.94	12.8	0.84	5.2
	3	0.94	12.6	0.94	3.5
	4	0.98	6.8	0.95	3.2
	5	0.92	14.5	0.80	5.8
B	1	0.98	7.2	0.79	4.7
	2	0.99	6.2	0.97	1.9
	3	0.95	10.5	0.98	1.8
	4	0.98	6.5	0.85	4.0
	5	0.97	9.2	0.92	4.1
C	1	0.99	5.9	0.95	2.5
	2	0.99	6.0	0.77	5.9
	3	0.99	4.2	0.81	5.3
	4	0.99	6.1	0.95	3.2
	5	0.98	7.2	0.91	4.1

doi:10.6342/NTU202503732

To evaluate the Single-Side reconstruction method across all specimens we generated growth curves for height and span over time for each plant. The fifteen plants are organized into three rows labeled A, B and C.

Fig. 4-21 shows the height and span growth curves for the five plants in row A measured in centimeters. All five entered a rapid growth phase beginning around April 13 and lasting until April 27. Thereafter the rate of height increase slowed as each plant neared its maximum. Plant 1 grew from 20 centimeters on April 9 to 174 cm by May 11. Plants 2 and 3 reached 172 cm. Plant 4 peaked at 167 cm after a slightly delayed surge, and Plant 5, despite greater variability in midseason measurements, converged near 171 cm. Canopy span expanded more slowly. Over the first two weeks the five plants' spans increased from approximately 30 cm to about 55 cm before leveling off. Then entered a saturation phase after April 25. Plants 1 and 5 achieved the widest canopies at 68 cm. Plant 4 reached 67 cm. Plants 2 and 3 followed a nearly identical trajectory. These curves confirm a shared growth rhythm with individual variation in the timing of rapid height gain and in final canopy breadth. Height growth peaked before canopy expansion plateaued, reflecting a transition from vertical shoot extension to lateral leaf expansion as the plants matured.

Fig. 4-22 presents the height and span growth curves for the five plants in Row B. The stages mirror those seen in Row A. From about April 13 until April 27 they entered a rapid growth period before the rate of height gain slowed as they neared their mature size. Plant 1 grew from around 20 cm on April 9 to about 168 cm by May 11. Plant 2 reached a similar final height of 173 cm but showed slightly greater day-to-day variability in the mid-season measurements. Plant 3 peaked at 166 cm and exhibited the greatest scatter in its data points during the exponential phase. Plant 4 followed closely behind at 172 cm, and Plant 5 rose to 171 cm with the steadiest trend line of the group. Canopy

doi:10.6342/NTU202503732

span in Row B expanded more gradually. During the first two weeks span increased from approximately 35 cm to about 52 cm. After April 25 the rate of canopy expansion slowed as leaves began to reach maximum size. Plants 1 and 3 both achieved final spans of about 65 cm, while Plant 2 extended slightly further to 67 cm. Plant 4 produced the smallest canopy at 64 cm and Plant 5 reached the greatest final span of 67 cm. These curves illustrate that although all five specimens share a common growth schedule, they vary in the timing and magnitude of both vertical and lateral development.

The five plants in Row C follow the same pattern seen in Rows A and B. In Fig. 4-23 presents the height and span growth curves for the five plants in Row C. They enter a rapid growth period from about April 13 to April 27, and then their height increase slows as they near maturity. Plant 1 grows from around 20 cm to one 174 cm. Plant 2 reaches 173 cm. Plant 3 rises to173 cm with slightly more scatter in its mid-season measurements. Plant 4 attains 173 and Plant 5 peaks at 172 cm. These final heights are broadly comparable to those in Rows A and B, with Row C showing marginally higher maxima and similar timing of the inflection point. Canopy span in Row C also expands more gradually. During the first two weeks it increases from approximately 30 cm to around 53 cm. After April 25 the rate of lateral expansion tapers off. Plant 1 finishes with a span of about 66 cm s. Plant 2 ends at 65 cm, Plant 3 at 70 cm. Plant 4's canopy reaches 64 cm and Plant 5 at 70 cm. These span values fall within the range observed in the other two rows. Overall, Row C's growth curves closely mirror the rhythm and amplitude of Rows A and B, confirming a consistent growth pattern across all fifteen plants with only minor variation in final size. Across all fifteen plants in Rows A, B, and C, the growth curves reveal a consistent pattern of rapid vertical growth from mid-April to late April, followed by a slowdown in height increase and gradual canopy expansion with only minor variations in timing and final size.

doi:10.6342/NTU202503732

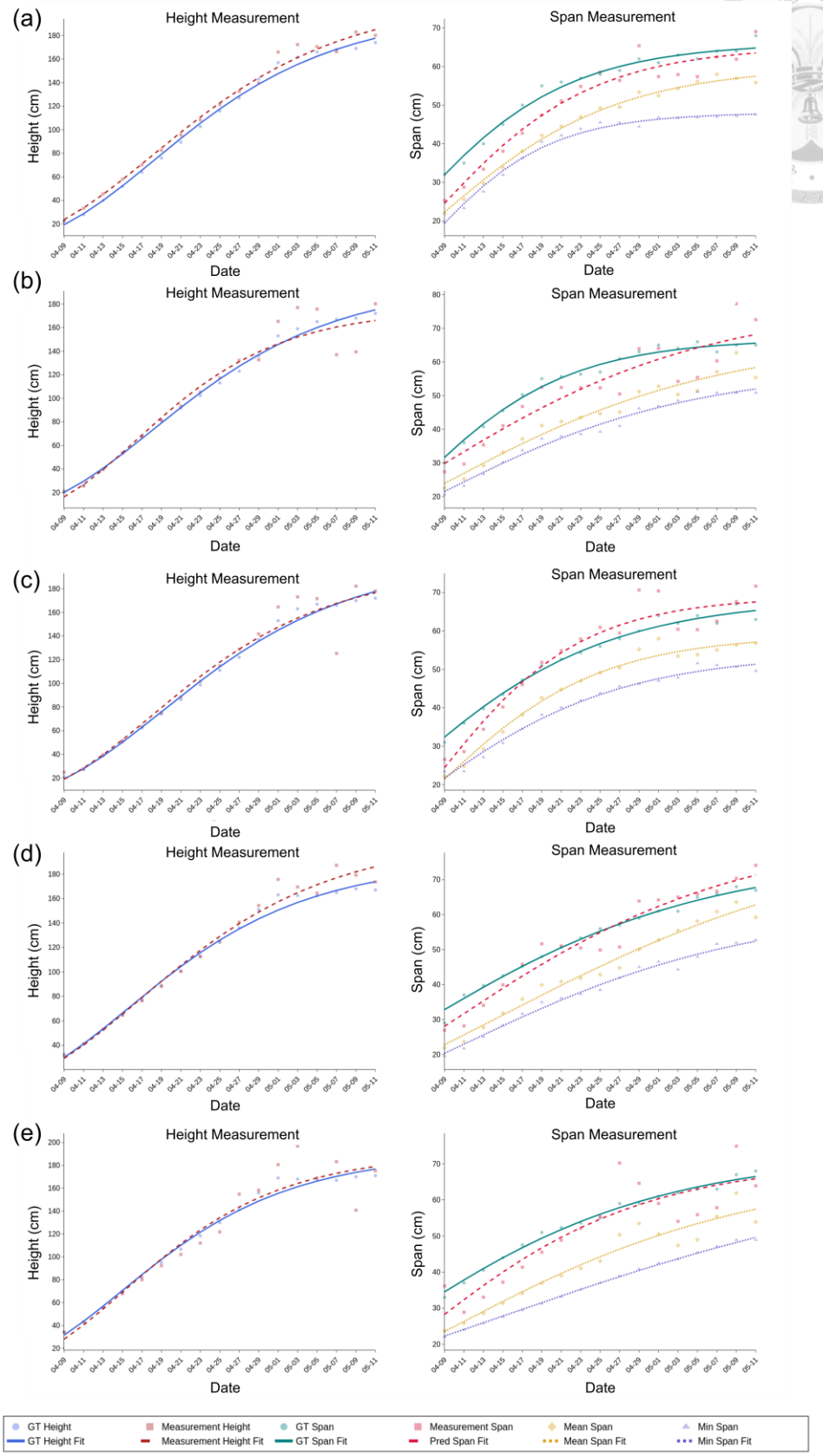


Fig. 4-21. Growth curves of height and span for Row A plants. Panels (a) to (e) correspond respectively to Plant A1 through Plant A5

doi:10.6342/NTU202503732

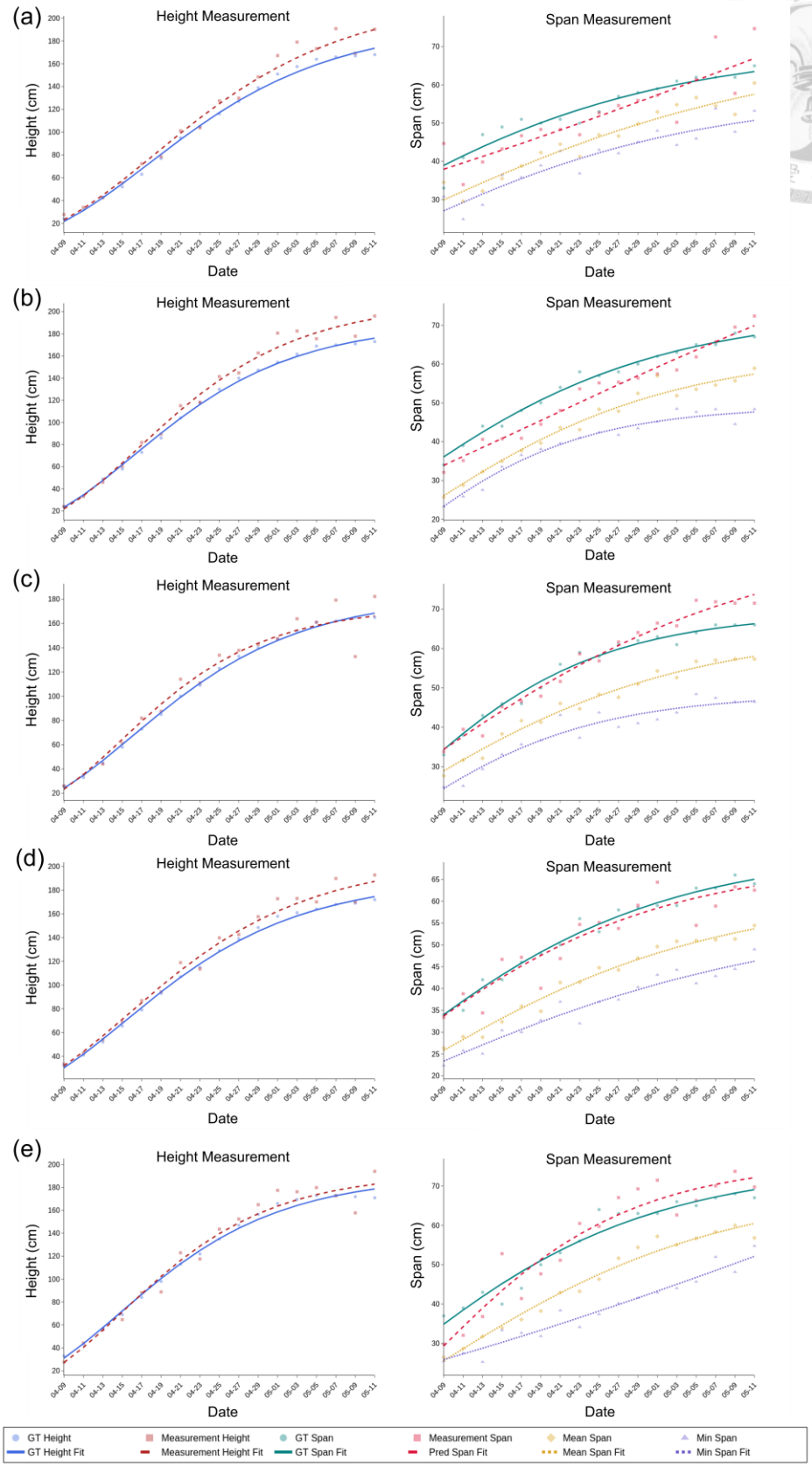


Fig. 4-22. Growth curves of height and span for Row B plants. Panels (a) to (e) correspond respectively to Plant B1 through Plant B5

doi:10.6342/NTU202503732

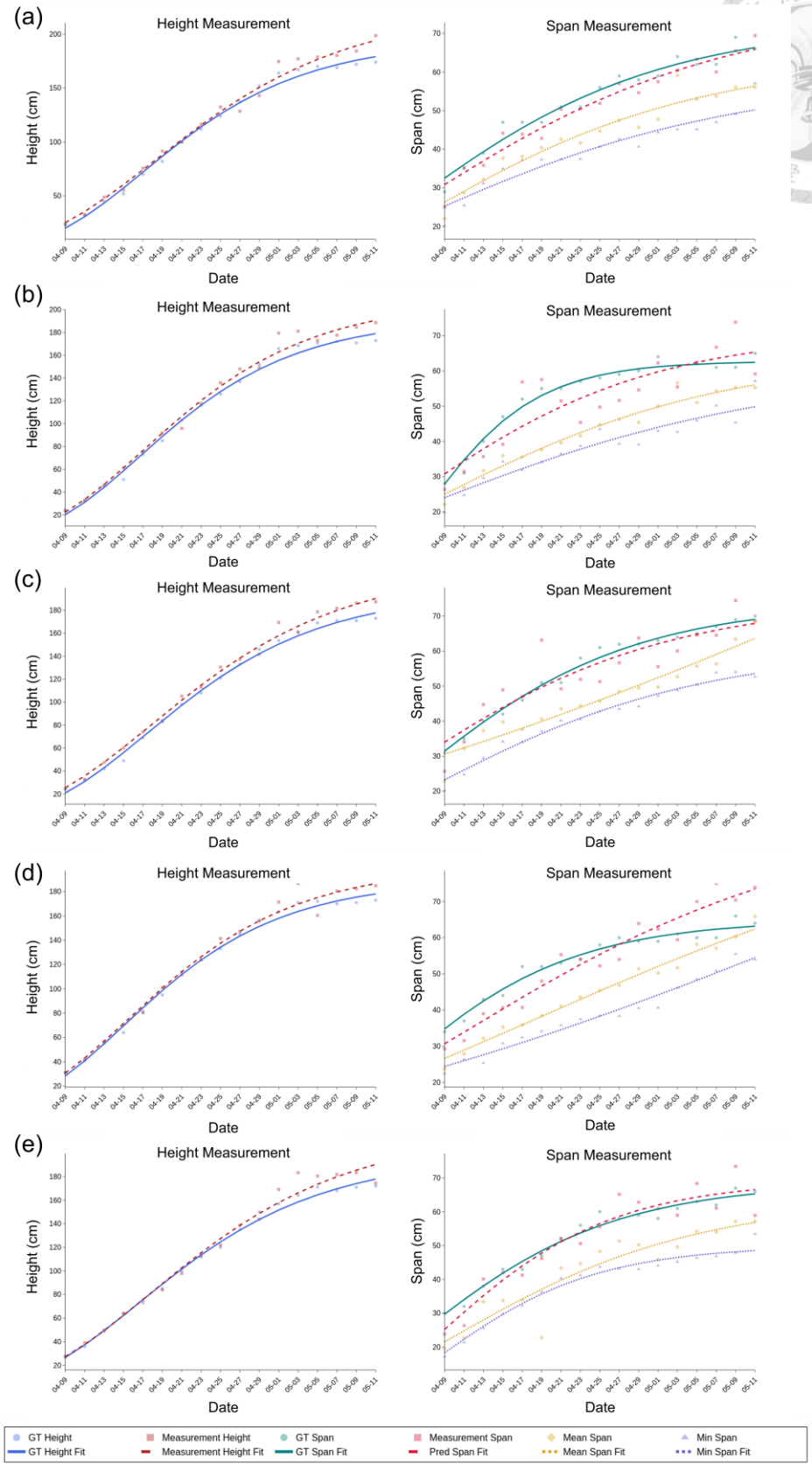


Fig. 4-23. Growth curves of height and span for Row C plants. Panels (a) to (e) correspond respectively to Plant C1 through Plant C5

doi:10.6342/NTU202503732

Fig. 4-24 presents the average height curve for all fifteen plants. The plants enter a steady growth phase that accelerates in mid-April before tapering off by early May. The measured height curve remains consistently above the ground truth curve and the gap widens slightly over time. Fig. 4-24 shows the average span rising from about 30 cm to 58 cm by late April and then moving into a slower saturation phase. The measured span curve lies above the ground truth and the difference between the mean and minimum span curves illustrates the variation among individual plants. Overall, the fifteen plants exhibit a uniform growth rhythm with a predictable bias in both height and span measurements.

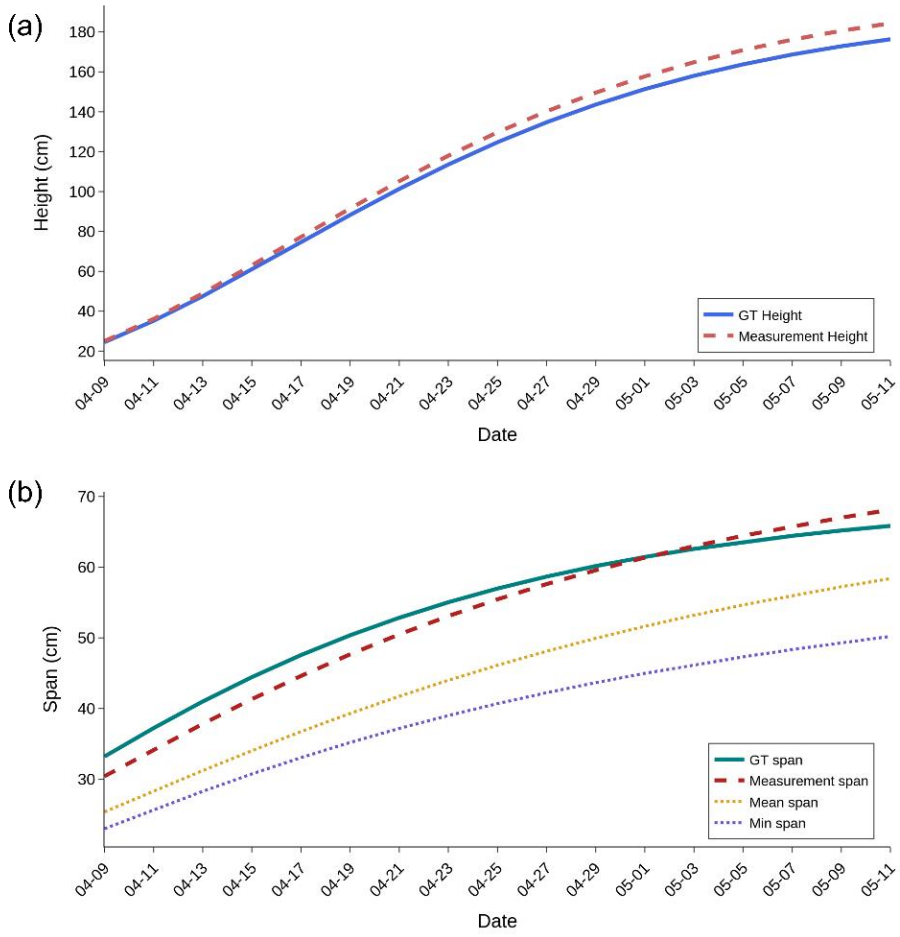


Fig. 4-24. Average growth curves for all fifteen plants: (a) Average height over time shown for ground truth and measured values, (b) Average canopy span over time showing ground truth, measured span, mean span, and minimum span curves

doi:10.6342/NTU202503732

CHAPTER 5
Conclusions and Suggestions

5.1 Conclusions

This research developed the first autonomous multi-UAV cooperative navigation system for greenhouse crop monitoring. Compared with conventional single-UAV solutions, this system demonstrates significant advancements in navigation accuracy, mission efficiency, and 3D phenotyping capability, establishing a scalable framework for precision agriculture in GPS-denied environments.

Multi-UAV autonomous navigation system

1. Developed a stable multi-UAV system with reliable communication and autonomous navigation using visual SLAM.
2. Proposed a loop-optimized mapping strategy that reduced mean absolute localization error by 39% compared to single-loop mapping, establishing a best practice for high-precision map generation.
3. Designed and evaluated three cooperative flight paths with UWB-based ground truth, providing quantitative guidelines for balancing accuracy and efficiency in collaborative missions. The parallel-aisle path shows an error range of 7 to 12 cm, the closed-loop path 5 to 9 cm, and the multi-altitude path 4 to 11 cm.
4. Multi-UAV deployment outperforms single UAV, reducing mission time by 73% and battery usage by 5% while maintaining localization accuracy across all flight paths.

doi:10.6342/NTU202503732

3D reconstruction of plant

1. Established a multi-view UAV imaging pipeline with Gaussian Splatting-based 3D reconstruction, achieving high-fidelity plant morphology models with PSNR up to 0.37, SSIM of 9.48, and LPIPS of 0.65.
2. Single-Side offers stable results, Merged captures both sides, and Three-Height performs best for tall plants.
3. Three-Height method achieves highest quality with PSNR 0.37, SSIM 9.48, and LPIPS 0.65.

Comparison of Plant Height and Canopy Span Across Three Reconstruction Methods

1. Estimated traits using point clouds from three reconstruction methods.
2. Single-Side method achieved the lowest height RMSE of 9.9 cm and the lowest span RMSE of 9.2 cm, demonstrating the most consistent measurement accuracy overall.

Per-Plant Measurements of Height and Canopy Span

1. Monitor growth for each plant and compare predicted and actual growth curves.
2. Calculate the R^2 and RMSE for 15 plants. Monitor growth for each plant and compare predicted and actual growth curves.

Plant height and canopy growth monitoring

1. Successfully monitored muskmelon growth dynamics using UAV-based 3D reconstruction

doi:10.6342/NTU202503732

2. Height increased rapidly in mid-April and stabilized in May, showing reliable growth measurement.
3. Canopy span grew slower, plateaued after late April, and showed higher variability.

5.2 Suggestions

1. Future work should integrate visual-inertial SLAM and deep-learning-based feature tracking to improve robustness under challenging lighting and dense foliage conditions.
2. In designing autonomous navigation for multi-UAV systems, future improvements could involve capturing images from closer distances and adopting more flexible imaging strategies. For instance, flight paths could be specifically planned to focus on capturing images of individual plants.
3. In terms of reconstruction, merged reconstruction tends to have a higher failure rate, particularly due to greater alignment difficulties. Future work could focus on increasing image overlap and experimenting with parameter adjustments in the structure from motion process to improve reconstruction success.

doi:10.6342/NTU202503732

References

Ampatzidis, Y., & Partel, V. (2019). UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sensing*, *11*(4). https://doi.org/10.3390/rs11040410

Asaamoning, G., Mendes, P., Rosário, D., & Cerqueira, E. (2021). Drone swarms as networked control systems by integration of networking and computing. *Sensors*, *21*(8). https://doi.org/10.3390/s21082642

Aslan, M. F., Durdu, A., Sabanci, K., Ropelewska, E., & Gültekin, S. S. (2022). A comprehensive survey of the recent studies with UAV for precision agriculture in open fields and greenhouses. *Applied Sciences*, *12*(3), 1047. https://www.mdpi.com/2076-3417/12/3/1047

Azzam, R., Boiko, I., & Zweiri, Y. (2023). Swarm cooperative navigation using centralized training and decentralized execution. *Drones*, *7*(3). https://doi.org/10.3390/drones7030193

Bagagiolo, G., Matranga, G., Cavallo, E., & Pampuro, N. (2022). Greenhouse Robots: ultimate solutions to improve automation in protected cropping systems—A review. *Sustainability*, *14*(11). https://doi.org/10.3390/su14116436

Bailey, T., Nieto, J., Guivant, J., Stevens, M., & Nebot, E. (2006, 9-15 Oct. 2006). Consistency of the EKF-SLAM algorithm. 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, https://doi.org/10.1109/IROS.2006.281644

Bouabdallah, S., Noth, A., & Siegwart, R. (2004, 28 Sept.-2 Oct. 2004). PID vs LQ control techniques applied to an indoor micro quadrotor. 2004 IEEE/RSJ

doi:10.6342/NTU202503732

International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566), https://doi.org/10.1109/IROS.2004.1389776

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., & Tardós, J. D. (2021). ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM. *IEEE Transactions on Robotics*, *37*(6), 1874-1890. https://doi.org/10.1109/TRO.2021.3075644

Chandra, A. L., Vikas Desai, S., Guo, W., & Balasubramanian, V. N. (2020). Computer vision with deep learning for plant phenotyping in agriculture: A survey. *arXiv e-prints*, arXiv:2006.11391. https://doi.org/10.48550/arXiv.2006.11391

Chang, Y., Cheng, Y., Manzoor, U., & Murray, J. (2023). A review of UAV autonomous navigation in GPS-denied environments. *Robotics and Autonomous Systems*, *170*, 104533. https://doi.org/10.1016/j.robot.2023.104533

Chen, W., Zhu, J., Liu, J., & Guo, H. (2024). A fast coordination approach for large-scale drone swarm. *Journal of Network and Computer Applications*, *221*, 103769. https://doi.org/10.1016/j.jnca.2023.103769

Chung, S. J., Paranjape, A. A., Dames, P., Shen, S., & Kumar, V. (2018). A Survey on aerial swarm robotics. *IEEE Transactions on Robotics*, *34*(4), 837-855. https://doi.org/10.1109/TRO.2018.2857475

Davison, A. J., Reid, I. D., Molton, N. D., & Stasse, O. (2007). MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(6), 1052-1067. https://doi.org/10.1109/TPAMI.2007.1049

Engel, J., Schöps, T., & Cremers, D. (2014). LSD-SLAM: Large-scale direct monocular SLAM.*Lecture Notes in Computer Science* ECCV 2014, Cham. https://doi.org/10.1007/978-3-319-10605-2_54

doi:10.6342/NTU202503732

Fiorani, F., & Schurr, U. (2013). Future scenarios for plant phenotyping. *Annual review of plant biology*, *64*. https://doi.org/10.1146/annurev-arplant-050312-120137

Gao, T., Sheng, W., Zhang, Z., Li, H., & Zhang, M. (2024). Greenhouse phenotyping measurement techniques and systems: A review. In Y. Yun, W. Sheng, & Z. Zhang (Eds.), *Advanced Sensing and Robotics Technologies in Smart Agriculture* (pp. 43-59). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-6441-9_3

Gompertz, B. (1825). XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. *Philosophical transactions of the Royal Society of London*(115), 513-583. https://doi.org/10.1098/rspl.1815.0271

Gu, Y., Jin, X., Xiang, R., Wang, Q., Wang, C., & Yang, S. (2020). UAV-based integrated multispectral-LiDAR imaging system and data processing. *Science China Technological Sciences*, *63*(7), 1293-1301. https://doi.org/10.1007/s11431-019-1571-0

Guédon, A., & Lepetit, V. (2023). SuGaR: Surface-aligned gaussian splatting for efficient 3D mesh reconstruction and high-quality mesh rendering. *arXiv e-prints*, arXiv:2311.12775. https://doi.org/10.48550/arXiv.2311.12775

Hu, J., Bruno, A., Zagieboylo, D., Zhao, M., Ritchken, B., Jackson, B., Chae, J. Y., Mertil, F., Espinosa, M., & Delimitrou, C. (2018). To centralize or not to centralize: A tale of swarm coordination. *arXiv e-prints*, arXiv:1805.01786. https://doi.org/10.48550/arXiv.1805.01786

Hunt Jr, E. R., & Daughtry, C. S. T. (2018). What good are unmanned aircraft systems for agricultural remote sensing and precision agriculture? *International Journal*

doi:10.6342/NTU202503732

of Remote Sensing, *39*(15-16), 5345-5376. https://doi.org/10.1080/01431161.2017.1410300

Huynh, H. X., Tran, L. N., & Duong-Trung, N. (2023). Smart greenhouse construction and irrigation control system for optimal Brassica Juncea development. *PLOS ONE*, *18*(10), e0292971. https://doi.org/10.1371/journal.pone.0292971

James, C., Chandra, S. S., & Chapman, S. C. (2025). A scalable and efficient UAV-based pipeline and deep learning framework for phenotyping sorghum panicle morphology from point clouds. *Plant Phenomics*, 100050. https://doi.org/10.1016/j.plaphe.2025.100050

Jamshidpey, A., Wahby, M., Allwright, M., Zhu, W., Dorigo, M., & Heinrich, M. K. (2024). Centralization vs. decentralization in multi-robot sweep coverage with ground robots and UAVs. *arXiv e-prints*, arXiv:2408.06553. https://doi.org/10.48550/arXiv.2408.06553

Ju, C., & Son, H. I. (2018). Multiple UAV systems for agricultural applications: Control, implementation, and evaluation. *Electronics*, *7*(9), 162. https://www.mdpi.com/2079-9292/7/9/162

Khosiawan, Y., & Nielsen, I. (2016). A system of UAV application in indoor environment. *Production & Manufacturing Research*, *4*(1), 2-22. https://doi.org/10.1080/21693277.2016.1195304

Kim, J., Kim, S., Ju, C., & Son, H. I. (2019). Unmanned Aerial Vehicles in agriculture: A review of perspective of platform, control, and applications. *IEEE Access*, *7*, 105100-105115. https://doi.org/10.1109/ACCESS.2019.2932119

Krul, S., Pantos, C., Frangulea, M., & Valente, J. (2021). Visual SLAM for indoor livestock and farming using a small drone with a monocular camera: A feasibility study. *Drones*, *5*(2). https://doi.org/10.3390/drones5020041

doi:10.6342/NTU202503732

Li, D., Shi, G., Li, J., Chen, Y., Zhang, S., Xiang, S., & Jin, S. (2022). PlantNet: A dual-function point cloud segmentation network for multiple plant species. *ISPRS Journal of Photogrammetry and Remote Sensing*, *184*, 243-263. https://doi.org/10.1016/j.isprsjprs.2022.01.007

Li, R., Wang, S., & Gu, D. (2018). Ongoing evolution of Visual SLAM from geometry to deep learning: challenges and opportunities. *Cognitive Computation*, *10*(6), 875-889. https://doi.org/10.1007/s12559-018-9591-8

Li, Y., Liu, J., Zhang, B., Wang, Y., Yao, J., Zhang, X., Fan, B., Li, X., Hai, Y., & Fan, X. (2022). Three-dimensional reconstruction and phenotype measurement of maize seedlings based on multi-view image sequences. *Frontiers in plant science*, *13*, 974339. https://doi.org/10.3389/fpls.2022.974339

Loayza, K., Lucas, P., & Peláez, E. (2017, October 16-20). A centralized control of movements using a collision avoidance algorithm for a swarm of autonomous agents. 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), Salinas, Ecuador. https://doi.org/10.1109/ETCM.2017.8247496

Lu, Y., Zhucun, X., Gui-Song, X., & Zhang, L. (2018). A survey on vision-based UAV navigation. *Geo-spatial Information Science*, *21*(1), 21-32. https://doi.org/10.1080/10095020.2017.1420509

Mur-Artal, R., Montiel, J. M. M., & Tardós, J. D. (2015). ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, *31*(5), 1147-1163. https://doi.org/10.1109/TRO.2015.2463671

Mur-Artal, R., & Tardós, J. D. (2017). ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, *33*(5), 1255-1262. https://doi.org/10.1109/TRO.2017.2705103

doi:10.6342/NTU202503732

Pan, L., Baráth, D., Pollefeys, M., & Schönberger, J. L. (2024). Global structure-from-motion revisited. *arXiv e-prints*, arXiv:2407.20219. https://doi.org/10.48550/arXiv.2407.20219

Pantos, C., Hildmann, H., & Valente, J. (2023). Experimental connectivity analysis for drones in greenhouses. *Drones*, *7*(1). https://doi.org/10.3390/drones7010024

Pardossi, A., Paola, G., Fernando, M., Marinone, A. F., Carla, M., Giovanni, S., & Vernieri, P. (2000). The influence of growing season on fruit yield and quality of greenhouse melon (Cucumis melo L.) grown in nutrient film technique in a Mediterranean climate. *The Journal of Horticultural Science and Biotechnology*, *75*(4), 488-493. https://doi.org/10.1080/14620316.2000.11511274

Park, T., Lee, J., Oh, H., Yun, W.-J., & Lee, K.-W. (2025). Optimizing indoor farm monitoring efficiency using UAV: yield estimation in a GNSS-denied cherry tomato greenhouse. *arXiv preprint arXiv:2505.00995*. https://doi.org/10.48550/arXiv.2505.00995

Pieruschka, R., & Schurr, U. (2019). Plant phenotyping: past, present, and future. *Plant Phenomics*, *2019*, 7507131. https://doi.org/10.34133/2019/7507131

Qi, C. R., Liu, W., Wu, C., Su, H., & Guibas, L. J. (2017). Frustum PointNets for 3D object detection from RGB-D data. *arXiv e-prints*, arXiv:1711.08488. https://doi.org/10.48550/arXiv.1711.08488

Rejeb, A., Abdollahi, A., Rejeb, K., & Treiblmaier, H. (2022). Drones in agriculture: A review and bibliometric analysis. *Computers and Electronics in Agriculture*, *198*, 107017. https://doi.org/10.1016/j.compag.2022.107017

Roldán, J. J., Joossen, G., Sanz, D., Del Cerro, J., & Barrientos, A. (2015). Mini-UAV based sensory system for measuring environmental variables in greenhouses. *Sensors*, *15*(2), 3334-3350. https://doi.org/10.3390/s150203334

doi:10.6342/NTU202503732

Schlegel, D., Colosi, M., & Grisetti, G. (2018, 21-25 May 2018). ProSLAM: Graph SLAM from a Programmer's Perspective. 2018 IEEE International Conference on Robotics and Automation (ICRA), https://doi.org/10.1109/ICRA.2018.8461180

Schönberger, J. L., & Frahm, J. M. (2016, 27-30 June 2016). Structure-from-Motion revisited. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), https://doi.org/10.1109/CVPR.2016.445

Shi, W., van de Zedde, R., Jiang, H., & Kootstra, G. (2019). Plant-part segmentation using deep learning and multi-view vision. *Biosystems Engineering*, *187*, 81-95. https://doi.org/10.1016/j.biosystemseng.2019.08.014

Singh, R., & Singh, G. S. (2017). Traditional agriculture: a climate-smart approach for sustainable food production. *Energy, Ecology and Environment*, *2*(5), 296-316. https://doi.org/10.1007/s40974-017-0074-7

Sumikura, S., Shibuya, M., & Sakurada, K. (2019). *OpenVSLAM: A versatile visual SLAM framework* proceedings of the 27th ACM International Conference on Multimedia, Nice, France. https://doi.org/10.1145/3343031.3350539

Teshome, F. T., Bayabil, H. K., Hoogenboom, G., Schaffer, B., Singh, A., & Ampatzidis, Y. (2023). Unmanned aerial vehicle (UAV) imaging and machine learning applications for plant phenotyping. *Computers and Electronics in Agriculture*, *212*, 108064. https://doi.org/10.1016/j.compag.2023.108064

Trujillo, J.-C., Munguia, R., Guerra, E., & Grau, A. (2018). Visual-based SLAM configurations for cooperative multi-UAV systems with a lead agent: An observability-based approach. *Sensors*, *18*(12). https://doi.org/10.3390/s18124243

doi:10.6342/NTU202503732

Vetrella, A. R., Fasano, G., Accardo, D., & Moccia, A. (2016). Differential GNSS and vision-based tracking to improve navigation performance in cooperative multi-UAV systems. *Sensors*, *16*(12). https://doi.org/10.3390/s16122164

Wang, Q., Tan, Y., & Mei, Z. (2020). Computational methods of acquisition and processing of 3D point cloud data for construction applications. *Archives of Computational Methods in Engineering*, *27*(2), 479-499. https://doi.org/10.1007/s11831-019-09320-4

Williams, D., Macfarlane, F., & Britten, A. (2024). Leaf only SAM: A segment anything pipeline for zero-shot automated leaf segmentation. *Smart Agricultural Technology*, *8*, 100515. https://doi.org/10.1016/j.atech.2024.100515

Zafari, F., Gkelias, A., & Leung, K. K. (2019). A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, *21*(3), 2568-2599. https://doi.org/10.1109/COMST.2019.2911558

Zarei, A., Li, B., Schnable, J. C., Lyons, E., Pauli, D., Barnard, K., & Benes, B. (2024). PlantSegNet: 3D point cloud instance segmentation of nearby plant organs with identical semantics. *Computers and Electronics in Agriculture*, *221*, 108922. https://doi.org/10.1016/j.compag.2024.108922

Zhai, Z., Martínez, J. F., Beltran, V., & Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, *170*, 105256. https://doi.org/10.1016/j.compag.2020.105256

Zhang, C., Tang, C., Wang, H., Lian, B., & Zhang, L. (2025). Data set for UWB cooperative navigation and positioning of UAV cluster. *Scientific Data*, *12*(1), 486. https://doi.org/10.1038/s41597-025-04808-0

doi:10.6342/NTU202503732

Zhang, H., Ma, H., Mersha, B. W., Zhang, X., & Jin, Y. (2023). Distributed cooperative search method for multi-UAV with unstable communications. *Applied Soft Computing*, *148*, 110592. https://doi.org/10.1016/j.asoc.2023.110592

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *arXiv e-prints*, arXiv:1801.03924. https://doi.org/10.48550/arXiv.1801.03924

Zhang, X., Bai, Y., & He, K. (2023). On countermeasures against cooperative fly of UAV swarms. *Drones*, *7*(3). https://doi.org/10.3390/drones7030172

Zhao, J., Bodner, G., & Rewald, B. (2016). Phenotyping: Using machine learning for improved pairwise genotype classification based on root traits [Original Research]. *Frontiers in plant science*, *Volume 7 - 2016*. https://doi.org/10.3389/fpls.2016.01864

doi:10.6342/NTU202503732