Smart Agricultural Technology 14 (2026) 102024

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Neural 3D reconstruction and immersive VR visualization of row crops across phenological growth stages

Shambhavi Joshi, Juan Di Salvo, Yanben Shen, Mozhgan Hadadi, Venkata Naresh Boddepalli, Talukder Z. Jubery, Soumik Sarkar, Arti Singh, Baskar Ganapathysubramanian, Asheesh K. Singh, Adarsh Krishnamurthy *

Iowa State University, Ames, 50011, Iowa, USA

ARTICLE INFO

Keywords:
BBCH growth stages
Neural radiance fields (NeRF)
3D Gaussian splatting
3D visualization
Virtual reality (VR)

ABSTRACT

Plant phenotyping in precision agriculture increasingly requires high-fidelity three-dimensional reconstruction and accessible visualization methods. This study presents an integrated pipeline combining Neural Radiance Fields (NeRF), 3D Gaussian Splatting (G-Splat), and Virtual Reality (VR) visualization for comprehensive plant analysis across developmental stages. We collected multi-view imagery of finger millet, proso millet, mungbean, and field pea under controlled greenhouse conditions, aligning data acquisition with standardized BBCH phenological scales. Camera pose estimation was performed using GLOMAP, followed by reconstruction via both Nerfacto and G-Splat implementations. Quantitative evaluation using PSNR, SSIM, and LPIPS metrics revealed complementary strengths of the two approaches: G-Splat achieved superior structural fidelity, while NeRF provided enhanced perceptual realism. Both reconstruction methods were successfully integrated into an immersive VR greenhouse environment deployed on Meta Quest headsets, maintaining consistently high framerates. This framework establishes a practical foundation for incorporating neural reconstruction and immersive technologies into agricultural phenotyping workflows, supporting both research applications and educational engagement.

1. Introduction

Three-dimensional (3D) reconstruction of plants is increasingly critical in precision agriculture, enabling detailed analyses of plant morphology, growth, productivity, and stress responses that exceed the capabilities of conventional 2D imagery [1]. Accurate reconstruction of plant structures is essential for extracting precise phenotypic traits, which directly influence breeding decisions and stress evaluation [2]. As the demand for data-driven approaches in crop management continues to grow, robust and accessible 3D reconstruction methods have become fundamental tools for advancing both research and education in agricultural sciences. Traditional reconstruction techniques, such as photogrammetry and Light Detection and Ranging (LiDAR), have served as the primary methods for generating 3D plant models. Photogrammetry can produce high-resolution models but suffers from significant limitations: labor-intensive workflows, sensitivity to environmental variations, and difficulty reconstructing fine plant details under complex occlusions or low-texture scenarios typical of agricultural environments [2,3]. Similarly, while LiDAR provides precise spatial data, it remains costly, often yields limited resolution for thin plant structures such as leaves and stems, and poses deployment challenges in dynamic field settings [4,5]. These limitations have motivated the exploration of alternative approaches capable of capturing complex plant geometries with greater efficiency and fidelity.

Recent advancements in implicit volumetric modeling have introduced transformative capabilities for 3D reconstruction. Neural Radiance Fields (NeRF) encode scenes into continuous neural representations by training networks to predict color and density from spatial coordinates and viewing angles across multiple RGB images [6]. More recently, 3D Gaussian Splatting (G-Splat) has emerged as a complementary approach, rendering radiance fields using view-dependent 3D Gaussians to enable real-time performance while preserving fine structural detail [7]. NeRFs specifically have demonstrated enhanced reconstruction fidelity for capturing complex plant geometries under natural imaging conditions [8]. Several NeRF variants, including Mip-NeRF [9], Instant-NGP [10], and Nerfacto [11], have further improved reconstruction accuracy, computational speed, and robustness to variable imaging conditions, making them increasingly viable for agricultural applications.

Despite these advances, significant gaps remain in applying neural reconstruction methods to agricultural research. First, existing work

* Corresponding author.
E-mail address: adarsh@iastate.edu (A. Krishnamurthy).

https://doi.org/10.1016/j.atech.2026.102024
Received 19 January 2026; Received in revised form 17 March 2026; Accepted 18 March 2026
Available online 20 March 2026

lacks standardized temporal alignment with biologically meaningful growth stages, limiting the ability to make consistent developmental comparisons. Second, rigorous quantitative validation across diverse plant architectures and developmental stages remains limited, with few studies systematically comparing reconstruction quality using established image quality metrics. Third, while high-fidelity 3D models provide valuable data, their utility for education and collaborative research is constrained without accessible, interactive visualization platforms. Immersive Virtual Reality (VR) environments offer a promising solution by enabling researchers, educators, and students to explore plant structures at natural scale, manipulate viewing perspectives, and compare phenotypes across growth stages in an intuitive spatial context. However, integrating neural reconstructions with deployable VR systems presents substantial technical challenges, including coordinate system transformations, rendering optimization, and packaging for standalone devices.

Standardized phenological scales provide essential frameworks for addressing the temporal alignment challenge. The BBCH scale (Biologische Bundesanstalt, Bundessortenamt, and Chemical industry) is an internationally recognized system for coding phenological growth stages across diverse crop species [12]. The BBCH scale employs a decimal code from 0 to 99, describing successive developmental stages from germination through senescence based on visible morphological characteristics. This standardization enables consistent communication among researchers, facilitates comparative studies across experiments and locations, and provides a common reference for agronomic decision-making. Aligning 3D reconstruction datasets with BBCH stages establishes biologically meaningful temporal reference points, allowing quantitative assessment of morphological changes and trait expression patterns throughout plant development. Such alignment is particularly valuable for phenotyping studies where developmental stage directly influences trait measurements and their interpretation.

To address these gaps, we have developed a pipeline for capturing plants at different growth stages, reconstructing the 3D geometry, and visualizing them in an immersive virtual environment. The primary contributions are:

- A streamlined reconstruction workflow combining pose estimation with NeRF (Nerfacto) and 3D Gaussian Splatting, validated across multiple crop species (finger millet, proso millet, mungbean, and field pea) at key growth stages.
- BBCH-aligned data collection protocol providing biologically meaningful temporal reference points that enable consistent developmental analysis and phenotypic comparison.
- Integration of reconstructed models into Unreal Engine with precise camera-matched rendering, enabling quantitative evaluation through direct comparison of rendered and real images using image metrics.
- An immersive VR greenhouse environment deployed on standalone headsets (Meta Quest series), demonstrating practical integration of neural reconstructions for real-time interactive exploration in research and educational settings.

This framework establishes a practical and scalable foundation for incorporating neural 3D reconstruction and immersive visualization into agricultural phenotyping workflows. This work provides insights into how different neural rendering approaches balance structural detail, perceptual realism, and rendering efficiency for plant phenotyping tasks by systematically comparing NeRF and 3D Gaussian Splatting within a unified reconstruction and visualization pipeline. The framework further integrates BBCH-aligned data collection with an Unreal Engine-based camera-matched comparison workflow and immersive VR visualization, enabling both quantitative evaluation and intuitive exploration of plant development across growth stages. We provide guidance on method selection based on application requirements by comparing the complementary strengths of NeRF and G-Splat: perceptual realism versus geometric precision and rendering efficiency. The VR deployment bridges the gap between high-fidelity modeling and accessible visualization, thereby supporting broader adoption of digital technologies in precision agriculture.

The remainder of this paper is organized as follows: Section 2 describes our plant growth protocols and data acquisition methods; Section 3 details the 3D reconstruction pipeline using NeRF and G-Splat; Section 4 presents the VR greenhouse implementation; Section 3.5 explains our quantitative evaluation methodology; Section 5 discusses results and performance analysis; and Section 6 concludes with future directions.

2. Data acquisition

2.1. Plant growth protocol

Data collection was carried out under controlled greenhouse conditions to ensure consistent environmental parameters and high-quality imaging across key growth stages. Multiple genotypes of millet, mung beans, and peas were selected for their structural diversity and suitability for evaluating 3D reconstruction methods across varying plant architectures.

2.1.1. Millet

We selected two distinct millet genotypes: Finger millet (Eleusine coracana) and Proso millet (Panicum miliaceum), due to their contrasting growth habits and significant agricultural relevance, making them ideal for validating reconstruction methods across varying plant structures. Finger millet, characterized by dense foliage and complex branching, and Proso millet, exhibiting more open and sparse morphology, collectively provided a comprehensive basis for us to evaluate reconstruction accuracy and robustness [13].

The seeds were uniformly sown at a depth of 2.5 cm in 4-gallon buckets filled with Sungro potting mix (S.S. #1 - F1P), which we chose specifically for its consistent performance in supporting robust root development. We standardized initial plant nutrition using Osmocote (14-14-14), providing a balanced delivery of essential nutrients (nitrogen, phosphorus, potassium) crucial for healthy and uniform plant growth. After seedling emergence, we performed systematic thinning to maintain exactly one plant per container, eliminating inter-plant competition effects. We provided supplemental nitrogen fertilization using 1g of urea (46-0-0) at three weeks post-emergence to encourage vigorous vegetative growth. We arranged the plants in a completely randomized design (CRD), employing five replicates per millet genotype. Additionally, we randomized container locations bi-weekly within the greenhouse to mitigate positional biases and ensure even exposure to micro-environmental conditions. We strictly monitored and regulated environmental parameters, maintaining a consistent temperature between 70–75°F (approximately 21–24°C) and relative humidity between 60–65%. We documented phenological growth stages three times per week according to the internationally recognized BBCH (Biologische Bundesanstalt, Bundessortenamt, und Chemische Industrie) scale [12], ensuring standardized and comprehensive tracking of plant developmental progress.

2.1.2. Mung beans

For mung beans, seeds were planted in mid-December 2024, with three seeds per bucket at a depth of 1 inch, and later thinned to one plant per bucket. Sungro (S.S. #1 - F1P) potting mix was used, and Osmocote (14-14-14) was fertilized every 10 days, beginning 15 days after planting. This experiment utilized three replications per genotype, arranged randomly in a Completely Randomized Design. Greenhouse conditions were held at 70–75°F and 60–65% relative humidity.

2.1.3. Peas

Field peas, including the genotypes CDC Amarillo and Delta, were planted simultaneously with mung beans, with three seeds per bucket at

(a) (b) (c)

Fig. 1. (a) RAUBAY 360° spinner turntable for 360° plant imaging; (b) Mid-height view for capturing intermediate growth stages; (c) Overhead view for documenting taller plant structures.

a depth of 2inches. CDC Amarillo, developed by the Crop Development Center (CDC) at the University of Saskatchewan, is a yellow cotyledon field pea cultivar known for its good lodging resistance, medium-sized round seeds, and high yield potential [14]. Delta is a semi-leafless yellow field pea variety recognized for its shorter vine length and improved standability at harvest [15]. After emergence, plants were thinned to one per bucket and grew in Sungro (S.S. #1 - F1P) potting mix. Fertilization followed the same schedule as that of mung beans, with Osmocote (14-14-14) applied every 10 days, starting 15 days after planting. Each genotype had three replications in a Completely Randomized Design, with buckets positioned randomly. The greenhouse was maintained at 68°F and 60–65% relative humidity.

2.2. Plant imaging

We employed two complementary methods for data collection for accurate 3D plant reconstruction: Polycam-based photogrammetry and traditional camera video recording.

Polycam Photogrammetry: Polycam, an iPhone-based photogrammetry application, was used to capture 25–50 high-resolution images per plant using an iPhone 14 Pro equipped with a LiDAR sensor [16]. During acquisition, the device was moved around each plant at varying heights and distances to sample diverse viewpoints. Polycam provides real-time pose estimation using LiDAR-assisted sensor fusion, enabling immediate feedback on viewpoint distribution and helping identify coverage gaps before leaving the greenhouse. Although Polycam's internal pose estimation is independent of external markers, the consistent imaging setup across plants allowed a meaningful comparison with our Structure-from-Motion workflow.

Traditional Camera Video Capture: To complement this, we employed a traditional camera to capture high-resolution video footage at 4K and 60 frames per second for each plant. To maintain uniformity and ensure complete spatial coverage, we used a RAUBAY 360° spinner turntable (Fig. 1a), which facilitated systematic rotation for consistent video data acquisition. Videos were recorded from two distinct vertical viewpoints: mid-level (Fig. 1b) and overhead (Fig. 1c). This dual-height strategy ensured comprehensive coverage of structural complexity, particularly for taller millet plants at advanced growth stages.

Integrating these two complementary methods enabled us to leverage their respective strengths. Traditional video capture offers thorough and systematic coverage, while the photogrammetry of PolyCam provides quick and intuitive data collection features. This hybrid strategy provided robust datasets that formed the foundation for high-fidelity 3D plant reconstruction.

3. 3D reconstruction

Fig. 2 shows the different steps for our 3D reconstruction pipeline for plants. We will now explain the different steps in detail below.

3.1. Pose estimation and data processing

Following our data collection, individual video clips captured from multiple angles and heights were merged into continuous videos for each plant. These videos were then converted into image sequences using FFmpeg at 4 frames per second, extracting approximately 80–120 images per complete rotation, depending on plant size and turntable speed. The resulting image sets served as inputs for camera pose estimation. For each plant at each growth stage, we processed images from both capture methods: Polycam photogrammetry (25–50 images) provided initial pose estimates with LiDAR assistance, while the systematic video-derived frames (80–120 images) were processed through GLOMAP for high-precision pose estimation used in final reconstructions. This dual approach enabled us to leverage Polycam's rapid feedback during acquisition while ensuring comprehensive angular coverage through the turntable-based video capture.

We utilized GLOMAP, a globally optimized Structure-from-Motion (SfM) framework that builds upon COLMAP [17,18] and incorporates recent global optimization techniques [19] to enhance camera pose accuracy in low-texture environments. Compared to traditional COLMAP workflows, which required approximately 40–90 min per plant dataset on our workstation, GLOMAP completed pose estimation in under 10 min. These substantial reductions in processing time were accompanied by improved global alignment stability, yielding more accurate camera poses. In the context of plant phenotyping, where repetitive foliage and uniform containers often pose challenges for feature detection, these improvements made GLOMAP an ideal choice for efficiently generating camera poses required for high-fidelity NeRF and G-Splat reconstructions.

To enhance feature detection during GLOMAP processing, we applied colorful tape to the plant buckets (Fig. 3a). Each of the four top corners was marked with a distinct color: red, neon yellow, orange, and blue, chosen to avoid overlap with plant coloration and provide strong contrast against the bucket. For datasets such as peas and mungbeans, we additionally incorporated ArUco markers or patterned tape strips to further increase feature diversity. These visual anchors improved feature matching reliability, reduced pose-estimation failures, and strengthened alignment robustness in low-texture regions where foliage appeared repetitive from multiple viewpoints. While these markers enhanced reconstruction quality under controlled greenhouse conditions, we acknowledge this as a limitation for direct field deployment. Future work should validate whether comparable reconstruction quality can be achieved without such artificial markers, particularly for field phenotyping applications where such modifications are impractical.

This improved alignment is reflected in the resulting camera positions, visualized as red pyramids representing the calculated camera locations relative to the plant (Fig. 3b). Each pyramid corresponds to a precise viewpoint used during data capture, illustrating both the spatial

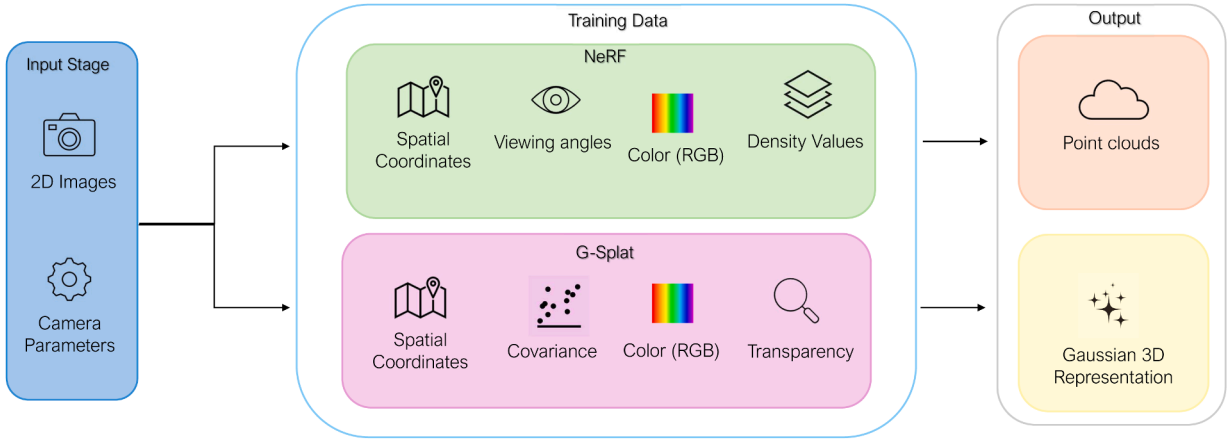


Fig. 2. NeRF/G-Splat reconstruction pipeline. Input images and camera parameters are processed using GLOMAP for estimating camera pose. NeRF models color and density for volumetric reconstruction, while G-Splat uses Gaussian covariance and transparency to generate efficient 3D representations.

(a) (b)

Fig. 3. (a) Plant bucket with colorful tape on corners to improve feature detection; (b) Visualization of GLOMAP-calculated camera poses (red pyramids) representing camera positions and orientations relative to the plant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distribution and orientation of the cameras. The dense and organized arrangement of these poses demonstrates comprehensive multi-angle coverage and reduced pose estimation errors. This high-quality alignment was crucial for downstream 3D reconstruction pipelines, enabling better structural accuracy and finer visual detail in complex plant geometries.

In parallel, we also utilized the Polycam application, which leverages LiDAR sensor data from the iPhone 14 Pro to estimate camera poses in real time during image acquisition. Although the colorful markers did not directly influence Polycam's internal pose estimation process, the same plants were used across both methods. This consistent setup enabled a fair comparison between the LiDAR-based and feature-based SfM pipelines. Polycam's immediate feedback during capture was valuable for identifying coverage gaps and guiding the selection of additional viewpoints. In contrast, GLOMAP provided more precise and globally consistent alignments, suitable for high-resolution NeRF and G-Splat reconstructions.

In addition to NeRF, the same image acquisition and pose estimation pipeline was used to generate G-Splat reconstructions. Although GLOMAP was not initially designed for splatting methods, its accurate pose output was directly compatible with both reconstruction pipelines, ensuring consistency across comparisons. To ensure a fair comparison between reconstruction methods, both NeRF and G-Splat were evaluated using the same image subsets per plant, providing identical viewpoints for quantitative analysis.

3.2. Neural radiance fields (NeRFs)

Neural Radiance Fields (NeRFs) have significantly advanced 3D reconstruction by enabling detailed volumetric modeling from collections of 2D images. NeRF models how light interacts at each spatial point (x, y, z) and a viewing or ray direction, generating continuous 3D representations without explicit geometry. This process involves training a neural network with spatial coordinates and camera poses obtained using Structure-from-Motion tools such as COLMAP [17,18] and predicting RGB colors and density values along a ray [6]. Recent studies underscore the extensive potential of NeRF techniques in agricultural phenomics, highlighting their effectiveness in capturing detailed plant morphology [20,21]. The NeRF rendering process utilizes multiple 2D images of the millet plant captured from various angles (left), with corresponding camera poses obtained using COLMAP or GLOMAP to ensure accurate alignment in 3D space. These inputs are processed through the NeRF model (center), where a neural network predicts RGB color values and density for each spatial point along a viewing ray (Fig. 4). The

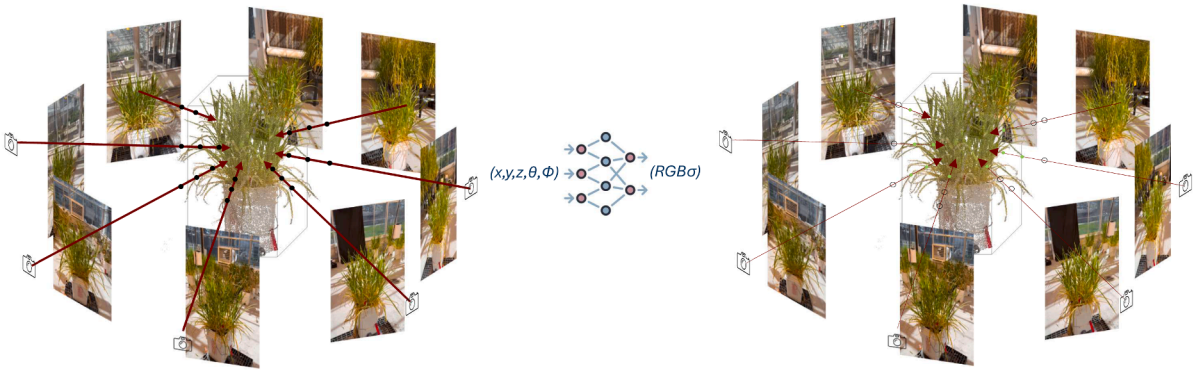


Fig. 4. NeRF 5D rendering process showing multiple 2D images of a millet plant captured from various angles, used to predict RGB color values and density for generating a detailed 3D reconstruction that accurately represents the structure of the plant.

(a) (b) (c)

Fig. 5. (a) Nerfstudio training viewer interface displaying the 3D plant reconstruction with point cloud visualization settings; (b) Real image of an open flower in proso millet; (c) Point cloud representation of the same flower.

Table 1
Training hyperparameters and computational requirements for NeRF (Nerfacto) reconstructions.

Parameter	Value
Training iterations	30,000
Learning rate (initial)	1×10^{-2}
Learning rate (final)	1×10^{-4}
LR scheduler	Exponential decay
Batch size (rays)	4096
Proposal network samples	256 + 96
Final network samples	48
GPU memory usage (training)	~3.6 GB
GPU utilization	50–90%
Training time per model	~20 min
Point cloud export time	~30 min (100M points)

reconstructed 3D model (right) demonstrates NeRF's ability to capture detailed plant structures.

In this study, we employed Nerfacto, a computationally efficient variant of NeRF optimized for general-purpose 3D reconstruction tasks [11]. Nerfacto employs multi-scale rendering and adaptive proposal sampling to optimize computational resource allocation. The system focuses on complex regions that require more detailed information while simplifying areas with less intricate details, resulting in more efficient rendering. This strategy substantially reduces processing time and computational load while preserving model accuracy and structural detail, key factors for precise plant 3D reconstructions [9,10].

Table 1 summarizes the training configuration and computational requirements for our Nerfacto reconstructions. Training was conducted on an NVIDIA GeForce RTX 3090 GPU (24 GB VRAM) paired with an Intel Core i9 processor and 128 GB system RAM. Using an exponential learning rate schedule that decayed from 1×10^{-2} to 1×10^{-4} over 30,000 iterations, each model trained for approximately 20 min, with GPU memory usage stabilizing around 3.6 GB. Although Nerfacto internally tracks performance metrics such as PSNR, we did not rely on convergence curves; instead, training was run for a fixed iteration count across all datasets to ensure consistent processing. Point cloud extraction required an additional ~30 min per plant due to the high-resolution exports (up to 100 million points).

We employed NeRFstudio, featuring an interactive viewer interface (Fig. 5a), for real-time monitoring during model training. Frameworks such as NeRFstudio streamline NeRF models, simplifying deployment and accelerating research for agricultural applications. We utilized the bounding-box cropping tools in NeRFstudio to precisely isolate plant structures, thereby eliminating unnecessary background noise and enhancing reconstruction quality. This proactive data refinement enhanced both processing efficiency and model fidelity. Post-processing steps included statistical outlier removal and normal computation via Open3D, which further improved point cloud accuracy and enhanced overall visual clarity. These steps were essential for refining the intricate morphological details of the plants.

Our NeRF pipeline successfully reconstructed detailed and accurate point clouds that captured various morphological traits, including intricate leaf structures and branching complexity in millet plants. However, we observed some limitations when reconstructing magnificent structures such as delicate flowers or thin foliage (Fig. 5c), primarily due to constraints inherent to image resolution and photogrammetric methods.

3.3. Post-processing and data cleanup

Following the initial generation of millet point clouds, we performed comprehensive data cleaning and post-processing to enhance model accuracy and ensure precise representation of plant morphology. Accurate and visually consistent point clouds are crucial for reliable

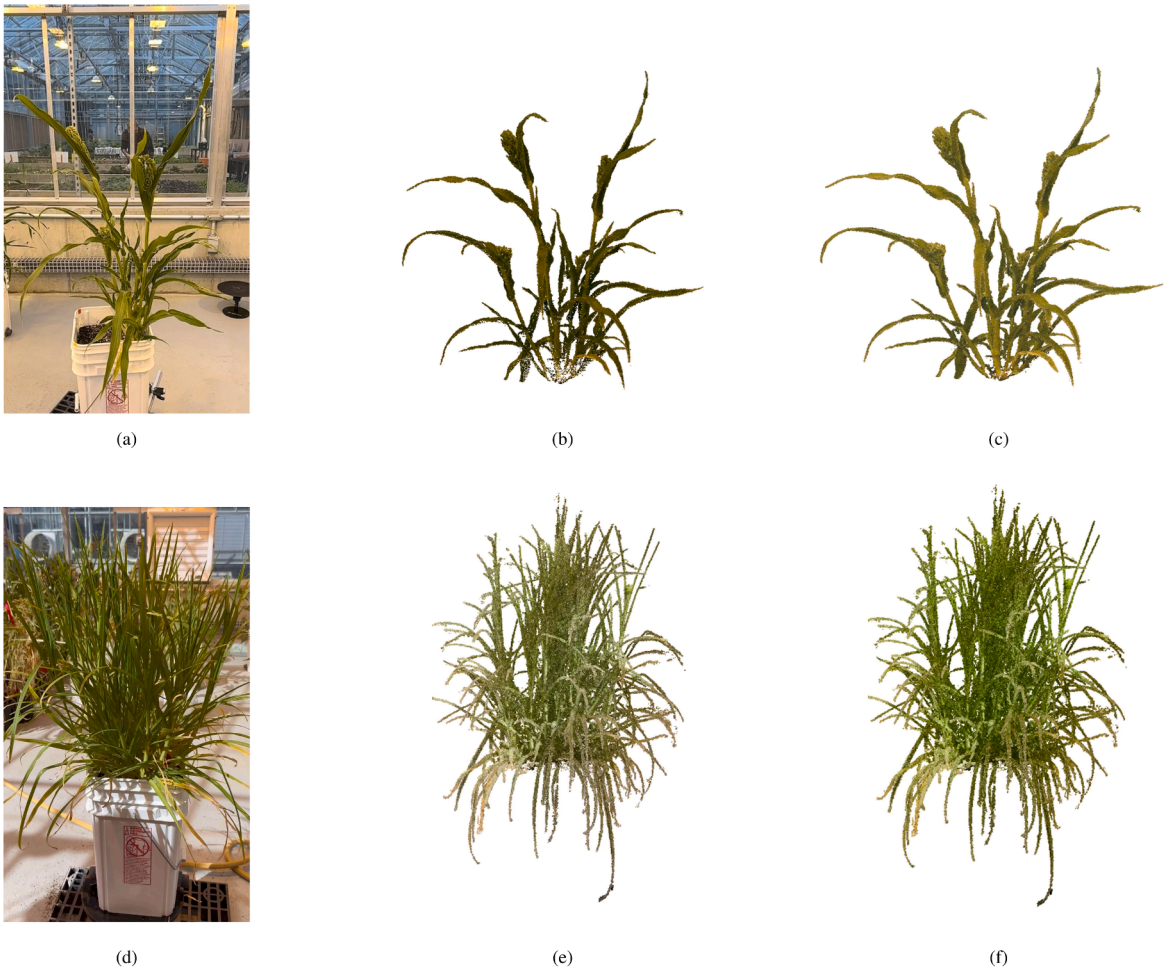


Fig. 6. Post-processing improvements showing real plant images (a and d), NeRF-generated point clouds (b and e), and color-corrected point clouds (c and f), enhancing visual consistency and model clarity.

quantitative and qualitative plant reconstructions, which significantly impact our subsequent interpretation and analysis.

Outlier Removal: Point cloud data often includes noise artifacts resulting from minor plant movements, reflective surfaces, and occlusions during image acquisition. To effectively mitigate these issues, we implemented statistical outlier removal (SOR) using the CloudCompare software [22]. Initially, we performed manual cleaning to remove major extraneous elements such as pots and background objects. Subsequently, SOR systematically removed points significantly deviating from their local neighborhoods. For rigorous cleaning scenarios demanding high noise reduction, we selected a nearest neighbor count of 50 and a standard deviation multiplier of 2. This combination effectively eliminated noise while preserving critical structural details in dense regions. Conversely, in scenarios requiring more delicate preservation of finer structural features, we employed a nearest neighbor count of 20 and a standard deviation multiplier of 1. This parameter choice provided a careful balance between noise reduction and retaining structural detail.

Color Correction: To address color inconsistencies caused by varying greenhouse lighting conditions, we applied a simple global color-shift adjustment. A target RGB value was sampled from a corresponding 2D reference image, and the nearest 3D point in the point cloud was matched to this location. The difference between the target color and the point cloud color at that location was then applied uniformly to all points:

$$\mathbf{c}_i' = \text{clip}\left(\mathbf{c}_i + \left(\mathbf{t} - \mathbf{c}_{ref}\right)\right) \quad (1)$$

where $\mathbf{c}_i$ is the original RGB of point i, $\mathbf{t}$ is the sampled image RGB, and $\mathbf{c}_{\text{ref}}$ is the RGB of the matched 3D point. This adjustment produced more consistent and natural color appearance across point clouds. This image-based correction effectively reduced lighting-induced shifts, resulting in more natural and consistent color representation across reconstructions. The adjusted point clouds were saved in '.las' format using 16-bit color encoding to preserve visual detail.

The effectiveness of our described post-processing and data cleaning approaches is evident in the improved quality and realism achieved in the final processed point clouds compared to their raw counterparts, as shown in Fig. 6. This ensured that the final point clouds accurately represented the structures of the plants and were optimized for our subsequent detailed analyses and visualization tasks.

3.4. 3D Gaussian splatting (G-Splat)

3D Gaussian Splatting (G-Splat) is a recent real-time rendering technique that reconstructs continuous radiance fields using explicit, point-based scene representations, rather than volumetric neural networks [7]. In contrast to NeRF's implicit formulation, G-Splat models the scene as a collection of anisotropic 3D Gaussians (Fig. 7a), each defined by its position, orientation, opacity, and color parameters. These Gaussians collectively approximate the surface geometry and radiance, allowing for direct rasterization through differentiable splatting. This approach enables photo-realistic rendering at interactive frame rates without the need for volumetric ray marching, significantly reducing computational cost [23,24].

(a) (b) (c)

Fig. 7. G-Splat reconstruction results. (a) Anisotropic 3D Gaussian representation of the plant showing explicit ellipsoid splats. (b) Final rendered reconstruction using the optimized Gaussian parameters. (c) Close-up view illustrating stable thin structures and high structural fidelity captured by the explicit Gaussian formulation.

Table 2
Key training settings and measurable reconstruction statistics for 3D Gaussian Splatting.

Parameter	Value
Training iterations	30,000
Spherical Harmonics degree (SH)	3
Average training time	45–60 min
GPU memory usage	3.6 GB
Final Gaussian count	0.7–3.5 million

Each Gaussian serves as a locally adaptive basis function that smoothly blends with neighboring splats, maintaining continuity and high visual fidelity even at lower sampling densities. During optimization, the parameters of these Gaussians are refined using gradient descent to minimize photometric error between rendered and reference images [7]. This explicit representation offers several advantages over volumetric methods, including eliminating the need for dense voxel grids, accelerating convergence, and supporting direct editing and real-time visualization.

In this study, we employed the open-source implementation released by the Inria GraphDeco group [7], which provides a CUDA-accelerated renderer optimized for Gaussian-based radiance fields. Table 2 details the training configuration used for G-Splat reconstructions. The pipeline begins with the same input datasets and camera poses generated through GLOMAP to ensure consistency with the NeRF reconstructions. The images and corresponding poses were processed using the G-Splat framework's standard training configuration, which optimizes Gaussian parameters via gradient descent using photometric loss functions. Training converged substantially faster than NeRF, typically completing in under one hour on the same hardware. The SH degree of 3 allows each Gaussian to model view-dependent color, improving shading realism on curved leaf surfaces. The adaptive densification strategy progressively refined Gaussian distributions in high-frequency regions, with final models containing 0.7–3.5 million Gaussians depending on plant complexity.

To further refine the reconstructions and remove stray points, we used the SuperSplat Editor [25] for lightweight manual cleanup and visual inspection of splatted scenes before final visualization.

The resulting reconstructions exhibit high spatial coherence and fine structural fidelity while maintaining real-time rendering capabilities (Fig. 7b). In particular, G-Splat effectively captures complex plant geometries, such as overlapping leaves and curved stems, that often introduce depth ambiguity in volumetric methods. The explicit point-based formulation also enabled a more stable representation of thin features (Fig. 7c), which are usually blurred or underrepresented in NeRF reconstructions due to volumetric smoothing.

Although G-Splat achieves substantial improvements in rendering speed and geometric sharpness, it has inherent limitations in modeling global illumination and view-dependent effects compared to NeRF [7]. These differences make the two approaches complementary: NeRF provides higher photometric accuracy through volumetric light modeling, while G-Splat delivers real-time performance and sharper structural detail. Together, they establish a practical balance between accuracy and efficiency, advancing 3D plant visualization and analysis for agricultural research.

3.5. Quantitative evaluation

The generation of 3D plant reconstructions necessitates a rigorous evaluation of their fidelity. To assess reconstruction quality, image-based metrics such as peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS) are employed. Rather than using independent ground-truth datasets such as terrestrial LiDAR scans, the original real images used to build the reconstructions are selected as the reference. This choice ensures that the evaluation is based on readily available data while providing a consistent benchmark for comparison. Although we acknowledge that this approach primarily evaluates rendering fidelity rather than geometric accuracy, no additional geometric measurements were collected for this study. As a result, our assessment relies exclusively on these quantitative image-based metrics.

We selected these three complementary metrics because: (1) PSNR evaluates pixel-level fidelity, which is important for color-based trait inspection; (2) SSIM captures structural similarity, which relates directly to morphological features; and (3) LPIPS aligns with human perceptual judgment, making it valuable for educational and visualization-focused applications. Together, these metrics provide a comprehensive evaluation of both visual realism and structural consistency for agricultural phenotyping tasks.

3.5.1. Coordinate system transformation

For these image-based metrics to be meaningful, the rendered views of the reconstructed models must be generated from the exact same camera poses as those used to capture the real images. This requires a careful alignment of coordinate systems. COLMAP, which provides the recovered camera poses, follows the OpenGL convention: a right-handed coordinate system where the camera forward direction is aligned with the negative Z-axis, $+X$ points to the right, and $+Y$ points upwards [26]. Unreal Engine, on the other hand, operates in a left-handed coordinate system where the forward axis is $+X$, the right axis is $+Y$, and the up axis remains $+Z$ [27]. Furthermore, COLMAP stores translations in meters, while Unreal Engine requires centimeters. Without addressing these differences, reconstructed camera trajectories and rendered images would be misaligned, leading to invalid metric calculations. Importantly, Unreal was chosen as the rendering platform to maintain continuity with the VR greenhouse framework, ensuring that evaluation and immersive visualization are performed within a unified environment.

An OpenGL-style camera-to-world matrix is defined as

$$T_{GL}^{cw} = \begin{bmatrix} R_{GL} & t_{GL} \\ 0 & 1 \end{bmatrix}, \quad R_{GL} \in SO(3),\ t_{GL} \in \mathbb{R}^3,$$

where R_{GL} denotes the rotation matrix and t_{GL} the translation vector [28]. To transform this into Unreal Engine coordinates, we apply a fixed change-of-basis matrix M:

$$M = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad M^{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The Unreal-space camera pose is obtained through conjugation, a form of similarity transformation used to express a matrix in a different basis [29]:

$$T_{UE(m)} = M\, T_{GL}\, M^{-1},$$

which yields the pose expressed in Unreal's axis conventions, but still in meters. To match Unreal's native scale, the translation vector is then multiplied by 100 to convert meters into centimeters.

$$\tilde{t}_{UE} = 100 \cdot t_{UE}, \qquad T_{UE}^{cw} = \begin{bmatrix} R_{UE} & \tilde{t}_{UE} \\ 0 & 1 \end{bmatrix}.$$

For reconstructed geometry, each point $p_{GL} \in \mathbb{R}^3$ of the point cloud in COLMAP coordinates (meters) is mapped into Unreal coordinates as

$$p_{UE} = 100 \cdot R_B\, p_{GL},$$

where R_B is the 3×3 rotation block of M and p_{UE} is the corresponding point in the transformed point cloud. If necessary, an additional reflection about the Y axis can be applied to account for importer-specific handedness issues:

$$y' = 2\bar{y} - y,$$

with $\bar{y}$ denoting the mean of the Y coordinates of the point cloud.

Finally, the camera intrinsics must be reformulated for Unreal's rendering pipeline. COLMAP provides focal lengths (f_x, f_y) in pixel units and image resolution (w, h). Unreal Engine requires field of view (FOV) and aspect ratio, computed as

$$\mathrm{FOV}_x = 2\arctan\left(\frac{w}{2f_x}\right), \qquad \mathrm{FOV}_y = 2\arctan\left(\frac{h}{2f_y}\right), \qquad AR = \frac{w}{h}.$$

When loaded into Unreal, the rotation matrices R_{UE} are converted into quaternions $q = (x, y, z, w)$ by

$$q = \mathrm{quat}(R_{UE}),$$

and subsequently into Rotators (pitch, yaw, roll) for compatibility with Unreal's orientation system.

To validate this transformation, the camera trajectories and point clouds can be visualized before and after conversion, clearly showing the change in coordinate systems. Such graphs provide intuitive confirmation that the mapping preserves geometric relationships while reconciling the differences between OpenGL and Unreal conventions. Additionally, the final Unreal visualization of the reconstructed point cloud, along with its aligned camera trajectory, can be showcased to demonstrate the success of the conversion pipeline.

3.5.2. Image quality metrics

The resulting matched image pairs were analyzed using the following metrics:

Peak Signal-to-Noise Ratio (PSNR)

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

where MAX_I represents maximum pixel intensity (255 for 8-bit images), and MSE is the Mean Squared Error between image pairs [30].

Structural Similarity Index Measure (SSIM)

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where:

- μ_x, μ_y: mean luminance values of images.
- σ_x, σ_y: standard deviations (contrast) of images.
- σ_{xy}: covariance between the images.
- C_1, C_2: stability constants [31].

Learned Perceptual Image Patch Similarity (LPIPS) Unlike PSNR and SSIM, which are based on pixel-level intensity and structural information, LPIPS measures perceptual similarity by comparing deep feature representations extracted from convolutional neural networks. Given two images x and y, their distance is defined as

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} \left\| w_l \odot \left(\hat{y}_{hw}^l - \hat{y}_{0hw}^l \right) \right\|_2^2$$

where:

- $\hat{y}_{hw}^l \in \mathbb{R}^{C_l}$ are the *channel-normalized* deep features of x at layer l and spatial location (h, w);
- $\hat{y}_{0hw}^l \in \mathbb{R}^{C_l}$ are the corresponding channel-normalized features of the reference image x_0;
- $w_l \in \mathbb{R}^{C_l}$ are learned channel-wise weights;
- H_l, W_l are the spatial height/width at layer l;
- $\odot$ denotes channel-wise (Hadamard) multiplication, and $\| \cdot \|_2$ is the ℓ_2 norm over channels.

LPIPS has been shown to correlate more strongly with human perceptual judgments of image similarity than traditional pixel-level metrics [32].

In addition to assessing visual quality using PSNR and SSIM, rendering performance and responsiveness in user interactions were quantitatively analyzed using the Meta Quest Developer Hub. Testing was conducted on a high-performance system equipped with an Intel Core i9-12900KF CPU, an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM, and 128 GB of DDR5 RAM, ensuring optimal performance for demanding VR applications. The system also utilized a 1 TB NVMe SSD for fast data loading, crucial for handling large-scale 3D point clouds in Unreal Engine. Rendering smoothness was evaluated by measuring frames per second (fps) and interaction latency was assessed by monitoring GPU rendering times (ms). Together, these performance metrics comprehensively validated both the visual realism and interactive practicality of the VR visualization system for precision agriculture applications.

4. Immersive virtual visualization

We used Unreal Engine to develop an immersive virtual greenhouse environment that connects reconstructed plant models with an interactive, spatially realistic experience. The current implementation was developed and tested using NeRF-based point cloud reconstructions, allowing users to explore detailed plant structures at an accurate scale. The refined NeRF point clouds were imported through the LiDAR Point Cloud Plugin, which enabled high-fidelity visualization of dense reconstructions within the virtual environment.

To maintain stable performance and visual quality, several rendering optimizations were implemented. Level of Detail (LOD) settings were adjusted to minimize abrupt changes in point density, using an LOD factor of 1 and an extended LOD distance multiplier of 0.7. The draw distance was set to 12,000 units to ensure distant plants remained visible, while an octree depth of 7 provided efficient spatial data management. A four-billion-point rendering budget preserved the density of the reconstructed models during interaction. Together, these settings strike a balance between real-time performance and visual realism.

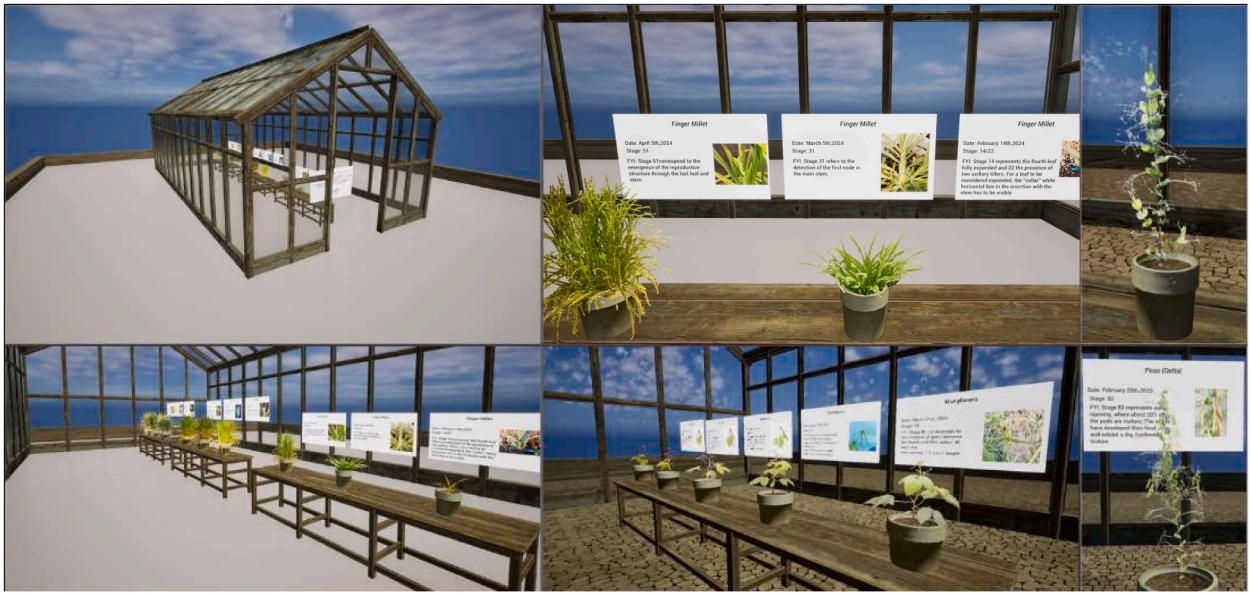


Fig. 8. Virtual greenhouse visualization showing reconstructed millet, mungbean, and pea plants with interactive growth-stage information panels. Each plant is paired with descriptive notecards displaying collection dates, growth stages, and visual references, providing an immersive environment for comparative analysis of plant development in 3D.

Lighting and interaction features were designed to replicate real greenhouse conditions. A dynamic skybox and physically based lighting ensured accurate color rendering and depth perception. The Niagara particle system was utilized to introduce subtle interactive elements, such as dynamic dust and responsive light reflections, thereby enhancing realism without compromising the frame rate. Within the VR interface, users can manually rotate plants, zoom in on specific regions, and display informational overlays that show phenotypic data. These features make the virtual greenhouse a valuable tool for both analysis and education.

While the NeRF-based models offered high visual detail, they were difficult to package and deploy for standalone VR systems. The LiDAR Point Cloud Plugin used for NeRF imports is not supported for packaging in Unreal Engine, and the large model sizes made real-time rendering impractical for lightweight headsets. To overcome these challenges, we integrated 3D Gaussian Splatting (G-Splat) reconstructions into the virtual greenhouse using the XScene Unreal Engine Plugin [33]. The plugin offers native support for importing and rendering Gaussian Splat scenes directly within Unreal Engine, thereby eliminating the dependency on external point cloud plugins. G-Splat models are computationally lightweight and can be fully packaged into standalone VR builds, making them suitable for devices such as the Meta Quest series. This setup allows smooth real-time exploration of dense plant scenes with consistent frame rates and minimal latency.

The virtual greenhouse can now be deployed as an executable application through the incorporation of G-Splat via XScene, expanding its usability beyond workstation environments. This advancement makes the system portable, scalable, and suitable for broader applications such as field demonstrations, student training, and collaborative visualization.

The integrated virtual greenhouse platform, which combines optimized visualization settings, interactive features, and immersive VR exploration, enhances both data interpretability and user engagement. The system provides a unified tool for visualizing plant structure, monitoring growth stages, and supporting interactive analysis through the balance of high-fidelity detail in NeRF reconstructions with the efficiency and deployability of G-Splat. Its flexibility makes it well-suited for research, education, and collaborative applications in precision agriculture, bridging the gap between detailed 3D reconstruction and accessible, real-time visualization.

5. Results

The reconstructed 3D plant models generated using NeRF are shown in Fig. 9, covering all five species across their full developmental progression. Each reconstruction corresponds to an individual plant and is annotated with its associated BBCH, providing a visual overview of morphological changes from early vegetative growth to reproductive maturity. These qualitative results demonstrate NeRF's ability to recover coherent canopy structure, smooth radiance transitions, and distinct species-level architectural traits. To further assess reconstruction fidelity beyond visual inspection, we performed a quantitative evaluation using the Unreal Engine-based comparison workflow.

Quantitative evaluation of reconstruction quality was performed using the Unreal Engine-based comparison workflow described in the quantitative evaluation section. This setup generated corresponding image pairs from real and rendered camera views, allowing consistent calculation of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). Since these comparisons rely on 2D reference images rather than independent 3D measurements (e.g., LiDAR scans or manual plant measurements), the metrics evaluate rendering fidelity and appearance consistency rather than direct geometric ground truth. PSNR represents pixel-level fidelity, SSIM captures structural similarity, and LPIPS measures perceptual difference using learned image features. Together, these metrics provide a comprehensive assessment of rendering fidelity and perceptual realism in the generated views, rather than direct validation of metric 3D geometric accuracy.

Two categories of evaluation were considered: unseen images, which were not used during training, and known images, which were included in the training dataset. As shown in Fig. 10, NeRF and G-Splat reconstructions are displayed side-by-side for both trained and unseen viewpoints, allowing direct visual comparison against the corresponding real images. This distinction enables an understanding of the performance of model generalization versus memorization under realistic visualization conditions.

Across all datasets, G-Splat achieved higher PSNR (13.9–18.7 dB) and SSIM (0.33–0.62) values compared to NeRF, indicating better structural consistency and edge preservation. In contrast, NeRF consistently achieved lower LPIPS values (0.09–0.18), demonstrating higher

Fig. 9. NeRF reconstructions for each plant species with corresponding BBCH labels shown beneath each reconstruction to indicate growth progression.

perceptual similarity and smoother visual continuity. For this verification, five sets of unseen images and five sets of known images were evaluated for each dataset. The quantitative values reported in Tables 3 and 4 represent the mean ± standard deviation across these image sets, providing an estimate of both the average reconstruction performance and its variability across viewpoints. This confirms a fundamental trade-off between geometric precision and photometric coherence: G-Splat prioritizes sharpness and local structure, while NeRF emphasizes global realism and radiance consistency.

For unseen viewpoints, G-Splat outperformed NeRF in structural metrics across most crops, particularly in the Mungbeans and Delta Peas datasets, which achieved the highest SSIM (0.62 and 0.59, respectively). NeRF, however, maintained stable perceptual quality across all datasets, producing smoother transitions and fewer visible artifacts even under occluded or complex canopy structures. This observation aligns with prior work showing that NeRF models tend to optimize for photometric consistency rather than exact pixel-level alignment [6,21,34].

When evaluating known views that were part of the training set, both models exhibited measurable improvements. NeRF showed increases of 0.3–1.3 dB in PSNR and 0.03–0.15 in SSIM, accompanied by lower LPIPS, confirming its ability to refine view-dependent features once radiance information is explicitly learned. G-Splat maintained nearly identical PSNR and SSIM values across known and unknown views, indicating strong generalization and minimal overfitting. However, its slightly higher LPIPS (0.36–0.48) suggests that while G-Splat accurately retains geometric structure, it sacrifices some perceptual smoothness in texture continuity. Despite these improvements, both approaches show limitations in reproducing exact geometric scale and perspective. Certain structures appear slightly shifted or distorted relative to real images, and some elements are partially occluded from specific viewpoints. This indicates that while both models capture overall structure and radiance, neither fully preserves precise spatial relationships.

SSIM has been shown to better capture luminance, local correlation, and structural relationships aligned with human visual perception [30,31,35]. The consistently higher SSIM values observed for G-Splat therefore indicate improved preservation of spatial detail and canopy morphology. Meanwhile, NeRF's lower LPIPS scores reflect higher perceptual realism, consistent with previous work emphasizing its optimization for appearance-based coherence rather than pixel-wise accuracy [6,9].

From an application standpoint, the distinction between these results becomes particularly relevant for agricultural visualization. NeRF reconstructions are advantageous for tasks requiring smooth visual representation, such as trait quantification, canopy visualization, and growth-stage comparison. Their ability to maintain radiance continuity makes them highly suitable for evaluating morphological traits and subtle color gradients in plant phenotyping [36–38]. On the other hand, G-Splat excels at generating sharp, geometrically consistent representations that can be efficiently rendered in real-time. The Gaussian-based representation offers exceptional performance in Unreal Engine, enabling interactive viewing and analysis on standalone headsets, such as the Meta Quest series.

Overall, these findings highlight the complementary nature of the two approaches. NeRF offers perceptual accuracy and photometric realism, providing reliable reconstructions that emphasize visual coherence and smooth color transitions. G-Splat, on the other hand, ensures strong structural fidelity, fast rendering, and scalability for large datasets. In this context, structural fidelity refers to the preservation of visible thin plant structures and local detail in rendered views, rather than independently validated metric-scale geometric accuracy. Together, they establish a balanced and versatile framework for accurate, interpretable, and immersive 3D plant reconstruction across multiple crop types and developmental stages, effectively bridging scientific analysis and interactive visualization within the VR greenhouse environment.

5.1. VR performance

To assess the real-time performance of reconstructed models in immersive visualization, we measured the system's frame rate, latency, and GPU utilization during interaction inside the VR greenhouse. The evaluation compared NeRF-based and G-Splat-based reconstructions, both visualized under identical environmental conditions to ensure fairness.

The NeRF-based models maintained a consistent 72 FPS at the Meta Quest 2's native 72 Hz refresh rate, with an average total frame time of

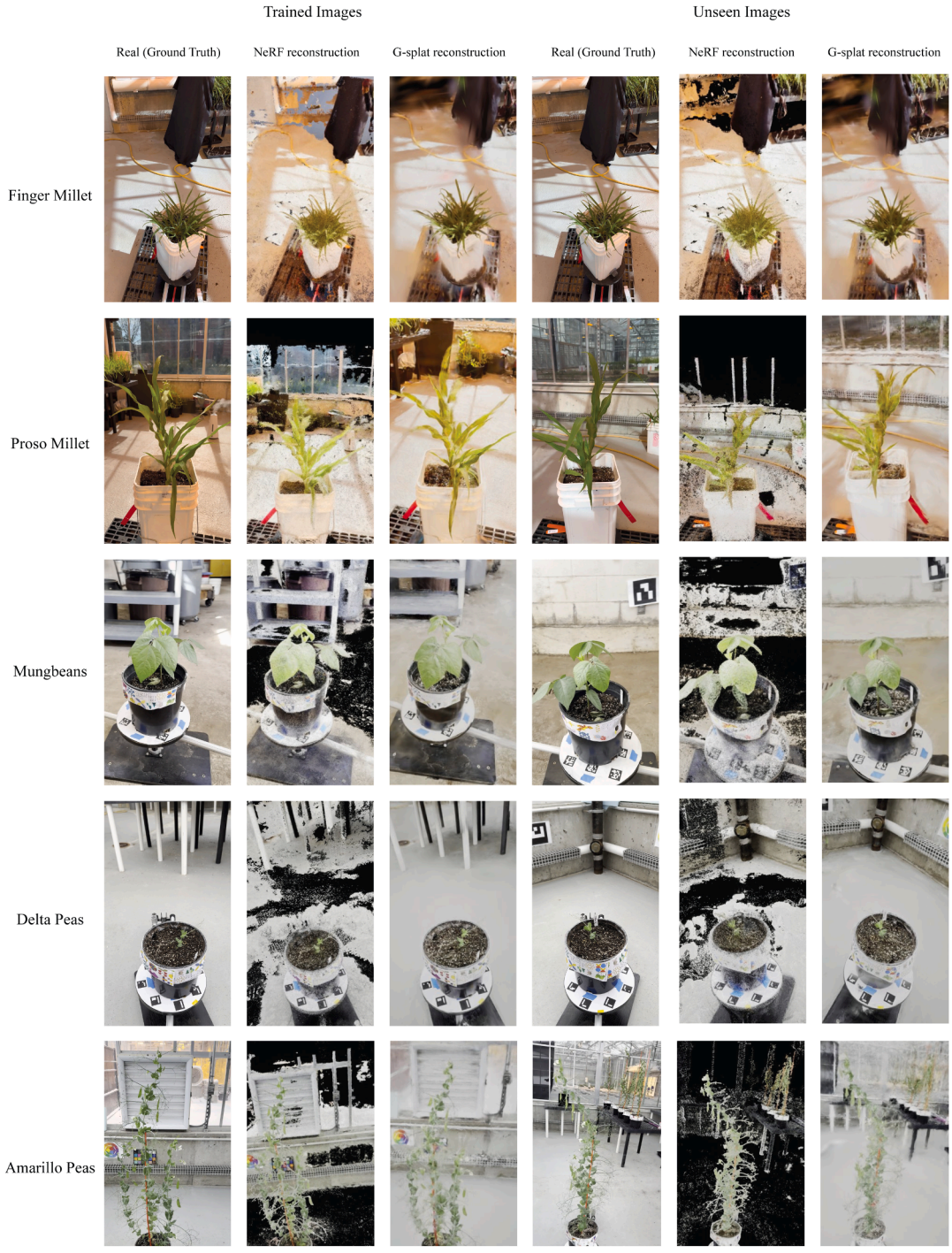


Fig. 10. Visual comparison of reconstructed scenes from trained (known) and unseen viewpoints for NeRF and G-Splat.

Table 3
Quantitative comparison of NeRF and G-Splat reconstructions on unseen (unknown) image views. Values are mean ± standard deviation.

	NeRF			G-Splat		
Unseen Images	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Finger Millet	14.534 ± 0.697	0.304 ± 0.141	0.115 ± 0.038	13.981 ± 0.956	0.332 ± 0.050	0.450 ± 0.057
Proso Millet	14.555 ± 0.682	0.228 ± 0.052	0.094 ± 0.032	13.244 ± 0.887	0.374 ± 0.020	0.448 ± 0.045
Delta Peas	15.024 ± 0.379	0.320 ± 0.022	0.178 ± 0.051	16.773 ± 1.796	0.596 ± 0.090	0.271 ± 0.099
Amarillo Peas	14.100 ± 1.266	0.049 ± 0.084	0.111 ± 0.021	16.944 ± 1.121	0.521 ± 0.030	0.403 ± 0.069
Mungbeans	14.594 ± 0.742	0.234 ± 0.080	0.092 ± 0.017	18.742 ± 1.219	0.620 ± 0.060	0.263 ± 0.062

Table 4
Quantitative comparison of NeRF and G-Splat reconstructions on trained (known) image views. Values are mean ± standard deviation.

	NeRF			G-Splat		
Known Images	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Finger Millet	14.487 ± 1.070	0.347 ± 0.139	0.090 ± 0.042	14.089 ± 0.312	0.354 ± 0.060	0.448 ± 0.037
Proso Millet	15.888 ± 0.475	0.382 ± 0.072	0.047 ± 0.022	13.249 ± 0.551	0.370 ± 0.030	0.481 ± 0.026
Delta Peas	15.445 ± 1.034	0.323 ± 0.061	0.148 ± 0.031	16.360 ± 1.545	0.596 ± 0.060	0.295 ± 0.097
Amarillo Peas	14.473 ± 0.643	0.136 ± 0.120	0.119 ± 0.047	17.582 ± 1.293	0.511 ± 0.060	0.382 ± 0.063
Mungbeans	14.620 ± 0.804	0.275 ± 0.061	0.088 ± 0.017	17.520 ± 1.327	0.564 ± 0.060	0.364 ± 0.014

13.88 ms. Latency analysis showed an App GPU Time of 1.69 ms and a Compositor GPU Time of 2.39 ms, both well below the 13.9 ms budget required for real-time performance. GPU utilization averaged 33 % at 525 MHz, and CPU usage remained moderate at 37 % (average) with a stable 1.38 GHz CPU frequency (Level 3), indicating smooth operation without thermal throttling. Rendering was performed at 1024 × 1024 per eye, with foveated rendering disabled to maintain a uniform test condition.

In comparison, the G-Splat models also maintained a stable 72 FPS, achieving an average frame time of 13.86 ms at a higher rendering resolution of 1840 × 1920 per eye. The App GPU Time (1.70 ms) and Compositor GPU Time (2.41 ms) were nearly identical to the NeRF case, while GPU utilization (33 %) and CPU usage (≈ 43 %) indicated balanced, efficient performance. The CPU frequency (1.48 GHz, Level 3) and shader occupancy (≈ 81 %) confirmed stable real-time rendering without performance drops.

Overall, both reconstructions achieved smooth, low-latency visualization on the standalone headset; however, G-Splat demonstrated higher rendering efficiency, sustaining equal performance at a significantly higher eye-buffer resolution. This highlights the scalability and real-time suitability of G-Splat for immersive visualization in VR-based plant phenotyping.

6. Conclusions

This study developed a unified pipeline integrating Neural Radiance Fields (NeRF), 3D Gaussian Splatting (G-Splat), and Virtual Reality (VR) visualization to advance plant phenotyping and precision agriculture. Using finger millet, proso millet, mungbean, and pea plants (Delta and Amarillo) as representative crops, we demonstrated how image-based 3D reconstructions can be transformed into immersive VR environments that reveal plant structure, growth stages, and morphology with high visual fidelity. The workflow combined greenhouse data collection, globally optimized camera pose estimation through GLOMAP, reconstruction via NeRF and G-Splat, and interactive rendering in Unreal Engine, linking quantitative modeling with intuitive visualization.

Quantitative evaluation based on PSNR, SSIM, and LPIPS, averaged across five sets of trained (known) and five sets of unseen images per dataset, highlighted the complementary strengths of both methods. G-Splat achieved higher PSNR (13.9–18.7 dB) and SSIM (0.33–0.62), indicating strong geometric fidelity, while NeRF achieved lower LPIPS (0.09–0.18), confirming smoother perceptual realism. These findings illustrate a trade-off between geometric precision and photometric coherence: G-Splat preserves fine structural detail, whereas NeRF maintains global radiance and appearance consistency. When visualized in VR, both models rendered efficiently with stable frame rates, enabling real-time interactive exploration.

Some challenges remain, particularly in translating camera poses from COLMAP or GLOMAP into Unreal Engine due to differences in coordinate systems. Future work will focus on automating this process and exploring faster reconstruction algorithms such as Instant-NGP, Mip-NeRF, or newer G-Splat variants to enhance quality and computational efficiency. Extending this pipeline to field-scale agricultural environments introduces several additional challenges beyond controlled greenhouse conditions. Outdoor deployments must account for variability in illumination, increased occlusions from surrounding vegetation, and wind-induced plant motion during image acquisition [39]. In larger agricultural settings, pose estimation may also become more difficult due to the absence of nearby visual anchors. However, many of these challenges can be addressed using established field imaging practices. For example, ground control points (GCPs), which are already widely used in drone mapping workflows, can be deployed across fields and aligned with GPS coordinates to improve camera pose estimation over larger areas [40]. Similarly, illumination variability can be mitigated by including a color calibration chart during data collection to provide a consistent reference for color correction [41]. Wind-related motion effects can also be reduced by capturing data during stable weather conditions or by using drone-based imaging to shorten acquisition time and minimize environmental variation during capture. With these adaptations, the proposed reconstruction and visualization framework has the potential to scale beyond greenhouse experiments toward larger field phenotyping scenarios.

Overall, this framework demonstrates that integrating neural reconstruction with immersive visualization provides a practical and scalable approach for 3D plant analysis. The combination of NeRF's perceptual realism, G-Splat's structural precision and rendering efficiency, and Unreal Engine's interactivity establishes a foundation for future phenotyping research and broader adoption of digital tools in agriculture.

CRediT authorship contribution statement

Shambhavi Joshi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Juan Di Salvo:** Writing – review & editing, Resources, Methodology, Investigation, Data curation; **Yanben Shen:** Writing – review & editing, Resources, Investigation, Data curation; **Mozhgan Hadadi:** Writing – review & editing, Methodology, Investigation; **Venkata Naresh Boddepalli:** Writing – review & editing, Resources, Investigation, Data curation; **Talukder Z. Jubery:** Writing – review & editing, Supervision, Methodology, Conceptualization; **Soumik Sarkar:** Writing – review & editing, Supervision, Conceptualization, Methodology; **Arti Singh:** Writing – review & editing, Supervision, Conceptualization, Methodology; **Baskar Ganapathysubramanian:** Writing – review & editing, Supervision, Conceptualization, Methodology; **Asheesh K. Singh:** Writing – review & editing, Supervision, Conceptualization, Methodology; **Adarsh Krishnamurthy:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Data availability

The code used for generating the 3D models in this work will be publicly available on GitHub (https://github.com/idealab-isu). Data will be made available on request.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (OpenAI), Claude (Anthropic), and Grammarly in order to improve the language, clarity, and readability of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the AI Institute for Resilient Agriculture (USDA-NIFA 2021-67021-35329), NSF BTT-EAGER IOS-1842097, and Iowa State University's Plant Science Institute.

Appendix A. BBCH growth stages

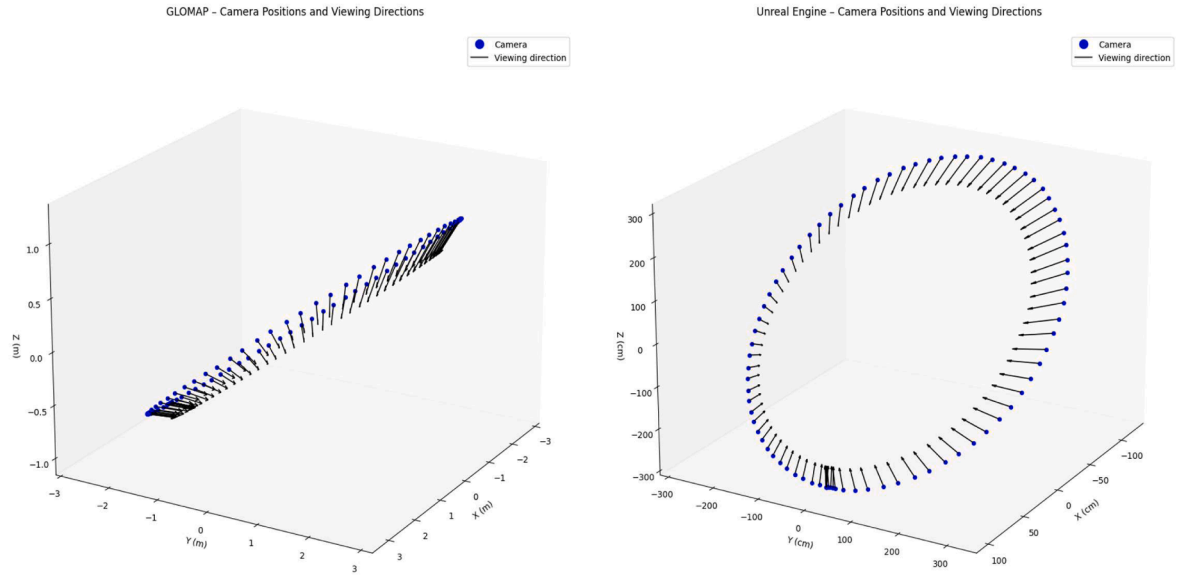
Table A.1

Table A.1
Growth stages, days after planting, and key morphological features documented for all species. Finger millet, proso millet, and both pea varieties were annotated using the BBCH decimal scale, covering leaf development, stem elongation, flowering, and fruit ripening. Mungbean stages follow the V/R system (VC-V2; R1-R7), which parallels BBCH developmental phases but uses legume-specific terminology. The table summarizes the phenological state of each species at its corresponding imaging date.

Species	BBCH Stage	Days After Planting	Key Morphological Features
Finger Millet	14/22	13	Fourth leaf fully expanded; two axillary tillers present
	31	32	First node detectable in the main stem
	51	64	Initial emergence of reproductive structure
	59	67	90% of reproductive structure emerged
	65	71	50% of anthers emerged (onset of flowering)
	69	77	90% of anthers emerged (end of flowering)
	85	92	Seed content soft but dry
	87	99	Seed content hard and dry
	92	104	Physiological maturity
Proso Millet	13	13	Third leaf fully expanded
	32/26/17	25	Two nodes detectable; six axillary tillers; seven leaves expanded
	55	29	Panicle 50% emerged
	65	36	50% flowers open; orange anthers visible
	69/85	53	90% flowers open; early soft-dry grain stage
	87	67	Seed content hard and dry
	92	67	Physiological maturity (black layer formation)
Mungbean	VC	9	Cotyledons expanded; unifoliate leaves unrolling
	V1	17	Unifoliate leaves fully expanded
	V2	22	First trifoliate leaf fully expanded
	R1	42	First open flower on upper nodes
	R2	47	Pod initiation (1–2 cm) between nodes 4–5
	R3	52	Pod elongation (5–7 cm); early seed formation
	R5	63	First pods turning black/brown
	R6	72	85–90% pods turned black/brown
	R7	79	100% pods mature (black/brown)
Delta Peas	14	15	4 leaves unfolded (leaf development)
	1/14	43	14 leaves unfolded at 14th node
	61	54	Beginning of flowering; 10% flowers open
	71	57	10% of pods at typical length; juicy when pressed
	74	72	40% of pods at typical length
	82	83	20% of pods ripe; seeds dry and hard
	85	90	50% of pods ripe; seeds dry and hard
	89	107	Fully ripe; all pods dry and brown
Amarillo Peas	14	15	4 leaves unfolded (leaf development)
	1/14	43	14 leaves unfolded at 14th node
	51	54	First flower buds visible outside leaves
	61	57	Beginning of flowering; 10% flowers open
	72	72	20% of pods at typical length; juice exudes if pressed
	81	83	10% of pods ripe; seeds dry and hard
	84	90	40% of pods ripe; seeds dry and hard
	89	107	Fully ripe; all pods dry and brown

Appendix B. Camera pose visualizations

Fig. B.1



(a) (b)

Fig. B.1. Visualization of camera trajectories before and after coordinate conversion. The first plot (a) shows the GLOMAP-estimated real-world camera positions and viewing directions expressed in the OpenGL convention used by COLMAP and NerfStudio, where the camera forward axis is $-Z$, right is $+X$, and up is $+Y$ in a right-handed, meter-scaled coordinate frame. The second plot (b) shows the same camera poses after transformation into Unreal Engine's left-handed coordinate system, where the forward axis becomes $+X$, right becomes $+Y$, and up becomes $+Z$, along with a required scale change from meters to centimeters. This conversion includes flipping axes to change handedness and applying a uniform $\times 100$ scale factor, ensuring that real-world GLOMAP trajectories correctly map to Unreal Engine for rendering-based evaluation.

References

[1] M. T. Chiu, X. Xu, Y. Wei, Z. Huang, A. G. Schwing, R. Brunner, H. Khachatrian, H. Karapetyan, I. Dozier, G. Rose, et al., Agriculture-vision: a large aerial image database for agricultural pattern analysis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2828–2838.

[2] S. Paulus, Measuring crops in 3D: using geometry for plant phenotyping, Plant Methods 15 (2019) 103.

[3] F. Okura, 3D modeling and reconstruction of plants and trees: A cross-cutting review across computer graphics, vision, and plant phenotyping, Breed. Sci. 72 (2022) 31–47.

[4] S. Paulus, J. Behmann, A.-K. Mahlein, L. Plümer, H. Kuhlmann, Low-cost 3D systems: suitable tools for plant phenotyping, Sensors 14 (2014) 3001–3018.

[5] N. Tilly, D. Hoffmeister, Q. Cao, S. Huang, V. Lenz-Wiedemann, Y. Miao, G. Bareth, Multitemporal crop surface models: accurate plant height measurement and biomass estimation with terrestrial laser scanning in paddy rice, J. Appl. Remote Sens. 8 (2014) 083671–083671.

[6] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, R. Ng, NeRF: representing scenes as neural radiance fields for view synthesis, Commun. ACM 65 (2021) 99–106.

[7] B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis, 3D Gaussian splatting for real-time radiance field rendering., ACM Trans. Graph. 42 (2023) 139–1.

[8] M. A. Arshad, T. Jubery, J. Afful, A. Jignasu, A. Balu, B. Ganapathysubramanian, S. Sarkar, A. Krishnamurthy, Evaluating neural radiance fields for 3D plant geometry reconstruction in field conditions, Plant Phenomics 6 (2024) 0235.

[9] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, P. P. Srinivasan, MIP-NeRF: a multiscale representation for anti-aliasing neural radiance fields, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5855–5864.

[10] T. Müller, A. Evans, C. Schied, A. Keller, Instant neural graphics primitives with a multiresolution hash encoding, ACM Trans. Graph. 41 (2022) 1–15.

[11] M. Tancik, E. Weber, E. Ng, R. Li, B. Yi, T. Wang, A. Kristoffersen, J. Austin, K. Salahi, A. Ahuja, et al., NeRFStudio: a modular framework for neural radiance field development, in: ACM SIGGRAPH 2023 Conference Proceedings, 2023, pp. 1–12.

[12] U. Meier, H. Bleiholder, L. Buhr, C. Feller, H. Hack, M. Heß, P. D. Lancashire, U. Schnock, R. Stauß, T. Van Den Boom, et al., The BBCH system to coding the phenological growth stages of plants–history and publications, J. Kulturpflanz. 61 (2009) 41–52.

[13] I. Amadou, M. E. Gounga, G.-W. Le, Millets, nutritional composition, some health benefits and processing, Emir. J. Food Agric. 25 (2013) 501–508.

[14] T. Warkentin, A. Vandenberg, B. Tar'an, S. Banniza, G. Arganosa, B. Barlow, S. Ife, J. Horner, D. de Silva, M. Thompson, et al., CDC Amarillo yellow field pea, Can. J. Plant. Sci. 94 (2014) 1539–1541.

[15] J. Smitchger, N. Weeden, Quantitative trait loci controlling lodging resistance and other important agronomic traits in dry field peas, Crop Sci. 59 (2019) 1442–1456.

[16] Polycam.Inc, Polycam, https://poly.cam/, 2024.

[17] J. L. Schonberger, J.-M. Frahm, Structure-from-motion revisited, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2016, pp. 4104–4113.

[18] J. L. Schönberger, E. Zheng, J.-M. Frahm, M. Pollefeys, Pixelwise view selection for unstructured multi-view stereo, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14, Springer, 2016, pp. 501–518.

[19] L. Pan, D. Baráth, M. Pollefeys, J. L. Schönberger, Global structure-from-motion revisited, in: European Conference on Computer Vision, Springer, 2024, pp. 58–77.

[20] K. Hu, W. Ying, Y. Pan, H. Kang, C. Chen, High-fidelity 3D reconstruction of plants using neural radiance fields, Comput. Electron. Agric. 220 (2024) 108848.

[21] M. Yao, Y. Huo, Y. Ran, Q. Tian, R. Wang, H. Wang, Neural radiance field-based visual rendering: A comprehensive review, arXiv preprint arXiv:2404.00714 (2024).

[22] CloudCompare, Cloudcompare, https://www.cloudcompare.org, 2024.

[23] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, A. Kanazawa, Plenoxels: radiance fields without neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5501–5510.

[24] M. Zwicker, H. Pfister, J. Van Baar, M. Gross, EWA splatting, IEEE Trans. Vis. Comput. Graph. 8 (2002) 223–238.

[25] SuperSplat, Supersplat editor, https://superspl.at/editor/, 2024. Accessed: 2025-10-04.

[26] NeRFstudio Team, Data conventions – NeRFstudio documentation, NeRFstudio, 2023. URL: https://docs.nerf.studio/quickstart/data_conventions.html.

[27] Epic Games, Coordinate system and spaces in unreal engine, Epic games, 2025. URL: https://dev.epicgames.com/documentation/en-us/unreal-engine/coordinate-system-and-spaces-in-unreal-engine.

[28] J. de Vries, Camera / View space (LearnOpenGL tutorial), 2025. URL: https://learnopengl.com/Getting-started/Camera.

[29] G. H. Golub, C. F. Van Loan, Matrix Computations, JHU press, 2013.

[30] A. Hore, D. Ziou, Image quality metrics: PSNR vs. SSIM, in: International Conference on Pattern Recognition, IEEE, 2010, pp. 2366–2369.

[31] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.

[32] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 586–595.

[33] Xverse Engine, XScene unreal engine plugin, https://github.com/xverse-engine/XScene-UEPlugin, 2025. Accessed: 2025-10-22.

[34] K. Zhang, G. Riegler, N. Snavely, V. Koltun, NeRF++: analyzing and improving neural radiance fields, arXiv preprint arXiv:2010.07492 (2020).

[35] Z. Wang, A. C. Bovik, Mean squared error: Love it or leave it? A new look at signal fidelity measures, IEEE Signal Process. Mag. 26 (2009) 98–117.

[36] F. Fiorani, U. Schurr, Future scenarios for plant phenotyping, Annu. Rev. Plant Biol. 64 (2013) 267–291.

[37] N. Harandi, B. Vandenberghe, J. Vankerschaver, S. Depuydt, A. Van Messem, How to make sense of 3D representations for plant phenotyping: A compendium of processing and analysis techniques, Plant Methods 19 (2023) 60.

[38] P. Sodhi, S. Vijayarangan, D. Wettergreen, In-field segmentation and identification of plant structures using 3D imaging, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2017, pp. 5180–5187.

[39] E. Omia, E. Park, D. Semyalo, R. Joshi, B. K. Cho, Advancements in 3D field-crop phenotyping using point clouds: a comparative review of sensor technology, target traits, and challenges under controlled and field conditions, Front. Plant Sci., 17, 1731852, 2026.

[40] S. Harwin, A. Lucieer, Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from unmanned aerial vehicle (UAV) imagery, Remote Sens., 4, 6, 1573–1599, 2012.

[41] S. Sunoj, C. Igathinathane, N. Saliendra, J. Hendrickson, D. Archer, Color calibration of digital images for agriculture and other applications, ISPRS J. Photogramm. Remote Sens., 146, 221–234, 2018.