agriculture

MDPI

Article

Automated 3D Phenotyping of Maize Plants: Stereo Matching Guided by Deep Learning

Juan Zapata-Londoño [1,2], Juan Botero-Valencia [2,*], Ítalo A. Torres [3,4], Erick Reyes-Vera [1] and Ruber Hernández-García [3,4]

[1] Grupo Automática, Electrónica y Ciencias Computacionales, Faculty of Engineering, Instituto Tecnológico Metropolitano—ITM, Medellin 050034, Colombia
[2] Grupo Sistemas de Control y Robótica, Faculty of Engineering, Instituto Tecnológico Metropolitano—ITM, Medellin 050034, Colombia
[3] Laboratory of Technological Research in Pattern Recognition—LITRP, Facultad de Ciencias de la Ingeniería, Universidad Católica del Maule, Talca 3480112, Chile
[4] Department of Computing and Industries, Facultad de Ciencias de la Ingeniería, Universidad Católica del Maule, Talca 3480112, Chile
* Correspondence: juanbotero@itm.edu.co

Abstract

Automated three-dimensional plant phenotyping is an essential tool for non-destructive analysis of plant growth and structure. This paper presents a low-cost system based on stereo vision for depth estimation and morphological characterization of maize plants. The system incorporates an automatic detection stage for the object of interest using deep learning techniques to delimit the region of interest (ROI) corresponding to the plant. The Semi-Global Block Matching (SGBM) algorithm is applied to the detected region to compute the disparity map and generate a partial three-dimensional representation of the plant structure. The ROI delimitation restricts the disparity calculation to the plant area, reducing processing of the background and optimizing computational resource use. The deep learning-based detection stage maintains stable foliage identification even under varying lighting conditions and shadowing, ensuring consistent depth data across different experimental conditions. Overall, the proposed system integrates detection and disparity estimation into an efficient processing flow, providing an accessible alternative for automated three-dimensional phenotyping in agricultural environments.

Keywords: three-dimensional phenotyping; plant growth and structure; non-destructive analysis; semi-global stereo matching

check for updates

Academic Editors: Adrian Gracia-Romero and Riccardo Rossi

Received: 11 November 2025
Revised: 10 December 2025
Accepted: 10 December 2025
Published: 12 December 2025

Citation: Zapata-Londoño, J.; Botero-Valencia, J.; Torres, Í.A.; Reyes-Vera, E.; Hernández-García, R. Automated 3D Phenotyping of Maize Plants: Stereo Matching Guided by Deep Learning. *Agriculture* **2025**, *15*, 2573. https://doi.org/10.3390/agriculture15242573

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Traditionally, phenotypic measurement and analysis have been laborious, costly, and time-consuming processes [1]. The use of three-dimensional (3D) imaging technologies has facilitated the study of plants in agricultural research, focusing on the systematic quantification of morphological and structural attributes, such as canopy architecture, leaf area, height, stem diameter, and biomass, among other relevant parameters [1–3]. These morphological traits serve as indicators of factors such as stress, yield, growth, and overall plant development [4,5]. Among 3D characterization methods, stereoscopic vision has established itself as an attractive alternative due to its balance between accuracy, low cost, and ease of implementation [6]. This approach captures the spatial and geometric relationships of plant structures, providing complementary information to two-dimensional methods [7–9].

Several studies have demonstrated the effectiveness of stereo vision in automated crop characterization [10,11]. Dandrifosse et al. [12] developed a stereo vision system to characterize wheat canopy architecture in the field, evaluating parameters such as height and leaf area. The results showed high accuracy compared to manual measurements, with an RMSE of 0.37 for leaf area and 97.1% agreement in canopy height estimation. According to Kim et al. [3], stereo vision also enables automated estimation of crop height, achieving high correlation with manual measurements (R^2 between 0.78 and 0.84). The study by Sampaio et al. [13] proposed a system based on RGB-D images that combines color (RGB) and depth (D), integrating segmentation and volumetric fusion to obtain accurate three-dimensional reconstructions of corn plants in dynamic conditions. On the other hand, Wen et al. [14] developed a stereo system to estimate the height of wheat stalks and adjust the position of a combine harvester's header in real time, achieving an average error of 5.5 cm compared to manual measurements.

The 3D data acquisition system developed in this work is based on a binocular stereoscopic vision scheme, supported by the favorable cost-benefit ratio widely documented in the literature. Stereoscopic systems are characterized by their low cost (i.e., $100–$1000 approximately), high video transmission speed, and ability to operate both indoors and outdoors, making them suitable for agricultural environments with variable lighting conditions [6]. These characteristics are essential factors in ensuring accessibility and efficiency in capturing large volumes of phenotypic data. Despite their advantages, the literature also reports an inherent weakness in stereo vision-based systems, which is associated with their high degree of dependence on calibration algorithms [15] and stereo correspondence algorithms [16–18], which can significantly affect the quality of the disparity map and, therefore, the accuracy of the three-dimensional reconstruction. However, by using an appropriate methodology that integrates robust stereo detection and correspondence, it is possible to mitigate these limitations and obtain consistent results even under variable environmental conditions.

Extracting phenotypic characteristics from 3D data presents various difficulties in outdoor environments [12]. Unlike controlled laboratory settings, field-captured images include environmental elements, such as soil and adjacent vegetation, which can be mistaken for the plant of interest and affect segmentation. Therefore, prior detection of the plant is an essential step, as it allows the region of interest (ROI) to be delimited and the analysis to be focused solely on the plant area, avoiding background interference [19]. However, acquisition under variable conditions of natural lighting, cloud cover, or projected shadows directly affects detection accuracy, making it difficult to correctly identify the plant relative to its surroundings.

In this context, this article introduces a low-cost, automated 3D phenotyping system that reconstructs and analyzes the morphology of maize plants in the field using stereo vision. The main novelty and contribution of this research lies in the direct integration of a deep learning-based detector into the stereo matching process, allowing the region of interest to be dynamically narrowed and the disparity calculation to be confined solely to the plant volume. This strategy reduces redundant background processing, reduces computation time, and improves the robustness of the depth map, enabling its execution on low-power embedded platforms. Rather than replacing CNN-based segmentation strategies, our proposal represents an alternative suited for scenarios where on-device execution and constrained computing environments must be prioritized.

The rest of the paper is structured into four sections. Section 2 describes the hardware and algorithms employed in the implementation of the system. Section 3 presents the experimental results along with a comparative analysis with a commercial reference system. Section 4 discusses the advantages, applicability, and limitations of the proposed approach

in real agricultural environments. Finally, the main contributions and future research directions are presented in Section 5.

2. Materials and Methods

2.1. Materials

2.1.1. Acquisition System

The acquisition system was designed to capture both color and depth information for the three-dimensional reconstruction of maize plants under real outdoor conditions. Its architecture is based on a binocular stereo camera module, IMX219-83 (Waveshare Electronics, Shenzhen, China), which integrates two Sony IMX219 image sensors, each with 8 megapixels. The sensors are mounted on a rigid support with a fixed baseline of 60 mm, enabling a balance between depth accuracy and field coverage suitable for medium-sized crop phenotyping. This camera was chosen for its compact design, low power consumption, and direct compatibility with embedded processing units such as the NVIDIA Jetson Nano, which facilitates real-time stereoscopic computation at the edge. Furthermore, the optical distortion of this camera is less than 1%, minimizing the need for extensive post-calibration correction.

On the other hand, the stereo module connects via dual CSI (Camera Serial Interface) ports using flexible flat cables (FFC) with ZIF connectors, ensuring the low-latency, high-bandwidth data transmission required for synchronous image acquisition. The main technical specifications of the camera are summarized in Table 1.

Table 1. Technical specifications and connection details of the IMX219-83 stereo camera.

Parameter	Value
Sensor	Sony IMX219
Resolution (per camera)	3280 × 2464 pixels (8 MP)
Sensor size (CMOS)	1/4 inch
Focal length	2.6 mm
Field of view (D/H/V)	83°/73°/50°
Distortion	<1%
Baseline length	60 mm
Dimensions	24 mm × 85 mm
Connection interface	Dual CSI (Camera Serial Interface) ports
Cable type	Flat Flexible Cable (FFC) with ZIF connectors

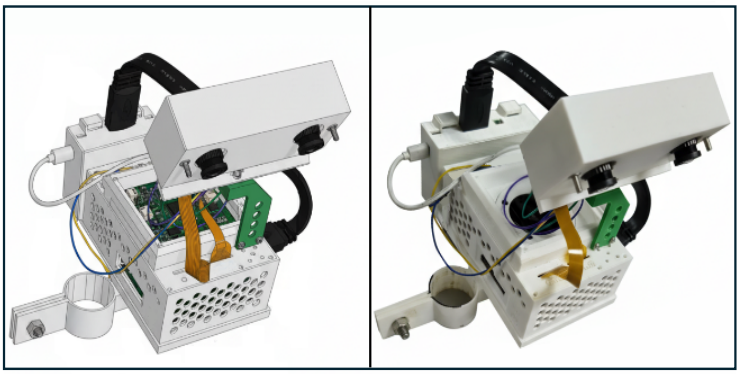
2.1.2. Embedded System

The NVIDIA® Jetson Nano (Developer Kit, NVIDIA Corporation, Santa Clara, CA, USA) was used as the integrated processing unit for managing image acquisition, synchronization, and edge computation. The IMX219-83 stereoscopic module was connected directly via the two CSI ports, allowing simultaneous capture of the left and right image streams without latency or frame misalignment. Image acquisition was performed at 30 FPS and a resolution of 1280 × 720. In addition, this device deploys Ubuntu 18.04 LTS with JetPack 4.6.4 (NVIDIA SDK), incorporating Python 3.8.0, OpenCV 4.12.0, and the Ultralytics YOLOv8 framework for detection tasks, along with CUDA 10.2 for GPU acceleration.

On the other hand, a 3.5-inch Waveshare® MPI3508 touchscreen was integrated for local interaction, allowing real-time visualization of acquisition parameters, frame synchronization status, and basic diagnostic metrics. The prototype was powered by an ADATA Power Bank PT100, providing autonomy during field measurements.

To ensure mechanical stability and consistent camera alignment, a custom 3D-printed housing was developed using PLA. The structure includes passive ventilation slots to prevent thermal throttling and mounting points for securing the device to tripods or fixed

supports, reducing vibrations during outdoor operation. This mechanical configuration was essential for enabling stable, reliable field measurements under variable environmental conditions. Figure 1 illustrates the assembly of the acquisition system, showing (a) the 3D model and (b) the assembled physical prototype, where the stereo module is positioned on the upper section of the frame and the touchscreen is mounted on the rear side to enhance operational usability.



(a) (b)

Figure 1. Assembly of the acquisition system: (**a**) 3D model and (**b**) physical prototype.

Table 2 summarizes the components that constitute the stereo acquisition system along with their respective costs. The value assigned to the structure was estimated based on the material usage reported by the 3D-printing software, which indicated that 226.83 g of PLA filament were required for its fabrication; this mass was then used to calculate its proportional cost according to the filament's price per kilogram. The remaining components are commercially available modules and accessories, whose listed prices were included in the overall system cost estimate.

Table 2. Component Costs of the Stereo Acquisition System.

Component	Price (USD)	Reference
Structure	3.68	[20]
Display	35.99	[21]
Jetson Nano Module	186.24	[22]
IMX219-83 Camera Module	47.50	[23]
Battery	25.00	[24]
Total	298.41	

The complete system was assembled at a total cost of USD 298.41, including both the stereo acquisition module and the on-board processing unit. This cost is significantly lower than that of commercial alternatives such as the ZED-2i (USD 519) or the Intel RealSense series (USD 272–419), which provide only the acquisition module without integrated processing capabilities. Moreover, several of these commercial systems require external high-performance hardware, such as GPU-equipped devices, to execute their processing pipelines. In contrast, the proposed system integrates acquisition and efficient processing within a single low-power platform, thereby reducing hardware requirements and overall system cost.

2.1.3. Stereo Calibration

Stereo calibration was performed using a printed checkerboard pattern with a 6-by-8 configuration of internal corners, where each square was 35 mm, and the regular arrangement provided a stable reference geometry for the system. The calibration procedure was

implemented using the OpenCV library, based on the method proposed by Zhang [25], which estimates the intrinsic and extrinsic camera parameters from multiple views of a flat pattern observed from different positions, without requiring prior knowledge of the three-dimensional geometry of the environment. During calibration, the pattern was placed at different positions and orientations within the camera field of view (FoV), spanning approximately 80 cm to 120 cm, as shown in Figure 2. Thus, we collected 11 image pairs for calibration. This variability allowed sufficient information to be captured, ensuring numerical stability and accuracy in the calculation of the calibration parameters.

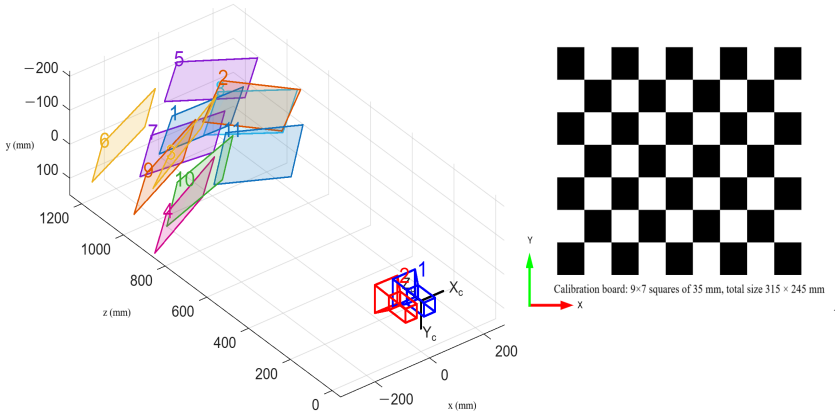


Figure 2. Spatial distribution of the chessboard pattern positions and orientations used during stereo calibration. Each marker (on the left) represents one of the captured views used to estimate the intrinsic and extrinsic parameters of both cameras (on the right). The inset shows the calibration chessboard pattern with an 8 × 6 grid of internal corners used as the geometric reference.

The intrinsic parameters of each camera (focal distance, principal point, and radial and tangential distortion coefficients) were estimated independently. In contrast, the extrinsic parameters (rotation matrix and translation vector) were calculated simultaneously to characterize the spatial relationship between the two sensors. The calibration results yielded a projection matrix and a rectification matrix that ensure the geometric alignment of the stereo pairs. From this data, the Q reprojection matrix was derived and used to reconstruct the scene in three dimensions.

The calibration quality was evaluated using the mean reprojection error per image, calculated for each camera. This metric allowed us to quantify the discrepancy between the positions of the corners detected in the images and the projections estimated by the calibration model, serving as an indicator of the calibration accuracy. The results obtained in the calibration process are illustrated in Figure 3, where a homogeneous distribution of errors with values between 0.3 and 0.5 pixels for both cameras can be observed, and an overall average error of 0.376 pixels. These results confirm that the deviations between the measured and projected points are less than the pixel size, ensuring accurate and stable reprojection. The slight variations observed between some image pairs can be attributed to local differences in corner detection or small mechanical disturbances during acquisition. Overall, the errors obtained confirm a stable geometric calibration that is suitable for three-dimensional reconstruction and metric analysis applications.

Finally, the calibration parameters were stored in .npz (NumPy compressed) files, allowing reuse and subsequent verification during the reconstruction and disparity calculation stages.

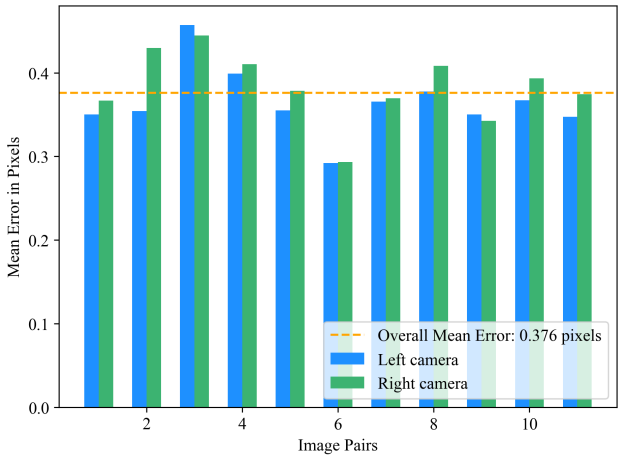


Figure 3. Average reprojection error per image for each camera during calibration.

2.1.4. Data Acquisition

Data acquisition was performed using the stereo system described above, installed at 97 cm above ground level. The target plant was located approximately 94 cm in front of the camera, ensuring complete coverage of the foliage within the field of view of both sensors. The geometric arrangement of the system and the plant is illustrated in Figure 4. During capture, the system was operated under natural lighting conditions and low to moderate wind. Although the influence of wind on 3D reconstruction has been documented, the literature indicates that establishing a direct correlation between wind speed and reconstruction errors is not straightforward [26]. Nevertheless, wind-induced leaf motion has been shown to introduce inaccuracies, typically resulting in errors below 3% in plant trait estimation [12]. For this reason, data were collected under low wind conditions to minimize motion-related artifacts in the disparity maps and ensure reliable height measurements. The data were stored in Portable Network Graphics (PNG) format to prevent information loss.

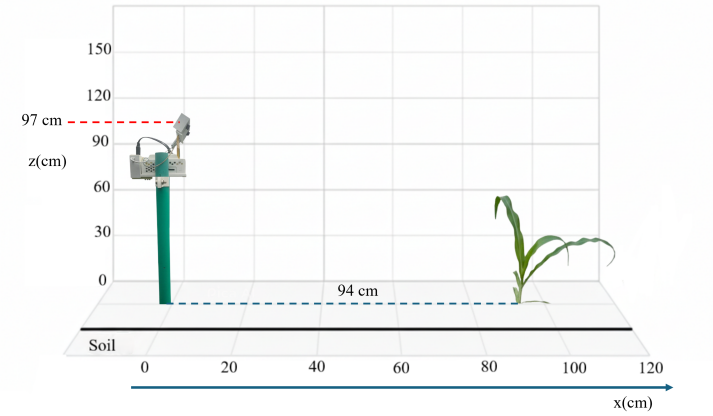


Figure 4. Spatial layout of the acquisition system: location of the stereo camera and the plant.

In order to evaluate the performance of the system and obtain three-dimensional data representative of phenotypic development, an experimental corn (*Zea mays* L.) plot was established in the open field. A regional creole variety of yellow corn from the department of Antioquia (Colombia) was used, grown under conventional management without fertilizers. The plantation consisted of six plants, distributed in two rows of three plants each. This number of plants was selected because the study focused on assessing the response and performance of the proposed stereo-vision system rather than on characterizing biological variability. In the early vegetative stages (V4–V7), maize

exhibits highly consistent structural traits across individuals. The separation between plants within each row was 75 cm, while the distance between adjacent rows was 120 cm, as shown in Figure 5.

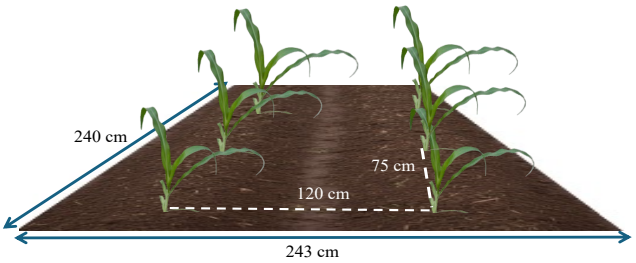


Figure 5. Layout of the maize plants in the experiment.

The measurements were taken over a period of approximately 3 weeks (18 days), with one measurement per plant per day. Additionally, on two random days (22 and 28 August 2025), images were acquired at two times of day, in the morning and in the afternoon, resulting in a total of 19 samples per plant across different vegetative development stages, from the V4 to V7 phases.

2.2. Methods

The three-dimensional reconstruction procedure was carried out through a sequence of steps aimed at transforming the pairs of captured stereo images into spatial representations. The general flow of the process is illustrated in Figure 6. First, a deep learning model was used to detect the ROI in the left image, corresponding to the area covered by the plant. Subsequently, this ROI was geometrically projected onto the right image using the intrinsic and extrinsic parameters obtained during system calibration, thereby ensuring spatial correspondence between the two views. Based on these regions, a stereo-matching algorithm was applied to estimate a disparity map, from which a three-dimensional point cloud describing the spatial geometry of the foliage was derived.

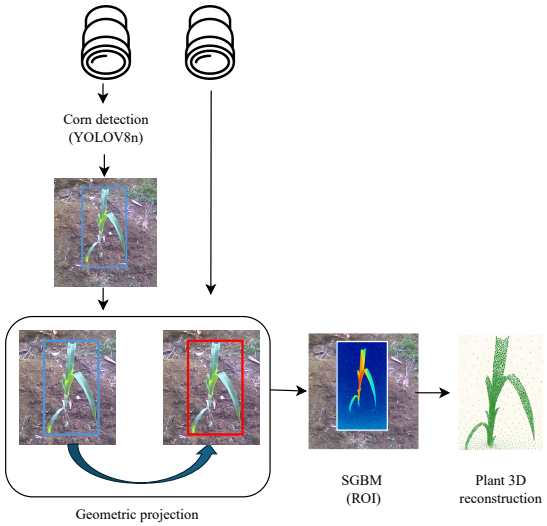


Figure 6. General processing pipeline for 3D reconstruction from stereo image pairs. The system integrates automatic plant detection, ROI projection, stereo matching, and 3D reprojection stages.

2.2.1. Automatic Detection of the Plant

For the model training stage, corn plants were manually annotated using the Computer Vision Annotation Tool (CVAT) [27], as illustrated in Figure 7. It is important to note that in

this part of the process, each plant was delimited by a bounding box covering the entire visible foliage region.



Figure 7. Manual annotation of a maize plant using the CVAT tool, showing the bounding box defining the region of interest.

The dataset comprises 544 RGB images acquired using the proposed stereo vision system. Although the system captures images in pairs, each image was treated independently for the 2D detection task. The captures were made at different times of the day in order to record the plants under various lighting conditions, including situations in which sunlight cast partial shadows on the foliage (see Figure 8). This variability allowed the model to generalize despite lighting changes typical of real agricultural environments. The acquisitions were made between 8:00 and 10:00 a.m. or between 4:00 and 5:00 p.m.



Figure 8. Examples of lighting variation during the acquisition period: (**a**) direct sunlight and (**b**) partial shadow conditions over the maize canopy.

Each plant was delimited by a polygon (bounding box) covering the entire area of the plant, as illustrated in Figure 7. The annotations are exported in YOLO format, generating a separate .txt file for each image. This text file contains the class label and the normalized bounding-box coordinates. To ensure reliable model evaluation, K-fold cross-validation (k = 5) was implemented. In this scheme, the total set of images was randomly divided into five equal-sized partitions. In each iteration, four subsets were used for training and one for validation, rotating systematically until all possible combinations were completed.

Training was performed using YOLOv8n [28], pre-trained with the COCO dataset [29]. Choice of the nano variant is consistent with recent studies on maize plant detection, which

highlight its optimal balance between accuracy and computational efficiency within the YOLOv8 family [19]. This deep learning model enables real-time instance detection and segmentation. In this study, it was used to automatically identify the regions corresponding to each corn plant in the dataset images.

The training process ran for 50 epochs with a learning rate of 0.001, a batch size of 16, and an input resolution of 640 × 640 pixels. The original images, with resolutions of 1280 × 720 and 640 × 360 pixels, were resized to the input size required by YOLOv8n to maintain dataset uniformity and ensure compatibility with the convolutional layers of the model. The pre-trained model was loaded from the local repository and tuned using the default Ultralytics optimizer [28].

All experiments were performed on a GNU Debian Linux system (kernel 4.19.98-1), with 2 × Intel Xeon Gold 6140, 128 GB RAM, and 4 × NVIDIA GPU GeForce 1080Ti. The results of each fold were stored in separate folders, allowing comparative analysis and quantification of performance variability across different subsets.

2.2.2. Disparity Calculation Within the Region of Interest

The Semi-Global Block Matching (SGBM) algorithm, implemented using the OpenCV library [30], was applied to the plant regions within the rectified stereo images. The stereo matcher was configured with a disparity search range of 96 levels (numDisparities = 96) and a correlation window of 5 pixels (blockSize = 5). To regularize disparity transitions between adjacent pixels, smoothness constraints were introduced through penalty terms P1 = 600 and P2 = 2400. Additional parameters were incorporated to improve the consistency and robustness of the disparity estimation, including a left–right consistency constraint (disp12MaxDiff = 1), a uniqueness ratio of 10, a speckle filtering module (20-pixel window and range of 2), and a pre-filtering cap value of 25. In full semi-global mode (STEREO_SGBM_MODE_HH), the algorithm aggregated matching costs along multiple epipolar paths, enabling the generation of a dense, spatially coherent disparity map (D). This map subsequently served as the foundation for the three-dimensional reconstruction of the plants and the generation of the corresponding point cloud.

Based on the disparity map D obtained, three-dimensional reconstruction is performed by transforming it into spatial coordinates (X, Y, Z) using the reprojection matrix Q through the OpenCV function cv2.reprojectImageTo3D(). This matrix, which encapsulates the epipolar geometry and the intrinsic and extrinsic parameters of the stereo system, enables disparity to be converted to depth. Conceptually, the depth Z is derived from the principle of triangulation, in which the disparity $d = d_2 - d_1$ (difference in position of a point between the left and right images) is inversely proportional to Z, modulated by the baseline x and the focal length f of the camera [31]. This relationship is illustrated in Figure 9, which shows how the variation in the position of a point between the two rectified views allows its depth coordinate to be reconstructed. The disparity information D obtained with SGBM is then projected using Q to generate the corresponding 3D point cloud, integrating the geometry and color of the detected ROI.

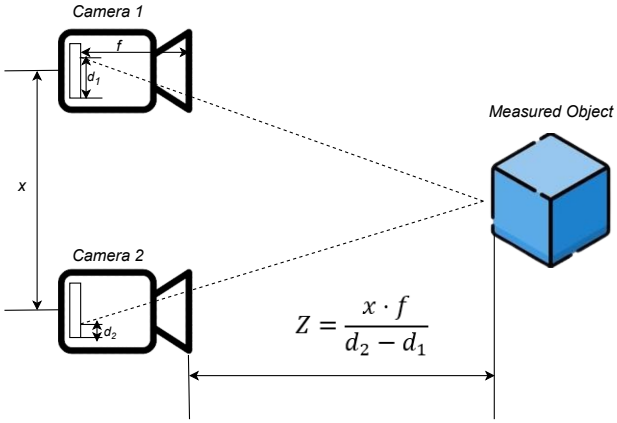


Figure 9. Schematic representation of the stereo vision principle. The disparity D between corresponding points in the left and right images is inversely proportional to the depth Z, according to the triangulation geometry defined by the focal length f and baseline b of the stereo camera.

2.2.3. Estimation of Height and Volume Based on 3D Reconstruction

The morphological parameters of the plant were calculated from the 3D reconstructed point cloud. Initially, the Random Sample Consensus (RANSAC) algorithm was applied to detect and fit a plane representing the ground beneath the plant. Due to local irregularities in the detected plane, a fixed offset of 2.5 cm was applied relative to the plane model to establish a consistent reference surface for height estimation. This reference was used uniformly for all plants, ensuring consistent height measurements despite variations in the ground surface. The points belonging to the ground plane were subsequently removed, retaining only those associated with the plant structure.

The plant height was determined by calculating the perpendicular distance from the top point of the point cloud to the adjusted ground plane, using the plane equation obtained by RANSAC. To reduce the influence of noise or outliers at the top of the cloud, the highest point was defined as the 95th percentile of the Z coordinates, representing the height below which 95% of the points lie. The distance was calculated perpendicular to the ground plane using its normal vector, ensuring the measurement is consistent with the actual orientation of the plane and independent of camera tilt. The plant volume was determined using a voxel-based spatial discretization approach. The three-dimensional space occupied by the point cloud was divided into a regular mesh of uniform-sized voxels, with each voxel containing at least one point from the cloud being considered occupied. The total volume was obtained by multiplying the number of occupied voxels by the volume of each voxel, thereby allowing the physical plant volume to be estimated from the spatial distribution of 3D points.

2.2.4. Performance Evaluation Methodology

A comparative evaluation of the processing time for the entire stereoscopic analysis workflow was conducted, comprising automatic plant detection using a YOLOv8 segmentation model, disparity estimation with StereoSGBM, and three-dimensional reprojection. The comparisons were performed under two experimental conditions: full image processing and combined processing, restricted to regions of interest detected by YOLO. The procedure was implemented in Python, using the OpenCV library for disparity estimation and 3D reprojection, while YOLO was used to identify and isolate corn plant areas in each stereo system image. To ensure reproducibility and reduce variability in measured times, each condition was evaluated 5 times per day, and the measurements presented correspond to Plant 4. For each measurement day, the processing times for the full image (resolution of 1280 × 720 px, equivalent to 921,600 pixels) and for the combined

YOLO + SGBM + reprojection method were recorded, along with the total number of pixels processed within the detected regions.

Before the comparative analysis between individuals, the consistency of the proposed system was evaluated under changes in lighting and variable environmental conditions on the same day. To this end, height measurements were taken at two times (morning and afternoon) with an approximate interval of 8 h on 22 and 28 August. This procedure allowed us to analyze the sensitivity of the system to variations in solar incidence, the presence of shadows, and potential environmental microfluctuations. Following the detection of an outlier corresponding to Plant 1 on 28 August, which was affected by leaf damage caused by rain and wind, this record was excluded from the analysis to avoid bias. Moreover, we analyzed the relationship between plant height and three-dimensional volume using Pearson's correlation coefficient to evaluate the physiological consistency of morphological metrics.

2.2.5. Comparative Experimental Protocol

A comparative experimental protocol was established for quantitative performance evaluation, using a reference system based on the commercial ZED 2i stereo camera (Stereolabs Inc., San Francisco, CA, USA) [32]. To ensure geometric equivalence between the two systems, the reference camera was positioned at the same distance, height, and orientation as the proposed system, thereby guaranteeing spatial correspondence throughout the acquisition process. Captures were performed consecutively at the same time of day to mitigate variations caused by changes in illumination or natural plant movement.

For each of the six evaluated plants, a three-dimensional point cloud (`.ply`) was generated and exported using the ZED 2i camera. From these clouds, manual measurements of the total plant height were taken within the MATLAB LiDAR Viewer environment [33], determining the vertical distance between the highest point of the plant structure and the ground, as illustrated in the Figure 10. Additionally, a cube of known dimensions (10 cm) was placed in the frontal view of the scene, serving as a spatial reference to validate the scale and metric consistency of the 3D reconstructions. This procedure was repeated during three consecutive experimental sessions (4–6 September 2025), which allowed the precision and consistency of the proposed system to be evaluated by calculating the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) with respect to the reference system.

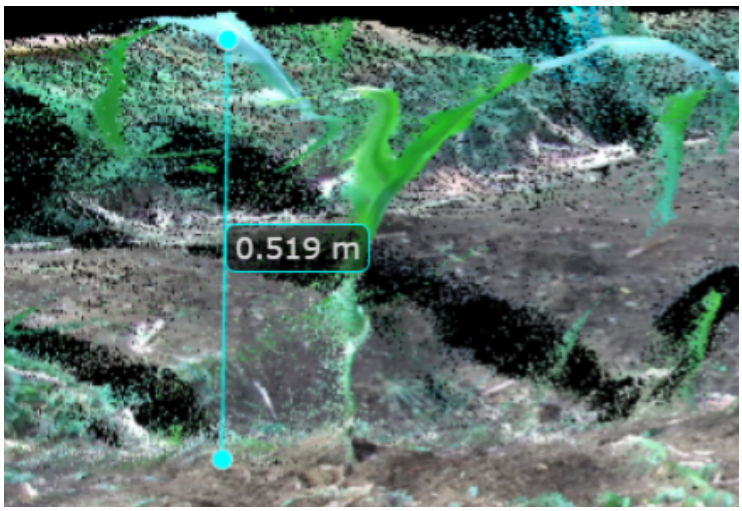


Figure 10. Manual height measurement process performed in MATLAB's LiDAR Viewer using the point cloud acquired with the ZED 2i camera.

3. Results

3.1. Performance of the Plant Detection Model

Figure 11 shows the evolution of the precision, recall, and mAP@0.5 metrics during the training of the YOLOv8n model, evaluated using five-fold cross-validation (k = 5). These metrics allow us to evaluate both the capacity of the model to correctly detect corn plants and its stability across different partitions of the dataset. During the first epochs (≈0–5), the metrics showed high dispersion, with average values below 0.5, reflecting the initial adjustment of the model weights. Starting in epoch 10, an increase was observed across all three metrics, reaching values close to 0.9, indicating a progressive improvement in the model's ability to discriminate the regions of interest correctly. At approximately epoch 20, precision, recall, and mAP@0.5 reached values close to 1.0 and remained virtually constant until epoch 50 was completed.

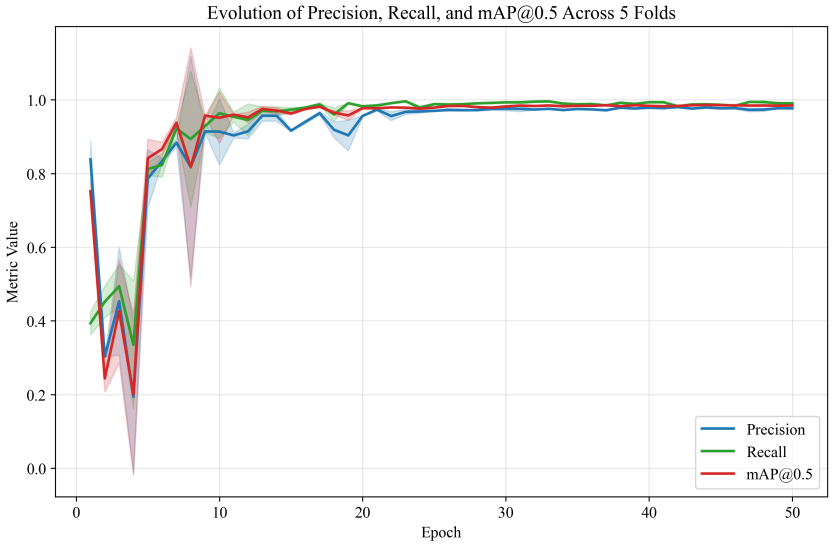


Figure 11. Evolution of the precision, recall, and mAP@0.5 metrics during the 50 training epochs of the YOLOv8n model. The shaded area represents the standard deviation across the five folds of cross-validation.

On the other hand, the gradual reduction in the shaded area (representing the standard deviation across the five folds) in the curves indicates low fold-to-fold variability, demonstrating stable and generalizable training. This behavior suggests that the morphological and textural characteristics of the plants were learned consistently, without noticeable overfitting to specific subsets of the dataset. The results above demonstrate the effectiveness and robustness of the YOLOv8n model in identifying corn plants under variable lighting conditions, including shadows and angular variations.

Figure 12 shows the evolution of processing time throughout the measurements. The blue line represents the performance of the complete image processing, which includes disparity estimation using SGBM and three-dimensional reprojection of the entire rectified image. In contrast, the orange line corresponds to the total combined processing time, which integrates automatic ROI detection using YOLOv8, stereoscopic disparity calculation, and 3D reprojection only within the detected regions.

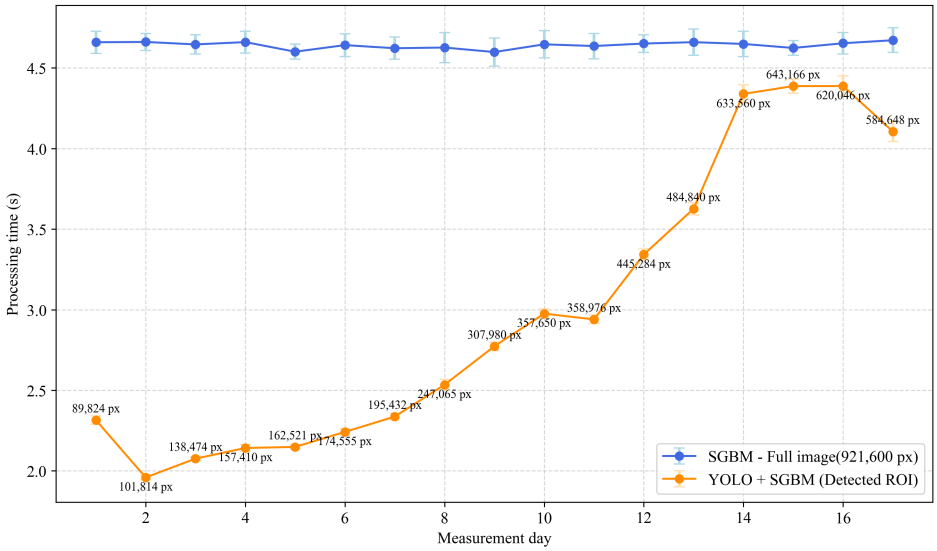


Figure 12. Daily processing time for stereoscopic flow. The blue line indicates complete processing, while the orange line shows the time required by the combined YOLO + SGBM + reprojection method for the detected ROIs.

During the first few days, when the plant coverage was low (approximately 90,000 to 200,000 pixels), the combined processing time was significantly reduced, reaching values close to 2 s per frame. As the plant grew and the number of detected pixels increased (exceeding 600,000 in the last few days), the execution time of the combined method gradually increased, approaching the time required to process the entire image. These results indicate that YOLO-based automatic detection significantly reduces computational time by restricting stereoscopic processing and 3D reprojection to regions of interest, while preserving spatial accuracy. The ROI was key to effectively isolate the plant, ensuring that morphological measurements were derived from the plant's structure while minimizing the influence of neighboring vegetation.

3.2. Analysis of Phenotyping Metrics

Figure 13 shows the temporal evolution of the height of six corn plants between 20 August and 6 September 2025, a period corresponding to the vegetative elongation phase. Each curve represents an individual plant, with daily measurements expressed in centimeters (cm) obtained from the point cloud reconstructed by detecting the ground plane using RANSAC and subsequently estimating the maximum vertical distance (Z_{max}) within each plant region. As shown in Figure 13, all plants exhibit a sustained increase in height throughout the monitoring period, demonstrating the system's ability to capture growth dynamics at daily resolution. However, the growth rate differs among the individuals analyzed during the experiment, which is attributable to both biological variability and microenvironmental effects (lighting, soil moisture, and leaf density). In particular, Plant 3 (green line) reached the highest recorded height, exceeding 60 cm, while Plant 6 (brown line) showed the least growth, remaining below 40 cm at the end of the observation period.

The negative fluctuations observed on specific dates do not correspond to measurement errors but rather to temporary structural alterations associated with leaf damage from rain and wind, which temporarily altered the aerial architecture and the vertical projection of the leaves. This behavior was confirmed by direct field visual inspection and reflects the sensitivity of the stereo vision system to record actual physical changes in plant morphology.

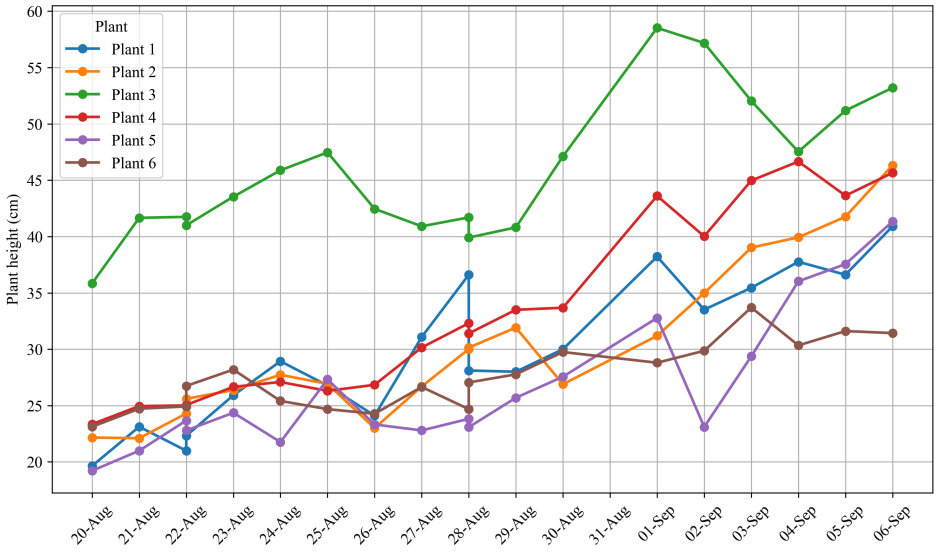


Figure 13. Temporal evolution of plant height for six maize individuals between 20 August and 6 September. Each curve represents the height extracted from the 3D point cloud of an individual plant.

Considering only valid measurements, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of 1.1 cm and 1.29 cm, respectively, were obtained, demonstrating high stability in the three-dimensional estimates. These differences, of the order of 1 cm, are consistent with the expected growth of plants over short intervals and can be attributed to both slight environmental fluctuations and actual biological growth during the observation period. In practical terms, these results demonstrate that the variation observed between consecutive measurements does not arise from the vision system but from environmental fluctuations or actual growth, confirming the robustness of the method to changes in natural lighting.

Figure 14 shows the temporal evolution of the estimated three-dimensional volume of the aerial part of the plants during the monitoring period. This parameter acts as a structural indicator of biomass, reflecting the three-dimensional expansion of the canopy. The results reveal a progressive increase in volume across all plants, with varying growth slopes among individuals. Plant 3 reached the maximum value, close to 430 cm^3, while Plant 5 showed the lowest volumetric development, less than 105 cm^3. The temporal trend reflects a pattern consistent with corn vegetative growth, characterized by rapid leaf expansion and a sustained increase in aerial volume. Non-invasive monitoring of this parameter using 3D reconstruction demonstrates the system's ability to quantify biomass dynamics with daily resolution.

Regarding the physiological consistency of morphological metrics, Pearson's correlation coefficient yielded a value of $r = 0.802$ between plant height and three-dimensional volume. These findings reveal a statistically significant and positive association ($p < 0.05$), demonstrating that proportional expansions in estimated biomass typically parallelled increases in plant height.

From a physiological perspective, this correspondence reflects a balance between stem elongation and leaf development, characteristic of vegetative growth of maize plants. However, the correlation was not perfect, suggesting differences in architecture between individuals. In some cases, growth was laterally oriented, with longer leaves or inclined stems, increasing volume without a significant change in height. This behavior is associated with morphophysiological responses to environmental factors, such as light direction, spatial competition, and mechanical stress caused by wind.

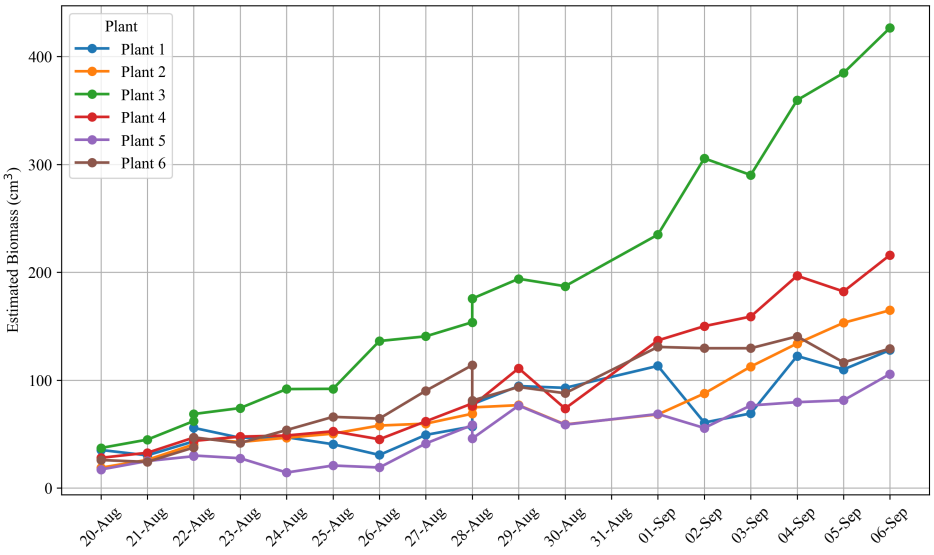


Figure 14. Temporal evolution of the estimated aerial biomass volume of maize plants. Each curve represents the voxel-based volumetric growth of an individual plant over the monitoring period.

3.3. Comparison Against a Commercial Reference System

Table 3 shows the MAE of the height measurements obtained with the proposed system compared to the ZED 2i camera. Three independent measurements were taken for each plant on 4–6 September. The MAE was calculated as the average absolute error (AE) between the measurements of the proposed system and those recorded with the ZED 2i camera, reflecting the average deviation from the reference system. The column "MAE per plant" indicates the average individual error for each plant, considering the three measurement dates. The results show that the proposed system reproduces height measurements with high fidelity compared to the ZED 2i, achieving an overall MAE of 1.48 cm, demonstrating its accuracy and consistency in phenotypic data acquisition.

Table 3. Mean Absolute Error (MAE) between height measurements obtained with the proposed system and the ZED 2i stereo camera. Each value represents the absolute difference for a single day, and the last column shows the MAE across the three measurement days (4–6 September 2025).

Plant	4 September	AE	5 September	AE	6 September	AE	MAE
1	37.76/37.5	0.26	36.61/39	2.39	40.91/40.17	0.74	1.13
2	39.93/40.2	0.27	41.76/40.5	1.26	46.33/43.2	3.13	1.55
3	47.53/49.5	1.97	51.18/51.2	0.02	53.21/51.9	1.31	1.10
4	46.66/47.4	0.74	43.63/45.05	1.42	45.66/45.2	0.46	0.87
5	36.03/38.7	2.67	37.55/41.6	4.05	41.34/42.3	0.96	2.56
6	30.34/33.1	2.76	31.61/33.7	2.09	31.43/31.2	0.23	1.69

Figure 15a presents the global correlation between the height measurements obtained with the proposed stereo-vision system and those recorded with the ZED 2i camera, pooling all 18 paired observations collected over three consecutive measurement days. Each point corresponds to a single plant on a given date, with ZED 2i measurements shown on the X-axis and those from the proposed system on the Y-axis. The resulting regression exhibits a strong linear relationship, with a coefficient of determination $R^2 = 0.931$, indicating that more than 93% of the variance in the proposed system's measurements is explained by the commercial reference device. The regression line lies close to the identity line ($y = x$), and the residuals remain limited across the entire height range, suggesting that the system maintains consistent accuracy for both shorter and taller individuals. The absence of any visible trend in the residual dispersion confirms the lack of proportional bias, while

the interspersed distribution of points across different days demonstrates stability under varying illumination and canopy configurations.

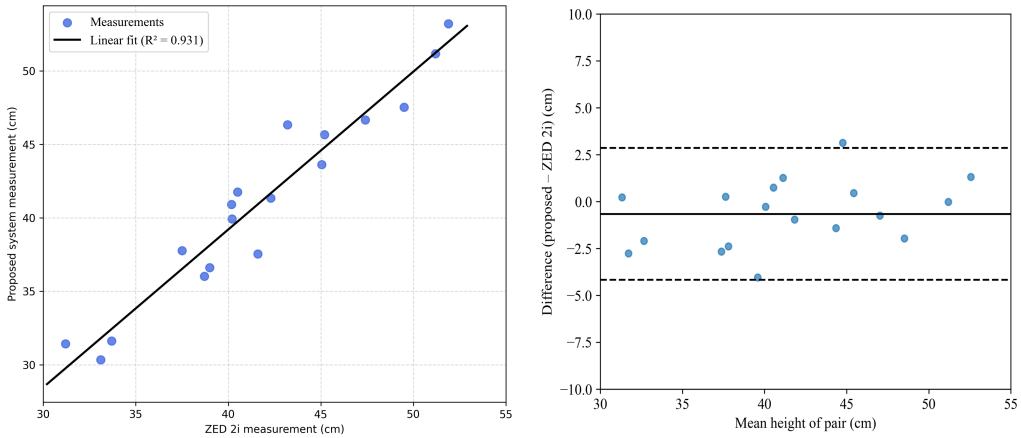


Figure 15. (**a**) Global correlation between the height measurements obtained with the proposed stereo vision system and the ZED2i camera across 4–6 September. Each point represents one plant on a given day. The solid line corresponds to the linear regression fit ($R^2 = 0.931$). (**b**) Bland–Altman plot comparing height measurements from the proposed stereo system and the ZED 2i camera ($n = 18$). The solid line indicates the mean bias (−0.66 cm), and the dashed lines show the limits of agreement (−4.18 cm and 2.85 cm).

To complement the correlation analysis, we quantified the agreement between the proposed system and the ZED 2i using several standard error metrics. The global MAE was 1.48 cm, and the RMSE was 1.87 cm, indicating that both average and larger deviations remained small. The MAPE of 3.67% shows that relative errors stayed below 4% across the full range of plant heights.

A Bland–Altman analysis was also performed to evaluate systematic effects. The mean bias was −0.66 cm (95% CI: [−1.56, 0.23] cm), and because the confidence interval includes zero, no significant systematic overestimation or underestimation is present. The limits of agreement ranged from −4.18 cm to 2.85 cm, with their respective confidence intervals contained within agronomically acceptable ranges. These limits indicate that 95% of the measurements fall within approximately ±3.5 cm of the reference device. The Bland–Altman plot (Figure 15b) further confirms an even distribution of differences across the height range, with no evidence of proportional bias.

4. Discussion

The obtained results demonstrate that the proposed stereo vision system is an efficient, accurate, and low-cost alternative for the three-dimensional phenotypic characterization of corn plants under real field conditions. The combination of automatic deep learning-based detection and SGBM enabled the automation of acquiring and analyzing morphological data, significantly reducing the manual intervention required in traditional photogrammetry or direct measurement methodologies.

The low average reprojection error (0.374 px) confirms robust geometric calibration, comparable to that reported for laboratory stereo systems [34]. This accuracy translates into a depth resolution of around 2–3 mm, validating the suitability of the system for applications involving the reconstruction of delicate plant structures. Likewise, the low residual vertical disparity and intercalibration stability demonstrate the mechanical rigidity of the assembly and the repeatability of the acquisition process, critical factors in field contexts where vibration and variable lighting often degrade the quality of stereo correspondence.

Additionally, the YOLOv8n model achieved accuracy, recall, and mAP values greater than 0.98, with a standard deviation across folds of less than 0.02, demonstrating stable convergence and strong generalization. These results are consistent with recent studies using YOLO architectures for crop detection and segmentation [19,35]. The observed interfold stability indicates that the model abstracts the distinctive morphological characteristics of corn across varying lighting and shade conditions, supporting its applicability in real agricultural settings.

On the other hand, reconstructing the point cloud using SGBM captured the plant geometry within the detected region of interest with high fidelity. The accuracy of the disparity map, along with subsequent filtering and reprojection using the Q matrix, enabled the generation of clean, dense three-dimensional clouds without significant loss of leaf information. The average height error (MAE $\approx$ 1.48 cm) is comparable to those reported in reference stereo systems such as the ZED 2i or RealSense D435i [12,36]. Moreover, the consistency of complementary agreement indicators—including RMSE, relative error, and the Bland–Altman analysis, which shows a small, non-significant bias and narrow limits of agreement—confirms that the reconstruction pipeline maintains stable performance across varying plant sizes and field conditions.

It is noteworthy that while both height and volume were extracted from the 3D reconstructions, direct validation was performed exclusively for height. Although we estimated the plants' height and volume, as described in Section 2.2.3, height estimation from 3D point clouds is inherently more objective and metrologically tractable than volume estimation. Particularly, plant height can be directly and unambiguously measured using conventional methods (rulers, lidar systems, or commercial stereo cameras like the ZED2i) as a reference measurement (ground-truth). In contrast, volumetric measurements of live plants are challenging and lack a universally accepted ground-truth methodology, and comparisons can be system- or algorithm-dependent. Besides, volume estimations depend on a comprehensive 3D reconstruction of the entire plant structure, which is influenced by accurate segmentation of all leaf surfaces and occlusions or overlapped regions. Therefore, the estimated biomass was included to corroborate that plant height does not follow a strictly linear pattern. Moreover, the positive correlation between height and volume (r = 0.802) confirms that increases in plant height typically parallel proportional increases in estimated biomass.

Finally, integrating automatic detection, stereo reconstruction, and morphological analysis on a low-power embedded platform opens new possibilities for real-time phenotyping and autonomous crop monitoring. In practical terms, the system can be scaled to mobile devices or terrestrial drones for continuous measurements without human intervention.

5. Conclusions

This study presented the design, implementation, and validation of a low-cost stereo vision system for 3D phenotyping of maize plants under real field conditions. The integration of deep-learning-based ROI detection (YOLOv8n) with SGBM enabled an automated, non-invasive workflow that captured structural plant traits with millimetric accuracy. The stereo calibration achieved sub-pixel reprojection errors, ensuring geometric stability and depth precision of approximately 2–3 mm within the working range. The detection model reached high performance with low cross-fold variability, demonstrating robust generalization to variations in illumination and shadow. Disparity-based 3D reconstruction yielded accurate height estimates (MAE = 1.1 cm; RMSE = 1.29 cm) and volumetric measurements strongly correlated with plant height, confirming the system's ability to quantify phenotypic growth dynamics reliably.

Compared with a commercial stereo camera, the proposed setup reproduced height measurements with high fidelity, confirming its metrological reliability and cost-effectiveness. A comprehensive method-comparison analysis was conducted in accordance with established metrological guidelines, and the global error metrics demonstrate that the system maintains centimeter-level accuracy across all measurement days. Likewise, the Bland–Altman analysis further validates its robustness, revealing a non-significant mean bias of −0.66 cm (95% CI: [−1.56, 0.23] cm) and narrow limits of agreement (−4.18 cm to 2.85 cm), with confidence intervals fully contained within agronomically acceptable ranges. These findings confirm that the system introduces neither systematic nor proportional error and that it performs consistently across the entire measured height range.

The obtained results position the proposed system as a viable alternative for low-cost field phenotyping, particularly in research or precision-agriculture contexts where portability, autonomy, and affordability are essential. The inclusion of a statistically rigorous agreement analysis significantly strengthens the system's validation and demonstrates its potential for reliable deployment in real-world agricultural conditions.

Finally, future work will focus on expanding the dataset to include different crop species and growth stages, enhancing segmentation accuracy through semantic and self-supervised learning, and integrating multispectral and thermal imaging for joint structural–physiological analysis. Additionally, deploying the platform on mobile or robotic units will enable large-scale, real-time phenotyping aligned with the vision of Agro 4.0 and smart farming.

Author Contributions: Conceptualization, J.Z.-L., J.B.-V. and E.R.-V.; methodology, J.Z.-L., J.B.-V. and E.R.-V.; software, J.Z.-L. and Í.A.T.; validation and formal analysis, J.Z.-L., J.B.-V., E.R.-V. and R.H.-G.; investigation, J.Z.-L. and Í.A.T.; resources, J.B.-V., E.R.-V. and R.H.-G.; data curation, J.Z.-L. and Í.A.T.; writing—original draft preparation, J.Z.-L., J.B.-V., Í.A.T. and E.R.-V.; writing—review and editing, J.B.-V., E.R.-V. and R.H.-G.; visualization, J.Z.-L. and Í.A.T.; supervision, J.B.-V. and E.R.-V.; project administration, J.B.-V. and E.R.-V.; funding acquisition, J.B.-V. and E.R.-V. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been carried out under contract RC130-2024, corresponding to project SIGP code 109922, co-financed by MINCIENCIAS, titled "Diversificación de fuentes de proteínas para uso alimentario mediante el empleo de terrazas de cultivo aeropónicas o hidropónicas, integradas con sistemas automatizados, inteligencia artificial y energía renovable para la creación de comunidades autosostenibles", led by the "Grupo de Sistemas de Control y Robótica COL0123701" of the "Instituto Tecnológico Metropolitano de Medellín".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in OSF at https://doi.org/10.17605/OSF.IO/MN6P9.

Acknowledgments: This work has been carried out under contract RC130-2024, corresponding to project SIGP code 109922, co-financed by MINCIENCIAS, titled "Diversificación de fuentes de proteínas para uso alimentario mediante el empleo de terrazas de cultivo aeropónicas o hidropónicas, integradas con sistemas automatizados, inteligencia artificial y energía renovable para la creación de comunidades autosostenibles", led by the "Grupo de Sistemas de Control y Robótica COL0123701" of the "Instituto Tecnológico Metropolitano de Medellín." Ítalo A. Torres acknowledges the funding from ANID–Subdirección de Capital Humano/Doctorado Nacional/2024–21241873.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the study's design, the collection, analysis, or interpretation of data, the writing of the manuscript, or the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CVAT	Computer Vision Annotation Tool
MAE	Mean Absolute Error
RANSAC	Random Sample Consensus
ROI	Region of interest
RMSE	Root Mean Square Error
SGBM	Semi-Global Stereo Matching
3D	Three-dimensional

References

1. Zapata-Londoño, J.; Botero-Valencia, J.; García-Pineda, V.; Reyes-Vera, E.; Hernández-García, R. A Comprehensive Review of Optical and AI-Based Approaches for Plant Growth Assessment. *Agronomy* **2025**, *15*, 1781. [CrossRef]
2. Gang, M.S.; Kim, H.J.; Kim, D.W. Estimation of greenhouse lettuce growth indices based on a two-stage CNN using RGB-D images. *Sensors* **2022**, *22*, 5499. [CrossRef]
3. Kim, W.S.; Lee, D.H.; Kim, Y.J.; Kim, T.; Lee, W.S.; Choi, C.H. Stereo-vision-based crop height estimation for agricultural robots. *Comput. Electron. Agric.* **2021**, *181*, 105937. [CrossRef]
4. Harandi, N.; Vandenberghe, B.; Vankerschaver, J.; Depuydt, S.; Van Messem, A. How to make sense of 3D representations for plant phenotyping: A compendium of processing and analysis techniques. *Plant Methods* **2023**, *19*, 60. [CrossRef]
5. Akhtar, M.S.; Zafar, Z.; Nawaz, R.; Fraz, M.M. Unlocking plant secrets: A systematic review of 3D imaging in plant phenotyping techniques. *Comput. Electron. Agric.* **2024**, *222*, 109033. [CrossRef]
6. Xiang, L.; Wang, D. A review of three-dimensional vision techniques in food and agriculture applications. *Smart Agric. Technol.* **2023**, *5*, 100259. [CrossRef]
7. Niknejad, N.; Bidese-Puhl, R.; Bao, Y.; Payn, K.G.; Zheng, J. Phenotyping of architecture traits of loblolly pine trees using stereo machine vision and deep learning: Stem diameter, branch angle, and branch diameter. *Comput. Electron. Agric.* **2023**, *211*, 107999. [CrossRef]
8. Bortolotti, G.; Piani, M.; Gullino, M.; Mengoli, D.; Franceschini, C.; Grappadelli, L.C.; Manfrini, L. A computer vision system for apple fruit sizing by means of low-cost depth camera and neural network application. *Precis. Agric.* **2024**, *25*, 2740–2757. [CrossRef]
9. Zhao, Y.; Zhang, X.; Sun, J.; Yu, T.; Cai, Z.; Zhang, Z.; Mao, H. Low-cost lettuce height measurement based on depth vision and lightweight instance segmentation model. *Agriculture* **2024**, *14*, 1596. [CrossRef]
10. Zhang, L.; Shi, S.; Zain, M.; Sun, B.; Han, D.; Sun, C. Evaluation of Rapeseed Leave Segmentation Accuracy Using Binocular Stereo Vision 3D Point Clouds. *Agronomy* **2025**, *15*, 245. [CrossRef]
11. Kobe, M.; Elias, M.; Merbach, I.; Schädler, M.; Bumberger, J.; Pause, M.; Mollenhauer, H. Automated workflow for high-resolution 4D vegetation monitoring using stereo vision. *Remote Sens.* **2024**, *16*, 541. [CrossRef]
12. Dandrifosse, S.; Bouvry, A.; Leemans, V.; Dumont, B.; Mercatoris, B. Imaging wheat canopy through stereo vision: Overcoming the challenges of the laboratory to field transition for morphological features extraction. *Front. Plant Sci.* **2020**, *11*, 96. [CrossRef]
13. Sampaio, G.S.; Silva, L.A.; Marengoni, M. 3D reconstruction of non-rigid plants and sensor data fusion for agriculture phenotyping. *Sensors* **2021**, *21*, 4115. [CrossRef]
14. Wen, J.; Yin, Y.; Zhang, Y.; Pan, Z.; Fan, Y. Detection of wheat lodging by binocular cameras during harvesting operation. *Agriculture* **2022**, *13*, 120. [CrossRef]
15. Liu, X.; Tian, J.; Kuang, H.; Ma, X. A stereo calibration method of multi-camera based on circular calibration board. *Electronics* **2022**, *11*, 627. [CrossRef]
16. Li, D.; Xu, L.; Tang, X.s.; Sun, S.; Cai, X.; Zhang, P. 3D imaging of greenhouse plants with an inexpensive binocular stereo vision system. *Remote Sens.* **2017**, *9*, 508. [CrossRef]
17. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 328–341. [CrossRef] [PubMed]
18. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *23*, 1222–1239. [CrossRef]
19. Chu, T.; Zha, H.; Wang, Y.; Yao, Z.; Wang, X.; Wu, C.; Liao, J. MaizeStar-YOLO: Precise Detection and Localization of Seedling-Stage Maize. *Agronomy* **2025**, *15*, 1788. [CrossRef]
20. Overture. PLA Filament 1.75 mm. Available online: https://www.amazon.com/Overture-Filamento-impresi (accessed on 9 December 2025).

21. Waveshare. 3.5 inch HDMI LCD. 2025. Available online: https://www.waveshare.com/3.5inch-hdmi-lcd.htm (accessed on 9 December 2025).
22. Corporation, N. Jetson Nano Developer Kit. 2025. Available online: https://www.digikey.com/en/products/detail/seeed-technology-co-ltd/102110417/11560461 (accessed on 9 December 2025).
23. Waveshare. 3D Stereo Camera IMX219 8MP 83°. 2025. Available online: https://www.digikey.com/en/products/detail/seeed-technology-co-ltd/114992270/12396915 (accessed on 9 December 2025).
24. ADATA. ADATA C100 10000 mAh Power Bank. 2025. Available online: https://www.adata.com/co/consumer/category/power-banks/power-bank-c100/ (accessed on 9 December 2025).
25. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]
26. Andújar, D.; Dorado, J.; Bengochea-Guevara, J.M.; Conesa-Muñoz, J.; Fernández-Quintanilla, C.; Ribeiro, A. Influence of Wind Speed on RGB-D Images in Tree Plantations. *Sensors* **2017**, *17*, 914. [CrossRef] [PubMed]
27. CVAT.ai Contributors. Computer Vision Annotation Tool (Online Version). 2025. Available online: https://app.cvat.ai/ (accessed on 25 September 2025).
28. Ultralytics. YOLOv8n: Nano Object Detection Model. Version 8.3.204. 2025. Available online: https://github.com/ultralytics/ultralytics (accessed on 30 September 2025).
29. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755. [CrossRef]
30. Foundation, O.S.V. OpenCV: Open Source Computer Vision Library. Version 4.12. 2025. Available online: https://github.com/opencv/opencv (accessed on 30 September 2025).
31. Vázquez-Arellano, M.; Griepentrog, H.W.; Reiser, D.; Paraforos, D.S. 3-D Imaging Systems for Agricultural Applications—A Review. *Sensors* **2016**, *16*, 618. [CrossRef]
32. Stereolabs. *ZED 2i Stereo Camera—Datasheet (Rev. 1.2)*; Stereolabs Inc.: San Francisco, CA, USA, 2022.
33. The MathWorks, Inc. Lidar Viewer (Lidar Toolbox)—MATLAB Online. Available online: https://www.mathworks.com/help/lidar/ref/lidarviewer-app.html (accessed on 1 November 2025).
34. Jia, Z.; Yang, J.; Liu, W.; Wang, F.; Liu, Y.; Wang, L.; Fan, C.; Zhao, K. Improved camera calibration method based on perpendicularity compensation for binocular stereo vision measurement system. *Opt. Express* **2015**, *23*, 15205–15223. [CrossRef]
35. Lu, C.; Nnadozie, E.; Camenzind, M.P.; Hu, Y.; Yu, K. Maize plant detection using UAV-based RGB imaging and YOLOv5. *Front. Plant Sci.* **2024**, *14*, 1274813. [CrossRef] [PubMed]
36. Song, P.; Li, Z.; Yang, M.; Shao, Y.; Pu, Z.; Yang, W.; Zhai, R. Dynamic detection of three-dimensional crop phenotypes based on a consumer-grade RGB-D camera. *Front. Plant Sci.* **2023**, *14*, 1097725. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.