Check for updates

"Canopy fingerprints" for characterizing three-dimensional point cloud data of soybean canopies

Therin J. Young[1†], Talukder Z. Jubery[2†], Clayton N. Carley[3], Matthew Carroll[3], Soumik Sarkar[1,2], Asheesh K. Singh[3], Arti Singh[3*] and Baskar Ganapathysubramanian[1,2*]

[1]Department of Mechanical Engineering, Iowa State University, Ames, IA, United States, [2]Translational AI Center, Iowa State University, Ames, IA, United States, [3]Department of Agronomy, Iowa State University, Ames, IA, United States

Advances in imaging hardware allow high throughput capture of the detailed three-dimensional (3D) structure of plant canopies. The point cloud data is typically post-processed to extract coarse-scale geometric features (like volume, surface area, height, etc.) for downstream analysis. We extend feature extraction from 3D point cloud data to various additional features, which we denote as 'canopy fingerprints'. This is motivated by the successful application of the fingerprint concept for molecular fingerprints in chemistry applications and acoustic fingerprints in sound engineering applications. We developed an end-to-end pipeline to generate canopy fingerprints of a three-dimensional point cloud of soybean [*Glycine max* (L.) Merr.] canopies grown in hill plots captured by a terrestrial laser scanner (TLS). The pipeline includes noise removal, registration, and plot extraction, followed by the canopy fingerprint generation. The canopy fingerprints are generated by splitting the data into multiple sub-canopy scale components and extracting sub-canopy scale geometric features. The generated canopy fingerprints are interpretable and can assist in identifying patterns in a database of canopies, querying similar canopies, or identifying canopies with a certain shape. The framework can be extended to other modalities (for instance, hyperspectral point clouds) and tuned to find the most informative fingerprint representation for downstream tasks. These canopy fingerprints can aid in the utilization of canopy traits at previously unutilized scales, and therefore have applications in plant breeding and resilient crop production.

KEYWORDS

fingerprints, terrestrial laser scanner, 3D point cloud, phenotypic fingerprint, soybean

1 Introduction

Soybean [*Glycine max* (L.) Merr.] canopy characteristics indicate crop growth, development, and health among other characteristics. Canopy traits have traditionally focused on 2-dimensional (2D) features, which is useful in certain instances, for example, canopy coverage (Purcell, 2000), which has frequently been collected with drone high throughput phenotyping (Guo et al., 2021). With the advent of high-throughput crop and plant phenotyping (Araus and Cairns, 2014; Yang et al., 2020; Guo et al., 2021; Jubery et al., 2021; Singh A. K. et al., 2021; Singh D. P. et al., 2021), plant scientists have been able to conduct large scale and time-series investigations on canopy coverage. Additionally, researchers have shown automated or semi-automated extraction of canopy traits; for example, canopy features, including height, shape, color, and texture, can be used for plant stress and disease assessment, estimating total biomass, leaf chlorophyll, and leaf nitrogen (Shiraiwa and Sinclair, 1993; Hunt et al., 2005; Pydipati et al, 2006; Jubery et al., 2016; Bai et al., 2018; Parmley et al., 2019; Parmley et al., 2019). Canopy morphology features, such as shape and size, impact light interception ability, which directly factors into the potential yield equation (Metz et al, 1984; Koester et al., 2014). Canopy characteristics, including height, shape, size, and color, can vary among developmental stages, genotypes, and environments (Virdi et al., 2021). Quantifying the canopy plasticity of a genotype due to changing environmental conditions and variability or similarity among genotypes is valuable for plant breeding applications (Sadras and Slafer, 2012). However, a major hurdle towards effective and full utilization of canopy features is the relatively slow pace of advancement of three-dimensional (3D) canopy features, which provide a "real-world" set of information.

Historically, digital cameras, hyperspectral cameras, and LIDAR have been used to take images and create point clouds of plants (Walter et al., 2019; Herrero-Huerta et al., 2020; Chiozza et al., 2021a; Chang et al., 2022) which are then used to characterize plant traits, leading to a composite plant canopy. Often, methods such as structure from motion or tomographic reconstruction methods are needed to render the 3D point clouds for these traits (Vandenberghe et al, 2018; Storey, 2020). Widely utilized characterization approaches are based on hand-crafted geometric measures, such as plant height, length, breadth, height, area, and volume. Although these geometric features are simple to interpret, they often do not comprehensively represent the spatial variability, for instance, between sample heights and the intricacy of the canopies. Several studies used latent feature representation methods, such as Principal component analysis (PCA) and Neural Network (NN), to characterize the canopy (Gage et al., 2019; Ubbens et al., 2020). Although these features can be used to capture the inherent complexity of the canopies, they are challenging to comprehend since they are difficult to relate to real geometry with low interpretability. There is interest in developing more detailed yet interpretable phenotypic traits for characterizing the crop canopy. Interpretable features are crucial to develop field-testable hypotheses for plant scientists. Most interpretable approaches concentrate on composite characteristics and do not account for individual trait

variations. An example approach that offers a middle ground between these two extremes is the elliptical Fourier transformation utilized to describe the complicated geometry of canopy structures (Jubery et al., 2016). However, the use of 3D point clouds can be more exhaustive and informative, motivating researchers to develop holistic phenotyping pipelines (end to end) as well as explore applications of the usage of information from these data. For example, 3D canopy generation has been successfully shown in wheat, *Triticum aestivum* (Paulus et al., 2013; Paulus et al., 2014), rice, *Oryza sativa* (Burgess et al., 2017; Zhu et al., 2018), and other crops (Vandenberghe et al 2018). These are exciting developments; however, there is still information lacunae on the creation of informative multiscale traits from 3D point cloud data. In this context, non-agricultural disciplines have reported a concept of fingerprinting using point cloud data (Koutsoukas et al., 2014; Spannaus et al., 2021; Wang et al., 2021), but this is lacking in crop production and broader agriculture.

Fingerprinting is a technique for the multiscale characterization of an object by computing a set of unique local characteristics or patterns. Fingerprinting successfully generates unique representations for complex objects in chemistry, geometry, and acoustics (Cano et al., 2005; Capecchi et al, 2020). It was successfully used for the retrieval, recognition, and matching tasks within large molecular and 3D shape databases (Fontaine et al., 2007). Fingerprinting facilitates the representation of a complicated, memory-intensive 3D point cloud as a hierarchically computed, low-dimensional vector. This vector captures both the geometric and topological characteristics of 3D shapes. Computational approaches to computing fingerprinting for 3D objects are broadly based on spectral and non-spectral methods. Spectral approaches utilize the eigenvectors and eigenvalues, referred to as the spectrum, of the Laplace-Beltrami (LB) operator applied to 3D shapes (Reuter et al., 2005). The spectrum is independent of the object's representation, including the parameterization method and spatial position. Other techniques were developed from LB, for example, Shape-DNA and Global Point Signature (Wu et al, 2022). Probabilistic fingerprinting (Mitra et al., 2006) is a non-spectral fingerprinting technique. It is suitable for determining partial matching between 3D objects. Here, the objects were separated into overlapping patches, unique descriptors were generated for each patch, the descriptors were hashed, and a random subset of the hashed descriptors with a predetermined vector size was chosen as the probabilistic fingerprint. Similar min-hashing techniques (random subset selections) were used to get structural similarity in larger data based on chemistry (Probst, 2018). Hashing aids in the compression of the fingerprinting representation, but this cannot be decoded and is less interpretable. There are application examples of the fingerprinting concept; for example, a phenotypic fingerprint of a soybean canopy was proposed to represent the temporal variation of coarse-scale geometric features, including canopy height and plant length (Zhu et al., 2020b), and it was employed to capture temporal dynamics, identify genotypes with comparable growth signatures, etc. Further, canopy fingerprints enabled large-scale evaluation of the environmental constraints and disturbances that shape the 3D structure of forest canopies (Jucker, 2022). However, thus far, there is no work to define and develop crop canopy fingerprints capturing

multi-scale geometric features that could be evaluated and applied in the future in crop modeling, and genomic prediction (Jarquin et al, 2016; Shook et al, 2021a; Shook et al., 2021b), or breeding decisions. Fingerprints are distinctly unique from traditional canopy architecture as they encompass the entire global canopy, while architecture traits are a composite of limited individual traits assessed together.

The major contribution of this work is to develop an end-to-end non-spectral interpretable fingerprint generation pipeline for 3D point cloud data of field-grown row crops (Figure 1). The pipeline includes point cloud noise removal, registration, plot extraction, and fingerprint generation. We illustrate this approach using a large-scale field experiment through a diversity panel of soybeans. Specifically, we report the construction of canopy fingerprints in soybean using geometric and topological features of the 3D point cloud obtained by a Terrestrial laser scanner (TLS). This is accomplished with an end-to-end pipeline to generate canopy fingerprints of a three-dimensional point cloud of soybean, which is simple to use for feature extraction and utilization in a myriad of applications, including modeling, genomic prediction, ideotype breeding, and cultivar development. For example, the development of unique canopy fingerprints could enable faster and more efficient screening of genetic material for identifying canopy relationships with yield traits (Liu et al, 2016), biotic stress traits such as disease and insects (Pangga et al, 2011), how various canopy levels impact planting density, light interception, and photosynthesis (Feng et al., 2016), enable novel meta-GWAS (Shook et al., 2021a) or improve how crop modeling could predict the ideal canopy fingerprint (Rötter et al., 2015), or fingerprint ideotype, which could then be screened across core collections (Glaszmann et al., 2010) to narrow the pool of experimental genotypes *in silico* prior to *in vivo* evaluation.

2 Materials and methods

2.1 Laser scanner

The TLS used in this study was Trimble TX5 (Trimble Inc., Sunnyvale, CA, USA) (Figure 1A). It is a small and light device (240 mm x 200 mm x 100 mm in size and 5 kg in weight) that can

perform measurements at speeds of 1 million points per second. The scanner collects data at a high angular resolution of 0.011 degrees, corresponding to a point spacing of 2 mm at a 10 m scanning range. The scanner emits a 3mm diameter and 905 nm wavelength laser beam and measures the distance between the scanner and the target using the phase-shift principle (Amann et al., 2001). The emitted laser beam is modulated at several frequencies, and the phase shift of all the returned modulations is assessed to increase the accuracy of the distance measurements while storing the intensity of the returned beam. The scanner covers a 360-degree x 300-degree field of view: 360 degrees on the vertical axis is achieved by rotating the scanner head, and a rotating mirror achieves 300 degrees on the horizontal axis. The scanner allows the acquisition of point clouds of 7.1 up to 710.7 million points (MP). The number of points corresponds to the resolution of the measurement. Additionally, the scanner has a built-in camera to capture RGB color values (up to 70 megapixels) and maps them to the corresponding point clouds.

2.2 Location, plant materials, and data collection

The experiment was done in a field frequently used for evaluating soybean iron deficiency chlorosis at Iowa State University's Agricultural Engineering/Agronomy Research Farm, IA, USA, at a latitude of 42.010 and a longitude of -93.735. Four hundred sixty-four soybean cultivars were included in this study. These accessions come from 35 countries and have crop maturities ranging from MG 0 to IV (Mourtzinis and Conley, 2017), along with variable seed weights and stem termination types. In May 2018, the cultivars were hand-planted in hill plots, three seeds per hill with 0.76 m spacing between each hill. Each plot consisted of a single hill replicated three times in the field, with each replication blocked together. No plants were thinned. Preparing a noise-free field is crucial for achieving accurate plant data. To minimize any interference from weeds, we conducted regular weeding at intervals throughout our study. Laser scanning was performed on all plots on the 9th and 10th of August. These scans were conducted 71 and 72 days after planting, and all plants had reached at least the
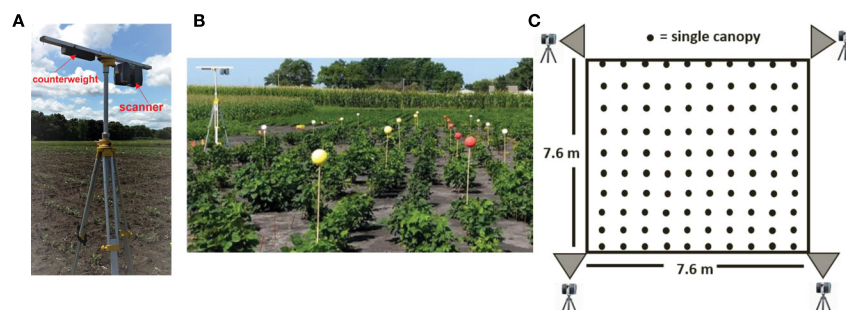

FIGURE 1
(A) Scanning platform: the scanner was mounted on a tripod in an inverted position with an extended bar and counterweight. **(B)** Placement of colored reference markers along the blocks. **(C)** Schematic of the scanning positions, block size, and canopy count per block.

reproductive R3 stage (Fehr et al., 1971) and were entirely opaque from the side, with no leaves visible from the opposite side.

The approximate size of the scanned field was 0.144 hectares (0.355 acres). During scanning, the field was divided into twenty-five 7.6 m x 7.6 m blocks, each containing 100 plots. The scanner was mounted on an 8 kg heavy-duty elevated tripod (Johnson Level, USA) at 2.1 m. For this height, the scanner can see the ground around the base of the farthest canopy within a block. This resulted in the typical scanner to ground distances between 2.1 to 11 m within the block. The device's scanning resolution was set at 0.5 (angular resolution 0.016 degree), and the expected point distance was 0.6 mm at 2.1 m and 3.1 mm at 11 m from the scanner.

To compensate for the scanner's field of view restriction of 150 degrees relative to the nadir, or lowest point under the observation lens, the scanner was mounted upside down using a 1.2 m-long bar, as illustrated in Figure 1A. This configuration allowed us to scan plots close to the tripod and position the scanner at the edge of each block. Scanning data was captured from four corners for each of the blocks: southwest (SW), southeast (SE), northeast (NE), and northwest (NW). The horizontal rotation limit of the laser scanner was set to 180 degrees, allowing two blocks to be scanned at once.

Before performing scans, Styrofoam spherical targets with a diameter of 0.127 m were placed within each block as reference markers to aid point cloud registration (Figure 1B), alignment, and plot identification. The spheres were painted yellow or red and mounted on 1.52 m-tall wooden dowels, which are 0.5 m taller than the expected maximum plant height. The dowels were manually pushed into the soil about 0.15 m deep. Each plot contained six reference markers. A white reference marker was placed at each corner of the block. The position of the reference markers was consistent across all scanned blocks.

Field experiments showed wind speeds to be lowest during the morning hours up until the early afternoon hours. When wind speeds exceeded 14.5 km h^{-1}, canopy movement exceeded the uncertainty acceptable for trait measurement. As a result, scanning took place between 9 a.m. and 2 p.m., or when wind speeds were 14.5 km h^{-1} or lower, to ensure the point cloud's quality was not compromised. While the optimal lighting condition for scanning is at noon, when sunlight is evenly distributed across the scanning area, field experiments demonstrated that overcast lighting also resulted in high-quality point cloud data. We

avoided operating the scanner in the early mornings or late afternoons when direct or bright sunlight reflected from plant materials would cause laser signal saturation, resulting in erroneous points synonymous with glare in 2D photography.

Validation data consisting of plant height and canopy area were collected on August 8th. Plant height was recorded as the distance between the soil line at the base of the stem and the topmost leaf. The canopy area was defined as the visible area of the canopy from the nadir described in detail below.

On each plot, plant height was measured manually with a meter stick. The canopy area was measured on a subset of the plots as follows: First, a digital camera (Zenmuse X5 camera with a lens focal length of 45 mm)) mounted on a drone (Matrice 600 Pro) captured RGB images of the plots flown at 30m with 80% overlap and stitched together using Pix 4-D stitching software. We used an in-house Python script to extract individual plots from the stitched orthomosaic image, using the geolocation data obtained from the ground control points (GCP) and the RTK GPS mounted on the UAV. Next, we converted the images from RGB to the HSV color space, and the canopy was separated from the ground by applying a threshold to the Hue (H) color channel. We experimented with different threshold values for the Hue channel and found that the hue value worked best for our case. The canopy area was then calculated by determining the total number of non-zero (canopy) pixels and converting this value to m^2. To obtain the conversion factor from pixels to m^2, we measured a predefined marker in the images.

2.3 Point cloud processing pipeline

The pipeline was built using Python 3.7.3 and various other programs and packages, including Autodesk Recap 4.2.2.15, Cloud Compare 2.12.4 (Girardeau-Montaut n.d.), Open3D 0.11.2 (Zhou et al., 2018), and MATLAB 2019a. MATLAB and Cloud Compare were used *via* the command line interface and incorporated into the Python script *via* the Python subprocess library. The specific tasks carried out by these packages are depicted in Figure 2. The point cloud data was converted from FLS (Faro) to PCD (Point Cloud Library) format using Autodesk Recap Pro and Cloud Compare. Then, point cloud processing and trait extraction were performed
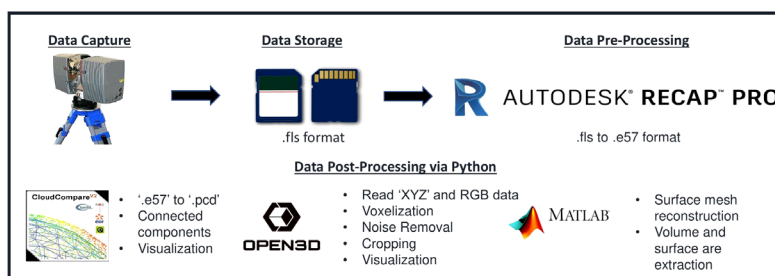

FIGURE 2
Data Processing Pipeline: Several applications were utilized in the pipeline. Autodesk recap pro was used to convert a scanner-vendor-specific file format to a generic one. CloudCompare and Open3D were employed for noise removal, voxelization, registration, and segmentation.

using Open3D, Cloud Compare, and MATLAB. This included cropping, voxelization, registration, noise removal, segmentation, and surface mesh reconstruction.

2.3.1 Point cloud file format conversion

The Trimble TX5 saves point cloud data in the FLS format, which is incompatible with the subsequent point cloud processing software. To ensure compatibility, the point cloud data format was converted from FLS to E57 using the Autodesk Recap Pro software. The E57 file format is a compact, vendor-independent format for storing point clouds, images, and metadata generated by 3D imaging systems such as laser scanners. Additionally, the E57 format retains the RGB component of the point cloud data. Finally, using Cloud Compare, the E57 files were converted to the PCD file format, which includes the Euclidean x, y, and z coordinates of each point and the RGB color value associated with each point.

2.3.2 Region of interest cropping

From the converted point cloud data, the region of interest, a block of the field, from each scan was automatically cropped out using the white-colored reference markers placed at the block's four corners. The markers were identified by separating points whose normalized R, G, and B color values are close to 1 and have a z coordinate (vertical direction, opposite of the gravity) value greater than 1 m. The z-value constraint was used to eliminate other white objects, such as orange and white plot stake identifiers. Then, the four white markers were identified as distinct objects using the connected components algorithm. Finally, the block was cropped out using the four markers' mean x and y coordinates (Figure 3A).

2.3.3 Homogenization

Due to the variable distances between the scanner and the plots, the point cloud density for a single plot captured from four different corners/perspectives varied (Figure 3B). This disparity can cause problems in mesh generation and skeletonization (Labussière et al., 2020; Xia et al., 2020). To reduce the disparity in point density, we voxelized the point clouds. The density of a point cloud is homogenized *via* uniform subsampling or voxel downsampling. We chose the latter method because it is more rigorous in ensuring uniform point distances and is invariant to the distribution of points within the sampling distance. It downsamples a point cloud uniformly using a regular voxel grid of 5 mm resolution. Briefly, voxels are used to group points, and each voxel generates an exact one point by averaging all points within an occupied voxel. Each point contains Euclidean x, y, and z coordinates and R, G, and B values.

2.3.4 Registration

Each block's four voxelized point clouds were co-registered and merged to form a single point cloud (Figure 3). The registration was carried out using the Cloudcompare 'Align' tool. We interactively identified a pair of color spheres in the point clouds, and then based on the center of the selected color spheres, the point clouds were aligned by rigid body transformation, ensuring the average root

mean squared values of the distances between the paired points after registration is less than 0.01 m. When the preceding procedure failed to produce satisfactory results, we used the iterative closest point (ICP) algorithm to achieve fine registration. The tool can register up to two-point clouds in a single registration. As a result, three registrations were necessary to merge the four perspectives into a single cloud. The final registered point cloud contained duplicate points, and their density was inconsistent. Therefore, it was voxel-downsampled to restore the uniform point density in the registered point cloud.

2.3.5 Noise removal

Due to the so-called edge effect, in which a laser beam is partially intercepted at an object's edge, and the remaining beam travels further to collide with other objects or passes through the canopy, phase-shift lidar instruments, such as the Trimble TX5, are more prone to generate noisy spurious points *via* range averaging. Additionally, poor co-registration of point clouds and wind-driven movement of the plants can introduce noisy points.
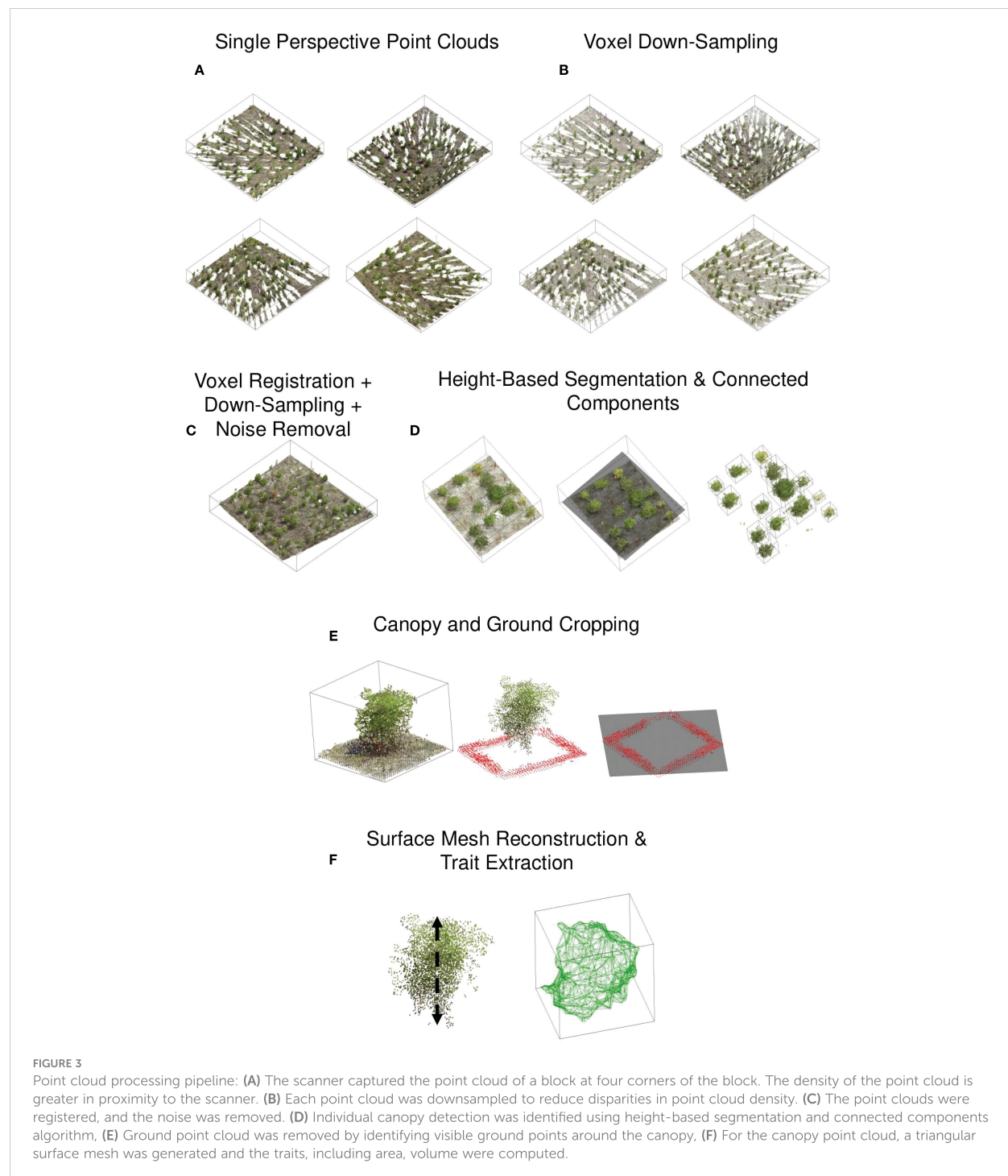
A statistical-outlier-removal algorithm was used to remove noise in the registered point cloud (Figure 3C). The algorithm begins by calculating the average distance between each point and its (k) closest neighbors. Then it discards points whose average distance exceeds a predefined threshold, $\mu+\alpha\sigma$. μ and σ denote the mean and standard deviation of the average distances, respectively, and α is a parameter that can be tuned. The smaller the value of α, the more aggressive the point removal. By monitoring the deviation of a trait value (canopy height) for various combinations, the number of nearest neighbors, k and α, were selected.

2.3.6 Plot segmentation and ground .removal

The visible ground points between the plots were used to segment each plot and remove the ground. A plane ($z = f(x, y)$) was fitted to the registered point cloud, and the points above the fitted plane were retained (Figures 3). The plane passes through the middle of each plot, and the points above the plane comprise the top portion of the canopies. Each canopy top was labeled using the connected component algorithm, and each component's mean x and y coordinate was considered the plot's approximate center. The surrounding points within a square band of width 0.1 m and inner length 0.1 m are the faithful ground points for each plot. Finally, a plane was fitted through the ground points, and the points above the plane were considered the canopy.

2.4 Trait extraction

Height, volume, and surface area were extracted from the segmented canopy point clouds (Figure 3). The canopy height was determined by subtracting the minimum z-value in the canopy points from the mean z-value of the top 3% of canopy points. The choice of using the top 3% was based on a heuristic approach, as it yielded the closest agreement with the ground truth values (See Supplemental Material (S1)). To calculate volume and surface area, canopy points were bound into a tight 'watertight'

FIGURE 3
Point cloud processing pipeline: **(A)** The scanner captured the point cloud of a block at four corners of the block. The density of the point cloud is greater in proximity to the scanner. **(B)** Each point cloud was downsampled to reduce disparities in point cloud density. **(C)** The point clouds were registered, and the noise was removed. **(D)** Individual canopy detection was identified using height-based segmentation and connected components algorithm, **(E)** Ground point cloud was removed by identifying visible ground points around the canopy, **(F)** For the canopy point cloud, a triangular surface mesh was generated and the traits, including area, volume were computed.

triangular mesh using MATLAB's trisurf algorithm, and the volume and surface area were calculated using Python's trimesh library. Traits of the projected 2D outline of the point clouds were also extracted. First, the 3d point cloud of the canopy was projected onto the plane of interest. The projected points' boundary/contour was considered the 2D canopy outline. Area, aspect ratio, roundness, circularity, and solidity of the outline are defined as (Jubery et al., 2016):

- aspect ratio = major axis of the best-fit ellipse on the outline: minor axis of the best-fit ellipse on the outline; the ratio of the major to the minor axis of the best-fitted ellipse on the outline;

- roundness = $4 * \text{Area}/(\text{pi} * \text{MajorAxis}^2)$; it indicates the closeness of the shape of the outline to a circle;

- circularity = $4 * \text{pi} * \text{Area}/(\text{Perimeter})^2$; it indicates the closeness of the form of the outline to a circle;

- solidity = Area/Convex Area; it is a measure of the compactness of the object.

2.5 Canopy fingerprints

Fingerprints are a way of representing complex physical objects mathematically. It can be used to illustrate various features on a local scale in a hierarchical and/or multi-scale way. Mathematical representation enables statistical or machine learning techniques to determine the similarity, signatures, and relationships between groups of objects. Here, we fingerprint a canopy by encoding it as a collection of sub-canopy-level features. For example, to fingerprint the shape of a canopy, we divide it into 2n+1 equally divided sections (sub-canopy). Here, we divide the canopy in the height direction into the 2n+1 sections. We then generate the signature of each sub-canopy using several geometric traits and normalize the traits concerning the traits of the center (n^{th}) sub-canopy. Finally, we represent normalized traits in a vector format to generate the fingerprint. Here 'n' is a tunable parameter that depends on the complexity of the canopies and intended downstream tasks involving the fingerprints (Figure 4).

3 Results and discussion

3.1 Parameters/conditions for TLS scanning and point cloud processing

Results showed that the point count (an indirect estimate of point density) of a canopy varies by its distance from the scanner, with as much as a 50% reduction when a canopy is close to the scanner versus when it is at the farthest possible distance from the scanner. However, this study circumvents the point density effect by registering multiple perspectives of the same scanned area. Thus, if the point density of a canopy near the scanner decreases as the scanner is moved farther away, the points lost can be recovered by scanning from a distinct perspective closer to the canopy (See Supplemental Material (S2)).

We used statistical outlier removal to reduce noise from the point cloud. Outlier selection is dependent on the values of two parameters: k, the number of neighbors, and alpha, the standard deviation ratio. We investigated the effect of 40 different combinations of these parameters on the extracted canopy height, including five values of the number of neighbors (k = 8, 16, 24, 32, 40) and eight levels of the standard deviation ratio.

The change in nearest neighbor parameter, k, from 8 to 16, average canopy height difference changed significantly after increasing (Figure 5). However, there were no significant changes in the average height difference for the remaining experiments (k = 32, k = 40). Additionally, following each experiment, visual analysis of the canopy point cloud revealed that most outliers were removed at k = 24. When k ≥ 24 and the standard deviation ratio was 0.075, no visually discernible changes in canopy structure occurred. Thus, the number of nearest neighbors, k, and the standard deviation ratio noise removal parameters were set at 24 and 0.075 for all noise removal tasks, respectively.

To evaluate the performance of co-registration, we determined the canopy top-view 2D projected area of a subsample of plants and compared it to the ground truth 2D projected area extracted from RGB images. By flattening the z values, the 3D point cloud of the canopy was projected onto the XY plane. A closed polyline represented the boundary/contour of the projected points, and its area was taken as the canopy area. An excellent agreement between the top-view canopy area and the ground truth area with R^2 = 0.826 (Figure 6).

3.2 Validation of the extracted canopy height

The results indicate that the extracted canopy height from the point cloud correlates with manual ground truth measurements taken on the same scanning day (Figure 7). Around 95% of the variation observed in extracted height values could be explained by a fitted linear least-squares regression model. Ground truth outliers in canopy height were defined as individuals with a Z-score greater than 2.5 compared to all samples' mean and standard deviation. Fewer than 2% of ground truth height measurements were considered outliers and were likely human errors in the collection. The deviation of heights from the manual measurements within the outliers ranged from 0.11 m to 0.43 m, with TLS measurements more frequently smaller than ground truth. After visualizing the outlier canopies' point cloud data, one explanation for the lower TLS height measurements is that occlusions between the measured canopy and neighboring canopies were not detected during data processing. However, manual height measurements of the canopies confirmed that the canopy segmentation and TLS height measurements were accurate (See Supplemental Material (S3)). Due to the low outlier rate (less
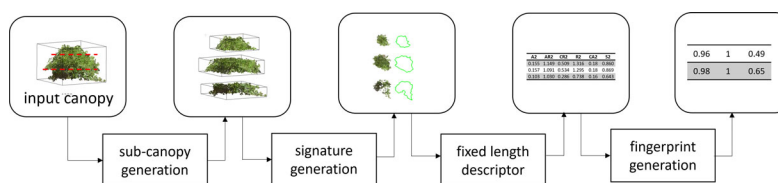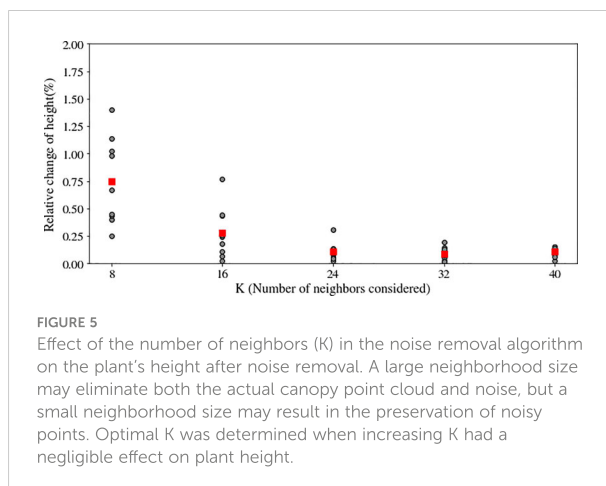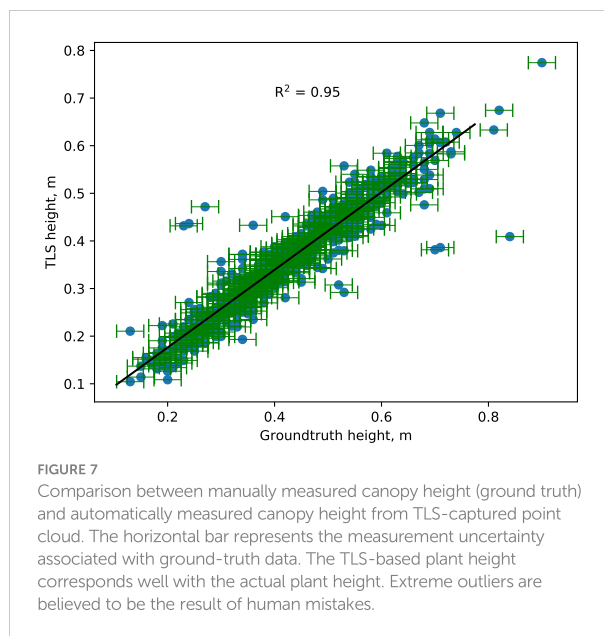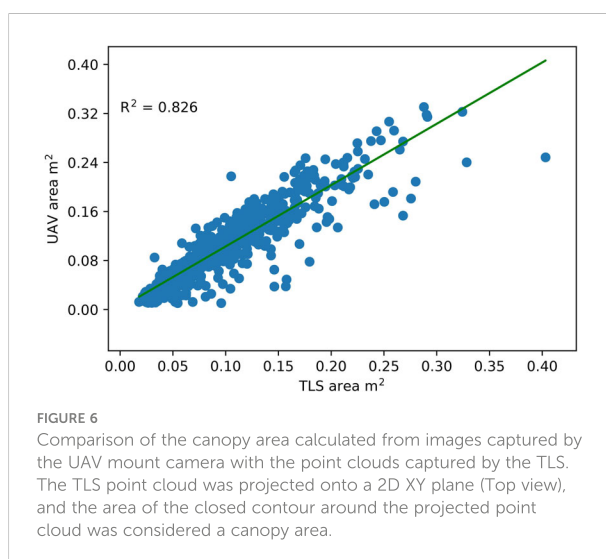

FIGURE 4
Overview of soybean canopy fingerprinting. Each canopy was subdivided into a predetermined number of sub-canopies, and the signature of each sub-canopy was extracted using several features, which were then arranged in a vector format and normalized with respect to the center sub-canopy features to represent the fingerprint of the canopy.

FIGURE 7
Comparison between manually measured canopy height (ground truth) and automatically measured canopy height from TLS-captured point cloud. The horizontal bar represents the measurement uncertainty associated with ground-truth data. The TLS-based plant height corresponds well with the actual plant height. Extreme outliers are believed to be the result of human mistakes.

than 2%) and high correlation (95%) between TLS and ground truth height data, TLS-based height extraction is a more robust method for canopy height measurement. This is useful because the TLS based method is automated.
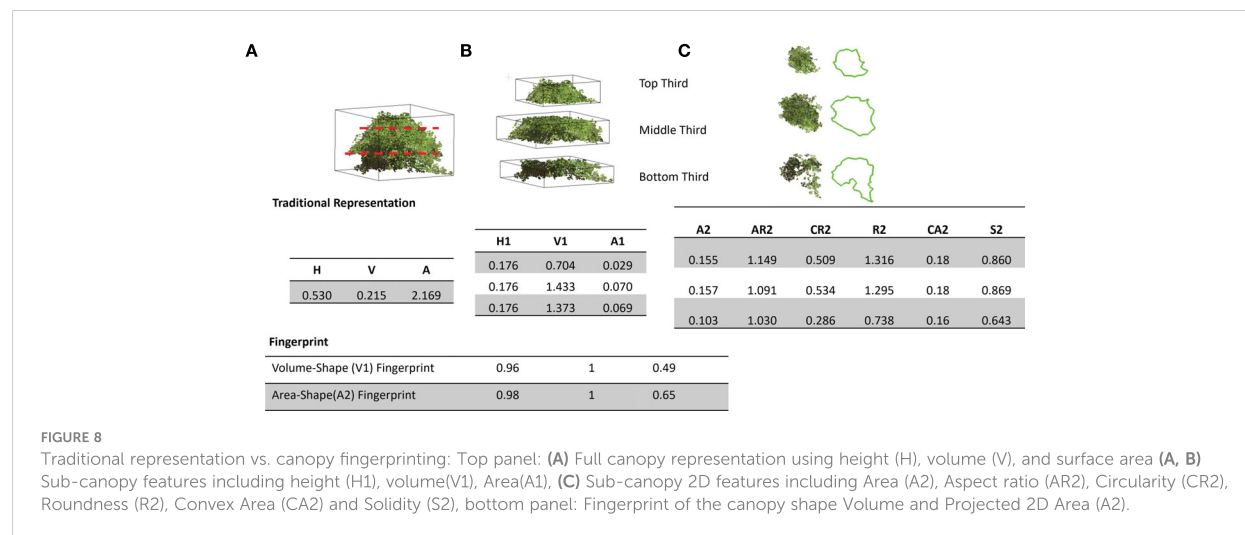
3.3 Fingerprinting and implications

Typical drone and LiDAR 3D point clouds are often limited to a top-down view of the plant canopy due to collection limitations. Figure 8 depicts the canopy's conventional representation, as often shown from the whole plant perspective, and then shows a fingerprint perspective. The fingerprint representation was created by dividing the canopy into three and nine sub-canopies (2n+1, where n = 1 and 4). With these canopy fingerprints, we can find similar-looking canopies for a given canopy or a desired/given shape. Splitting these canopies into sub-canopies enables new opportunities for phenomics and further genomic assessment of cultivars. Traditionally, only the labor and time-extensive method

of plant component partitioning would come close to this capability, but still lacked the ability for fingerprinting (Hintz and Albrecht, 1994; Raza et al., 2021). Sub-canopies paired with their fingerprints have the potential to further explore the unique relationships between certain fingerprint types or clusters with known canopy traits such as branching, leaf size, or leaf angle and their relationships with yield and yield component traits (Feng et al., 2018; Bianchi et al., 2020; Moro Rosso et al, 2021).

With digital canopy fingerprints, we can now query a given canopy (Figure 9). The canopy point clouds database was converted to a searchable fingerprint database. To query a given canopy, a fingerprint of the canopy was generated and then compared with the existing database of fingerprints to identify the possible match. This capability could enable further in-depth development and exploration of the germplasm resources for ideotypes (Kokubun, 1988; Evangelista et al., 2021; Lukas et al., 2022; Roth et al., 2022). While this work evaluated a single time point, a more in-depth and temporal fingerprint could be developed to evaluate the canopy growth and development across time. These temporal fingerprints could open new insights into agronomic traits (Pfeiffer and Pilcher, 1987; Shiraiwa and Sinclair, 1993), diseases, and pesticide applications (Hanna et al., 2008; Sikora et al., 2014; Nagasubramanian et al., 2019; Viggers et al, 2022), or even develop fingerprint responses to abiotic stress and amendments (Frederick et al., 1991; Anda et al., 2021; Shivani Chiranjeevi et al, 2021; Bonds et al., 2022). An example of searching for potential ideotypes is shown in Figure 10, where some canopies with monotonically increasing mass at the top or the bottom are identified using the fingerprint representation. To find a canopy of the desired shape, we constructed a fingerprint of the desired shape and then determined which canopy fingerprints have the least Euclidean distance to the desired fingerprint. This further enables the exploration of canopy fingerprints *in silico* not only in relation to proposed ideotypes but also as a complement to crop modeling.

FIGURE 6
Comparison of the canopy area calculated from images captured by the UAV mount camera with the point clouds captured by the TLS. The TLS point cloud was projected onto a 2D XY plane (Top view), and the area of the closed contour around the projected point cloud was considered a canopy area.

FIGURE 8
Traditional representation vs. canopy fingerprinting: Top panel: **(A)** Full canopy representation using height (H), volume (V), and surface area **(A, B)** Sub-canopy features including height (H1), volume(V1), Area(A1), **(C)** Sub-canopy 2D features including Area (A2), Aspect ratio (AR2), Circularity (CR2), Roundness (R2), Convex Area (CA2) and Solidity (S2), bottom panel: Fingerprint of the canopy shape Volume and Projected 2D Area (A2).
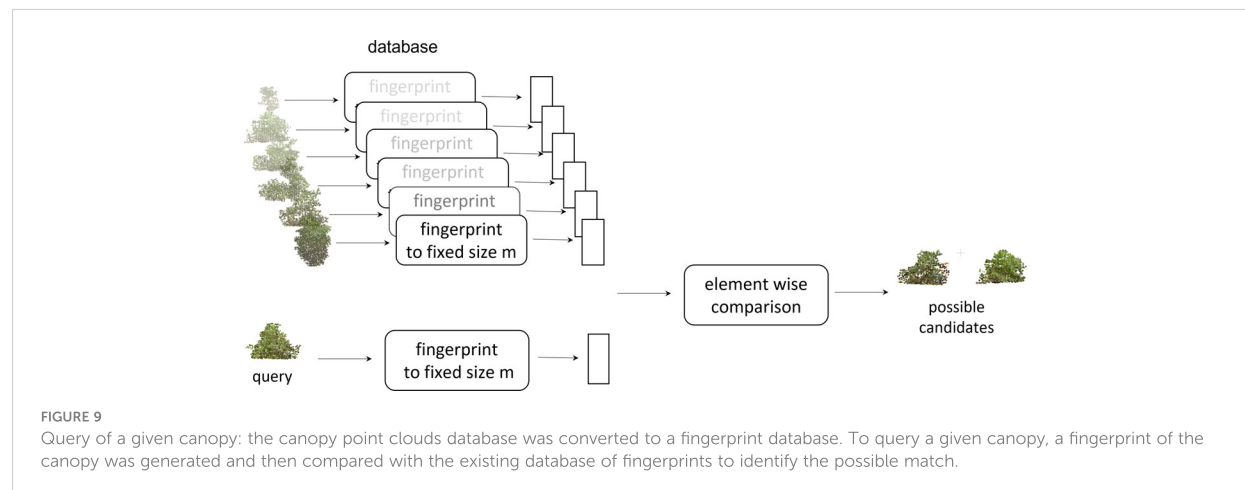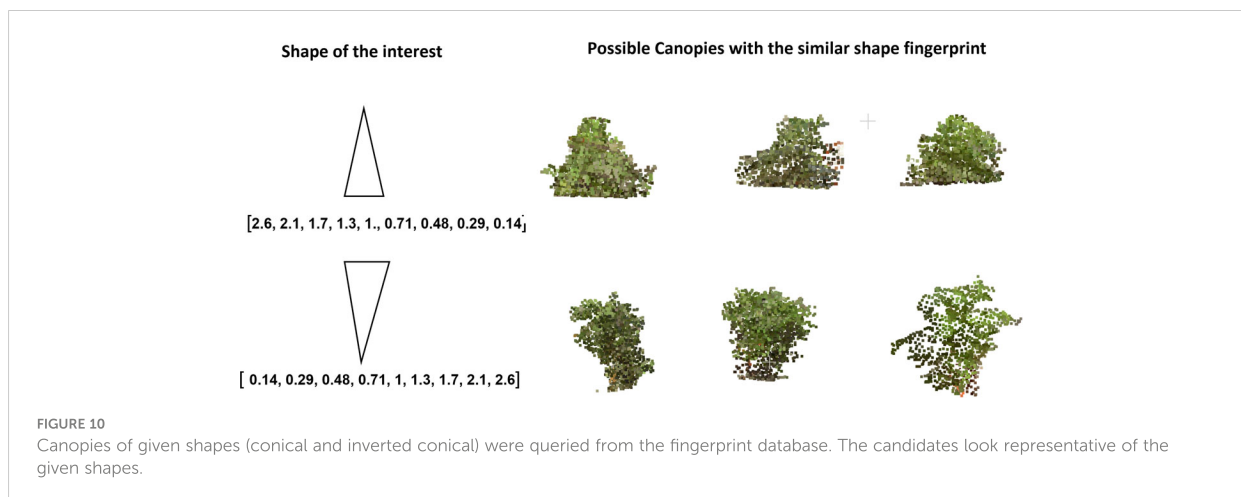
One of the core components of crop modeling is modeling the effect of light interception and radiation use efficiency of the canopy (Edwards et al, 2005; Singer et al., 2011; Zhu et al., 2020a). With canopy fingerprints integrated into a crop model, the theoretical evaluation of more genotypes in the models would be enabled, and stronger models could be developed and could also be expanded to explore environmental impacts and impacts on canopy fingerprints (Chiozza et al., 2021b; Krause et al., 2022).

An aspect of the fingerprinting to be further developed would be including the RGB data in the fingerprints. As apparent canopy color is related to soybean photosynthetic activity yield and plant health (Harrison et al, 1981; Rogovska et al, 2007; Naik et al., 2017; Yuan et al., 2019; Kaler et al., 2020; Rairdin et al., 2022). While the RGB data is already included within the voxels, additional work to evaluate the impact on fingerprint clustering due to color changes within each sublayer will be useful. Evaluating the color differences within each layer could provide an added trait assessment for radiation use efficiency relative to the amount of chlorophyll active in each canopy layer. Figure 11 shows PCA performed on the fingerprints and clustering. Each cluster's representative sample looks quite

different and shows that the fingerprint can be used to pick diverse samples with the shape analysis alone. However, the inclusion of color data and additional layers, such as horizontal sub-layering, could further enable more detailed fingerprints for assessment and query while still reducing the computation load required for searching canopy fingerprint databases. These methods can parse out canopy features (through their fingerprinting) for a more informative representation of the canopy and the role of various organs throughout the canopy on desired traits, e.g., seed yield. This enables the discovery of new relationships between canopy and organ level features and their impact on yield and yield component traits.

While this work is focused on fingerprints assessed from TLS laser point clouds, the concept of canopy fingerprints could also be implemented with any technology capable of building a full canopy 3D point cloud, such as structured light, space-carving, or full canopy structure from motion (Nguyen et al., 2015; Zhou et al., 2019; Das Choudhury et al., 2020). While we focus on canopy fingerprints, further work should be done to evaluate whole plant fingerprints, especially root system architecture (RSA) fingerprints. While 2D imaging is routine for RSA traits (Falk et al, 2020b; Jubery et al.,

FIGURE 9
Query of a given canopy: the canopy point clouds database was converted to a fingerprint database. To query a given canopy, a fingerprint of the canopy was generated and then compared with the existing database of fingerprints to identify the possible match.

FIGURE 10
Canopies of given shapes (conical and inverted conical) were queried from the fingerprint database. The candidates look representative of the given shapes.
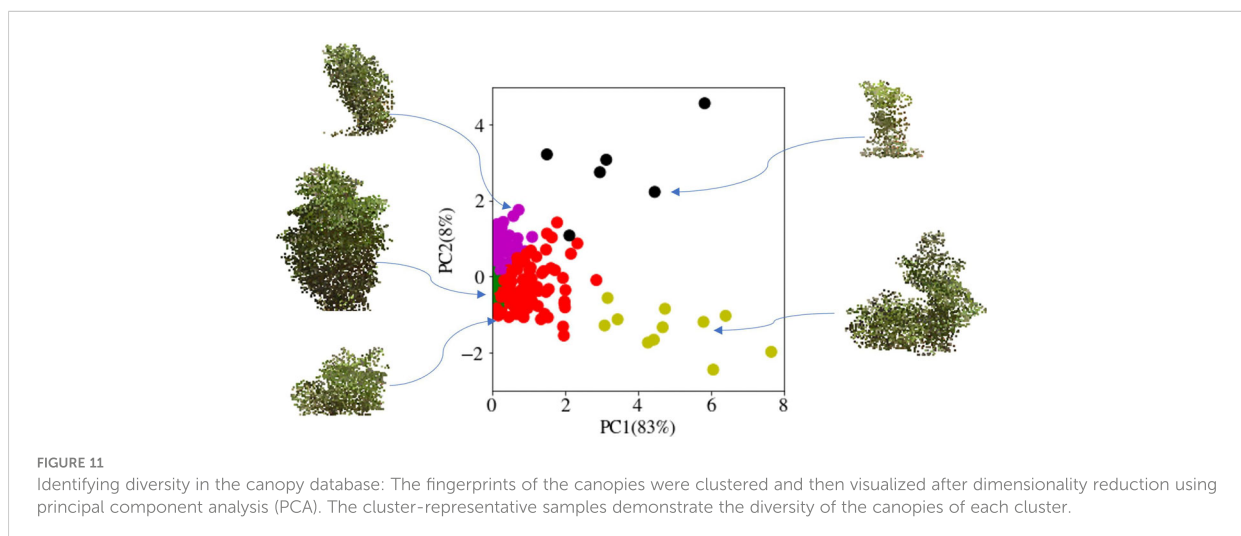
2021), there is tremendous interest in 3D imaging of root traits. RSA Fingerprints would further enable a whole plant analysis and efficient query system, and technology such as Xray-CT already enables dense 3D point clouds to be built of RSA (Gerth et al., 2021; Teramoto et al, 2021). Whole plant fingerprints could help meet the need for efficient RSA and canopy modeling, clustering, and assessment (Falk et al., 2020a; Carley et al., 2022a) while further exploring the root and shoot relationships to critical traits such as nodulation (Carley et al., 2022b). Irrespective of shoot or root fingerprints, there is tremendous potential for using this information to ID specific accessions and characterize germplasm collection (Azevedo Peixoto et al., 2017), cluster them based on their canopy features, develop relationships between agronomic, disease, or stress-induced traits, and modularize canopy features for their integration in trait development.

4 Conclusions

This study proposed an end-to-end fingerprint generation pipeline from a 3D point cloud of diverse soybean canopies grown on hill plots.

The pipeline includes point cloud noise removal, registration, plot extraction, and fingerprint generation. Canopy fingerprinting is a generic and powerful approach to constructing interpretable, multi-scale, and/or hierarchical geometric traits from 3D point cloud data. This approach is a useful middle ground between conventional approaches of extracting coarse scale (i.e., full canopy scale) geometric features that may not comprehensively capture the spatial distribution of the canopy and the more recent approaches of directly compressing the point cloud data that produce difficult to interpret features. The generated fingerprints were used to query canopies of specific shapes to the group and identify similar canopies, which could be useful for future work in further identifying the relationships between canopy, agronomic traits, and yield relative to proposed ideotypes in varying climate scenarios. Canopy or whole plant fingerprinting could be used as a pre-classifier for a complete shape-based retrieval system. It could be used as a pseudo-leveler for self-supervised model training (REF) or useful in situations of limited annotation to train ML models (Kar et al., 2021; Nagasubramanian et al., 2021). Fingerprints could be added as a semantic tag (metadata) to the point cloud and can be queried instead of opening the data, and

FIGURE 11
Identifying diversity in the canopy database: The fingerprints of the canopies were clustered and then visualized after dimensionality reduction using principal component analysis (PCA). The cluster-representative samples demonstrate the diversity of the canopies of each cluster.

can also be used for privacy-preserving deep models if data sharing is challenging (Cho et al., 2022). Fingerprinting also serves as a promising tool to store and quantify the inter-genotype or inter-environment variability. In combination with crop models and further development of voxel RGB data, these fingerprints could enable vast and rapid assessment of *in silico* genotypes for future experimentation in addition to the already improved searchability that fingerprint databases provide. As the fingerprint is based on simple sub-canopy level features, it has some limitations, and the proposed framework is sensitive to rigid transformations. If an upright plant becomes tilted, we get different fingerprint representations. However, this could be useful for estimating agronomic traits like lodging. Additionally, the vector of features as a function of plant height could be used for functional GWAS to explore putative loci with multi-scale canopy features. While our pipeline is built on TLS, future applications need to explore drone- and ground-based phenotyping (Gao et al., 2018; Guo et al., 2021; Riera et al., 2021). Plant phenotypic fingerprints serve as a novel opportunity to offer a diverse advantage to the future of high throughput phenotyping serving as a useful tool for data curation, cultivar selection, evaluation, and additional experimentation. Integration of canopy fingerprints with machine learning models can further advance the field of phenomics and cyber-agricultural systems (Singh et al., 2016; Singh et al., 2018; Singh A. K. et al., 2021).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

TY and TJ: Conceptualization; Data Collection; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing-original draft; Writing-review and editing. CC: Writing-original draft; Writing-review and editing. MC: Field experimentation; Data Collection; Writing-review and editing. SS: Funding acquisition; Resources; Supervision; Writing-review and editing. AKS: Funding acquisition; Resources; Experimentation; Writing-original draft; Writing-review and editing. AS: Conceptualization; Funding acquisition; Resources; Methodology; Supervision; Writing-review and editing. BG: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing-review and editing. All authors contributed to the article and approved the submitted version.

Funding

This project was supported by the Iowa Soybean Association, R.F. Baker Center for Plant Breeding, Plant Sciences Institute, USDA-NIFA Grants# 2017-67007-26151, 2017-67021-25965, AI Institute for Resilient Agriculture (USDA-NIFA #2021-67021-35329), COALESCE: COntext Aware LEarning for Sustainable

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1141153/full#supplementary-material

SUPPLEMENTARY FIGURE 1
R-squared (R^2) values of the predicted TLS height against the ground truth values, with respect to varying % of top canopy points. The results showed that the R^2 value was highest when canopy height was calculated using the top 3% of the canopy points.

SUPPLEMENTARY FIGURE 2
Variation of point count of a canopy relative to its distance from the scanner.

SUPPLEMENTARY FIGURE 3
Examples of height measurement differences between TLS and Manual. The images depict interactive TLS (3D point cloud)-based height measurement within the CloudCompare software.

SUPPLEMENTARY FIGURE 4
This histogram displays the distribution of canopies based on the number of points they contain. The data pertains to 464 soybean cultivars including 450 plant introduction (PI) lines that were studied, representing a diverse range of maturities, seed weights, and stem terminations, and originating from 35 different countries. Data was obtained from one or two replicates per cultivar.

References

Amann, M.-C., Bosch, T. M., Lescure, M., Myllylae, R. A., and Rioux, M. (2001). Laser ranging: A critical review of unusual techniques for distance measurement. *Optical Eng.* 40 (January), 10–19. doi: 10.1117/1.1330700

Anda, A., Simon, B., Soós, G., Teixeira da Silva, J. A., and Menyhárt, L. (2021). Water stress modifies canopy light environment and qualitative and quantitative yield components in two soybean varieties. *Irrigation Sci.* 39 (5), 549–665. doi: 10.1007/s00271-021-00728-0

Araus, JoséL., and Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* 19 (1), 52–615. doi: 10.1016/j.tplants.2013.09.008

Azevedo Peixoto, L. de, Moellers, T. C., Zhang, J., Lorenz, A. J., Bhering, L. L., Beavis, W. D., et al. (2017). Leveraging genomic prediction to scan germplasm collection for crop improvement. *PloS One* 12 (6), e01791915. doi: 10.1371/journal.pone.0179191

Bai, G., Jenkins, S., Yuan, W., Graef, G. L., and Ge, Y. (2018). Field-based scoring of soybean iron deficiency chlorosis using RGB imaging and statistical learning. *Front. Plant Sci.* 9 (July), 1002. doi: 10.3389/fpls.2018.01002

Bianchi, J. S., Quijano, A., Gosparini, C. O., and Morandi, E. N. (2020). Changes in leaflet shape and seeds per pod modify crop growth parameters, canopy light environment, and yield components in soybean. *Crop J.* 8 (2), 351–645. doi: 10.1016/j.cj.2019.09.011

Bonds, D., Koziel, J. A., De, M., Chen, B., Singh, A. K., and Licht, M. A. (2022). Dataset documenting the interactions of biochar with manure, soil, and plants: Towards improved sustainability of animal and crop agriculture. *Brown Univ. Digest Addict. Theory Application: Data* 7 (3), 325. doi: 10.3390/data7030032

Burgess, A. J., Retkute, R., Herman, T., and Murchie, E. H. (2017). Exploring relationships between canopy architecture, light distribution, and photosynthesis in contrasting rice genotypes using 3D canopy reconstruction. *Front. Plant Sci.* 8 (May), 734. doi: 10.3389/fpls.2017.00734

Cano, P., Batlle, E., Kalker, T., and Haitsma, J. (2005). A review of audio fingerprinting. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* 41 (3), 271–845. doi: 10.1007/s11265-005-4151-3

Capecchi, A., Probst, D., and Reymond, J.-L. (2020). One molecular fingerprint to rule them all: Drugs, biomolecules, and the metabolome. *J. Cheminformatics* 12 (1), 435. doi: 10.1186/s13321-020-00445-4

Carley, C. N., Chen, G., Das, K. K., Delory, B. M., Dimitrova, A., Ding, Y., et al. (2022a). Root biology never sleeps: 11th symposium of the international society of root research (ISRR11) and the 9th international symposium on root development (Rooting2021), 24-28 may 2021. *New Phytol.* 235 (6), 2149–2154. doi: 10.1111/nph.18338

Carley, C. N., Zubrod, M., Dutta, S., and Singh, A. K. (2022b). Using machine learning enabled phenotyping to characterize nodulation in three early vegetative stages in soybean. *Crop Sci.* 63, 204–226. doi: 10.1101/2022.09.28.509969

Chang, S., Lee, U., Kim, J.-B., and Deuk Jo, Y. (2022). Application of 3D-volumetric analysis and hyperspectral imaging systems for investigation of heterosis and cytoplasmic effects in pepper. *Scientia Hortic.* 302 (August), 111150. doi: 10.1016/j.scienta.2022.111150

Chiozza, M., Parmley, K., Higgins, R. H., Singh, A. K., and Miguez, F. (2021). "Assessment of the G X e X m interaction effect on the soybean LAI-yield relationship using apsim model: Implications for high-throughput phenotyping," in *ASA, CSSA, SSSA International Annual Meeting* (ASA-CSSA-SSSA) Salt Lake City, UT. https://scisoc.confex.com/scisoc/2021am/meetingapp.cgi/Paper/138146.

Chiozza, M. V., Parmley, K. A., Higgins, R. H., Singh, A. K., and Miguez, F. E. (2021). Comparative prediction accuracy of hyperspectral bands for different soybean crop variables: From leaf area to seed composition. *Field Crops Res.* 271 (108260), 1082605. doi: 10.1016/j.fcr.2021.108260

Cho, M., Nagasubramanian, K., Singh, A. K., Singh, A., Ganapathysubramanian, B., Sarkar, S., et al. (2022) *Privacy-preserving deep models for plant stress phenotyping.* Available at: https://openreview.net/pdf?id=8JkuZSH487G.

Das Choudhury, S., Maturu, S., Samal, A., Stoerger, V., and Awada, T. (2020). Leveraging image analysis to compute 3D plant phenotypes based on voxel-grid plant reconstruction. *Front. Plant Sci.* 11 (December), 521431. doi: 10.3389/fpls.2020.521431

Edwards, J. T., Purcell, L. C., and Karcher, D. E. (2005). Soybean yield and biomass responses to increasing plant population among diverse maturity groups: II. light interception and utilization. *Crop Sci.* 45 (5), 1778–1785. doi: 10.2135/cropsci2004.0570

Evangelista, J. S. P. C., Peixoto, M. Antônio, Coelho, I. F., Alves, R. S., Silva, F. F. e., de Resende, M. D. V., et al. (2021). Environmental stratification and genotype recommendation toward the soybean ideotype: A Bayesian approach. *Crop Breed. Appl. Biotechnol.* 21 (1). doi: 10.1590/1984-70332021v21n1a11

Falk, K. G., Jubery, T. Z., Mirnezami, S. V., Parmley, K. A., Sarkar, S., Singh, A., et al. (2020a). Computer vision and machine learning enabled soybean root phenotyping pipeline. *Plant Methods* 161(55), 1–19. doi: 10.1186/s13007-019-0550-5

Falk, K. G., Jubery, T. Z., O'Rourke, J. A., Singh, A., Sarkar, S., Ganapathysubramanian, B., et al. (2020b). Soybean root system architecture trait study through genotypic, phenotypic, and shape-based clusters. *Plant Phenomics.* doi: 10.34133/2020/1925495

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, glycine max (L.) Merrill 1. *Crop Sci.* 11 (6), 929–931. doi: 10.2135/cropsci1971.0011183X001100060051x

Feng, G., Luo, H., Zhang, Y., Gou, L., Yao, Y., Lin, Y., et al. (2016). Relationship between plant canopy characteristics and photosynthetic productivity in diverse cultivars of cotton (Gossypium hirsutum l.). *Crop J.* 4 (6), 499–508. doi: 10.1016/j.cj.2016.05.012

Feng, L., Raza, M. A., Li, Z., Chen, Y., Khalid, M. H. B., Du, J., et al. (2018). The influence of light intensity and leaf movement on photosynthesis characteristics and carbon balance of soybean. *Front. Plant Sci.* 9, 1952. doi: 10.3389/fpls.2018.01952

Fontaine, F., Bolton, E., Borodina, Y., and Bryant, S. H. (2007). Fast 3D shape screening of Large chemical databases through alignment-recycling. *Chem. Cent. J.* 1 (June), 12. doi: 10.1186/1752-153X-1-12

Frederick, J. R., Woolley, J. T., Hesketh, J. D., and Peters, D. B. (1991). Seed yield and agronomic traits of old and modern soybean cultivars under irrigation and soil water-deficit. *Field Crops Res.* 27 (1), 71–82. doi: 10.1016/0378-4290(91)90023-O

Gage, J. L., Richards, E., Lepak, N., Kaczmar, N., Soman, C., Chowdhary, G., et al. (2019). In-field whole-plant maize architecture characterized by subcanopy rovers and latent space phenotyping. *Plant Phenome J.* 2 (1), 1–115. doi: 10.2135/tppj2019.07.0011

Gao, T., Emadi, H., Saha, H., Zhang, J., Lofquist, A., Singh, A., et al. (2018). A novel multirobot system for plant phenotyping. *Robotics* 7 (4), 615. doi: 10.3390/robotics7040061

Gazzoni, DécioL., and Moscardi, Flávio *Effect of defoliation levels on recovery of leaf area, on yield and agronomic traits of soybeans1.* Available at: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/44655/1/pab158-96.pdf (Accessed January 5, 2023).

Gerth, S., Claußen, J., Eggert, A., Wörlein, N., Waininger, M., Wittenberg, T., et al. (2021). Semiautomated 3D root segmentation and evaluation based on X-ray CT imagery. *Plant Phenomics (Washington D.C.)* 2021 (February), 8747930. doi: 10.34133/2021/8747930

Girardeau-Montaut, D. *CloudCompare - open source project.* Available at: https://www.danielgm.net/cc/ (Accessed December 11, 2022).

Glaszmann, J. C., Kilian, B., Upadhyaya, H. D., and Varshney, R. K. (2010). Accessing genetic diversity for crop improvement. *Curr. Opin. Plant Biol.* 13 (2), 167–173. doi: 10.1016/j.pbi.2010.01.004

Guo, W., Carroll, M. E., Singh, A., Swetnam, T. L., Merchant, N., Sarkar, S., et al. (2021). UAS-based plant phenotyping for research and breeding applications. *Plant Phenomics (Washington D.C.)* 2021 (June), 9840192. doi: 10.34133/2021/9840192

Hanna, S. O., Conley, S. P., Shaner, G. E., and Santini, J. B. (2008). Fungicide application timing and row spacing effect on soybean canopy penetration and grain yield. *Agron. J.* 100 (5), 1488–1925. doi: 10.2134/agronj2007.0135

Harrison, S. A., Boerma, H. R., and Ashley, D. A. (1981). Heritability of canopy-apparent photosynthesis and its relationship to seed yield in soybeans 1. *Crop Sci.* 21 (2), 222–226. doi: 10.2135/cropsci1981.0011183X002100020004x

Herrero-Huerta, M., Eetu Puttonen, A. B., and Rainey, K. M. (2020). Canopy roughness: A new phenotypic trait to estimate aboveground biomass from unmanned aerial system. *Plant Phenomics.* doi: 10.34133/2020/6735967

Hintz, R. W., and Albrecht, K. A. (1994). Dry matter partitioning and forage nutritive value of soybean plant components. *Agron. J.* 86 (1), 59–625. doi: 10.2134/agronj1994.00021962008600010011x

Hunt, E.R., Cavigelli, M., Daughtry, C. S.T., Mcmurtrey, J. E., and Walthall, C. L. (2005). Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precis. Agric.* 6 (4), 359–785. doi: 10.1007/s11119-005-2324-5

Jarquin, D., Specht, J., and Lorenz, A. (2016). Prospects of genomic prediction in the USDA soybean germplasm collection: Historical data creates robust models for enhancing selection of accessions. *G3* 6 (8), 2329–2415. doi: 10.1534/g3.116.031443

Jubery, T. Z., Carley, C. N., Singh, A., Sarkar, S., Ganapathysubramanian, B., and Singh, A. K. (2021). Using machine learning to develop a fully automated soybean nodule acquisition pipeline (SNAP). *Plant Phenomics (Washington D.C.)* 2021 (July), 9834746. doi: 10.34133/2021/9834746

Jubery, T. Z., Shook, J., Parmley, K., Zhang, J., Naik, H. S., Higgins, R., et al. (2016). Deploying Fourier coefficients to unravel soybean canopy diversity. *Front. Plant Sci.* 7, 2066. doi: 10.3389/fpls.2016.02066

Jucker, T. (2022). Deciphering the fingerprint of disturbance on the three-dimensional structure of the world's forests. *New Phytol.* 233 (2), 612–617. doi: 10.1111/nph.17729

Kaler, A. S., Abdel-Haleem, H., Fritschi, F. B., Gillman, J. D., Ray, J. D., Smith, J. R., et al. (2020). Genome-wide association mapping of dark green color index using a diverse panel of soybean accessions. *Sci. Rep.* 1–11. doi: 10.1038/s41598-020-62034-7

Kar, S., Nagasubramanian, K., Elango, D., Nair, A., Mueller, D. S., O'Neal, M. E., et al. (2021) *Self-supervised learning improves agricultural pest classification.* Available at: https://openreview.net/pdf?id=3ndIvP2491e.

Koester, R. P., Skoneczka, J. A., Cary, T. R., Diers, B. W., and Ainsworth, E. A. (2014). Historical gains in soybean (Glycine max merr.) seed yield are driven by linear

increases in light interception, energy conversion, and partitioning efficiencies. *J. Exp. Bot.* 65 (12), 3311–3215. doi: 10.1093/jxb/eru187

Kokubun, M. (1988). Design and examination of soybean ideotypes. *JARQ* 21 (4), 237–243.

Koutsoukas, A., Paricharak, S., Galloway, W. R.J., Spring, D. R., IJzerman, A. P., Glen, R. C., et al. (2014). How diverse are diversity assessment methods? a comparative analysis and benchmarking of molecular descriptor space. *J. Chem. Inf. Modeling.* 230–242. doi: 10.1021/ci400469u

Krause, M. D., Dias, K. O.G., Singh, A. K., and Beavis, W. D. (2022). Using Large soybean historical data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. *bioRxiv.* 2022–04. doi: 10.1101/2022.04.11.487885

Labussière, M., Laconte, J., and Pomerleau, François (2020). Geometry preserving sampling method based on spectral decomposition for Large-scale environments. *Front. Robotics AI* 7 (September), 572054. doi: 10.3389/frobt.2020.572054

Liu, T., Wang, Z., and Cai, T. (2016). Canopy apparent photosynthetic characteristics and yield of two spike-type wheat cultivars in response to row spacing under high plant density. *PloS One* 11 (2), e01485825. doi: 10.1371/journal.pone.0148582

Lukas, R., Barendregt, C., Bétrix, C.-A., Hund, A., and Walter, A. (2022). High-throughput field phenotyping of soybean: Spotting an ideotype. *Remote Sens. Environ.* 269, 112797. doi: 10.1016/j.rse.2021.112797

Metz, G. L., Green, D. E., and Shibles, R. M. (1984). Relationships between soybean yield in narrow rows and leaflet, canopy, and developmental characters 1. *Crop Sci.* 24 (3), 457–462. doi: 10.2135/cropsci1984.0011183X002400030006x

Mitra, N. J., Guibas, L. J., Giesen, J., and Pauly, M. (2006). "Probabilistic fingerprints for shapes," in *Symposium on geometry processing* (Germany: Eurographics Association, Goslar), 121–130.

Moro Rosso, L. H., de Borja Reis, AndréF., and Ciampitti, I. A. (2021). Vertical canopy profile and the impact of branches on soybean seed composition. *Front. Plant Sci.* 12 (September), 725767. doi: 10.3389/fpls.2021.725767

Mourtzinis, S., and Conley, S. P. (2017). Delineating soybean maturity groups across the united states. *Agron. J.* 109 (4), 1397–14035. doi: 10.2134/agronj2016.10.0581

Nagasubramanian, K., Jones, S., Singh, A. K., Sarkar, S., Singh, A., and Ganapathysubramanian, B. (2019). Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods* 15 (1), 1–10. doi: 10.1186/s13007-019-0479-8

Nagasubramanian, K., Jones, S., Singh, A. K., Singh, A., Sarkar, S., and Ganapathysubramanian, B. (2021). Plant phenotyping with limited annotation: Doing more with less. *Earth Space Sci. Open Archive.*doi: 10.1002/essoar.10508358.1

Naik, H. S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., et al. (2017). A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods* 13 (April), 23. doi: 10.1186/s13007-017-0173-7

Nguyen, T. T., Slaughter, D. C., Max, N., Maloof, J. N., and Sinha, N. (2015). Structured light-based 3D reconstruction system for plants. *Sensors* 15 (8), 18587–16125. doi: 10.3390/s150818587

Pangga, I. B., Hanan, J., and Chakraborty, S. (2011). Pathogen dynamics in a crop canopy and their evolution under changing climate. *Plant Pathol.* 60 (1), 70–81. doi: 10.1111/j.1365-3059.2010.02408.x

Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sarkar, S., and Singh, A. K. (2019). Machine learning approach for prescriptive plant breeding. *Sci. Rep.* 9 (1), 171325. doi: 10.1038/s41598-019-53451-4

Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., and Singh, A. K. (2019). Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. *Plant Phenomics (Washington D.C.)* 2019 (July), 5809404. doi: 10.34133/2019/5809404

Paulus, S., Behmann, J., Mahlein, A.-K., Plümer, L., and Kuhlmann, H. (2014). Low-cost 3D systems: Suitable tools for plant phenotyping. *Sensors* 14 (2), 3001–3185. doi: 10.3390/s140203001

Paulus, S., Dupuis, J., Mahlein, A.-K., and Kuhlmann, H. (2013). Surface feature based classification of plant organs from 3D laserscanned point clouds for plant phenotyping. *BMC Bioinf.* 14 (July), 238. doi: 10.1186/1471-2105-14-238

Pfeiffer, T. W., and Pilcher, D. (1987). Effect of early and late flowering on agronomic traits of soybean at different planting dates 1. *Crop Sci.* 27 (1), 108–112. doi: 10.2135/cropsci1987.0011183X002700010027x

Probst, D. (2018). And Jean-Louis reymond A probabilistic molecular fingerprint for big data settings. *J. Cheminformatics* 10 (1), 665. doi: 10.1186/s13321-018-0321-8

Purcell, L. C. (2000). Soybean canopy coverage and light interception measurements using digital imagery. *Crop Sci.* 40 (3), 834–837. doi: 10.2135/cropsci2000.403834x

Pydipati, R., Burks, T. F., and Lee, W. S. (2006). Identification of citrus disease using color texture features and discriminant analysis. *Comput. Electron. Agric.* 52 (1), 49–59. doi: 10.1016/j.compag.2006.01.004

Rairdin, A., Fotouhi, F., Zhang, J., Mueller, D. S., Ganapathysubramanian, B., Singh, A. K., et al. (2022). Deep learning-based phenotyping for genome wide association studies of sudden death syndrome in soybean. *Front. Plant Sci.* 13 (October), 966244. doi: 10.3389/fpls.2022.966244

Raza, M. A., Gul, H., Yang, F., Ahmed, M., and Yang, W. (2021). Growth rate, dry matter accumulation, and partitioning in soybean (Glycine max l.) in response to defoliation under high-rainfall conditions. *Plants* 10 (8), 1497. doi: 10.3390/plants10081497

Reuter, M., Wolter, F.-E., and Peinecke, N. (2005). "Laplace-Spectra as fingerprints for shape matching," in *Proceedings of the 2005 ACM symposium on solid and physical modeling* (New York, NY, USA: Association for Computing Machinery), 101–106.

Riera, L. G., Carroll, M. E., Zhang, Z., Shook, J. M., Ghosal, S., Gao, T., et al. (2021). Deep multiview image fusion for soybean yield estimation in breeding applications. *Plant Phenomics (Washington D.C.)* 2021 (June), 9846470. doi: 10.34133/2021/9846470

Rogovska, N. P., Blackmer, A. M., and Mallarino, A. P. (2007). Relationships between soybean yield, soil pH, and soil carbonate concentration. *Soil Sci. Soc. America J. Soil Sci. Soc. America* 71 (4), 1251–1565. doi: 10.2136/sssaj2006.0235

Roth, L., Barendregt, C., Bétrix, C.A., Hund, A., and Walter, A. (2022). High-throughput field phenotyping of soybean: Spotting an ideotype. *Remote Sensing Environ.* 269, 112797. doi: 10.1016/j.rse.2021.112797

Rötter, R. P., Tao, F., Höhn, J. G., and Palosuo, T. (2015). Use of crop simulation modelling to aid ideotype design of future cereal cultivars. *J. Exp. Bot.* 66 (12), 3463–3476. doi: 10.1093/jxb/erv098

Sadras, V. O., and Slafer, G. A. (2012). Environmental modulation of yield components in cereals: Heritabilities reveal a hierarchy of phenotypic plasticities. *Field Crops Res.* 127 (February), 215–224. doi: 10.1016/j.fcr.2011.11.014

Shiraiwa, T., and Sinclair, T. R. (1993). Distribution of nitrogen among leaves in soybean canopies. *Crop Sci.* 33 (4), 804–885. doi: 10.2135/cropsci1993.0011183X003300040035x

Shivani Chiranjeevi, T. Y., Jubery, T. Z., Nagasubramanian, K., Sarkar, S., Singh, A. K., Singh, A., et al. (2021). Exploring the use of 3D point cloud data for improved plant stress rating.

Shook, J. M., Lourenco, D., and Singh, A. K. (2021a). PATRIOT: A pipeline for tracing identity-by-Descent for chromosome segments to improve genomic prediction in self-pollinating crop species. *Front. Plant Sci.* 12 (676269), 2095. doi: 10.3389/fpls.2021.676269

Shook, J. M., Zhang, J., Jones, S. E., Singh, A., Diers, B. W., and Singh, A. K. (2021b). Meta-GWAS for quantitative trait loci identification in soybean. *G3 (Bethesda Md.)* 11 (7). doi: 10.1093/g3journal/jkab117

Sikora, E. J., Allen, T. W., Wise, K. A., Bergstrom, G., Bradley, C. A., Bond, J., et al. (2014). A coordinated effort to manage soybean rust in north America: A success story in soybean disease monitoring. *Plant Dis.* 98 (7), 864–875. doi: 10.1094/PDIS-02-14-0121-FE

Singer, J. W., Meek, D. W., Sauer, T. J., Prueger, J. H., and Hatfield, J. L. (2011). Variability of light interception and radiation use efficiency in maize and soybean. *Field Crops Res.* 121 (1), 147–525. doi: 10.1016/j.fcr.2010.12.007

Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in plant science* 21 (2), 110–124. doi: 10.1016/j.tplants.2015.10.015

Singh, A.K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends in Plant Science* 23(10), 883–898. doi: 10.1016/j.tplants.2018.07.004

Singh, A. K., Singh, A., Sarkar, S., Ganapathysubramanian, B., Schapaugh, W., Miguez, F. E., et al. (2021). "High-throughput phenotyping in soybean," in *High-throughput crop phenotyping*. Eds. J. Zhou and H. T. Nguyen (Cham: Springer International Publishing), 129–163.

Singh, D. P., Singh, A. K., and Singh, A. (2021). *Plant breeding and cultivar development* (San Diego, CA: Academic Press).

Spannaus, A., Law, K. J.H., Luszczek, P., Nasrin, F., Micucci, C. P., Liaw, P. K., et al. (2021). Materials fingerprinting classification. *Comput. Phys. Commun.* 266 (September), 108019. doi: 10.1016/j.cpc.2021.108019

Storey, E. A. (2020). Mapping plant growth forms using structure-from-Motion data combined with spectral image derivatives. *Remote Sens. Lett.* 11 (5), 426–435. doi: 10.1080/2150704X.2020.1730467

Teramoto, S., Tanabata, T., and Uga, Y. (2021). RSAtrace3D: Robust vectorization software for measuring monocot root system architecture. *BMC Plant Biol.* 21 (1), 3985. doi: 10.1186/s12870-021-03161-9

Ubbens, J., Cieslak, M., Prusinkiewicz, P., Parkin, I., Ebersbach, J., and Stavness, I. (2020). Latent space phenotyping: Automatic image-based phenotyping for treatment studies. *Plant Phenomics (Washington D.C.)* 2020 (January), 5801869. doi: 10.34133/2020/5801869

Vandenberghe, B., Depuydt, S., and Van Messem, A. (2018). How to make sense of 3D representations for plant phenotyping: A compendium of processing and analysis techniques. doi: 10.31219/osf.io/r84mk

Viggers, J. N., Kandel, Y. R., and Mueller, D. S. (2022). The impact of fungicide application method on soybean canopy coverage, frogeye leaf spot, and yield. *Plant Health Progress.*doi: 10.1094/php-03-22-0024-rs

Virdi, K. S., Sreeka, S., Dobbels, A., Haaning, A., Jarquin, D., Stupar, R. M., et al. (2021). Branch angle and leaflet shape are associated with canopy coverage in soybean. *Res. Square.*doi: 10.21203/rs.3.rs-806530/v1

Walter, J. D.C., Edwards, J., McDonald, G., and Kuchel, H. (2019). Estimating biomass and canopy height with LiDAR for field crop breeding. *Front. Plant Sci.* 10 (September), 1145. doi: 10.3389/fpls.2019.01145

Wang, L., Zhao, Lu, Liu, X., Fu, J., and Zhang, A. (2021). SepPCNET: Deeping learning on a 3D surface electrostatic potential point cloud for enhanced toxicity classification and its application to suspected environmental estrogens. *Environ. Sci. Technol.* 9958–9967. doi: 10.1021/acs.est.1c01228

Wu, Y., Wu, T., and Yau, S.-T. (2022). Surface eigenvalues with lattice-based approximation in comparison with analytical solution. doi: 10.48550/arXiv.2203.03603

Xia, S., Chen, D., Wang, R., Li, J., and Zhang, X. (2020). Geometric primitives in LiDAR point clouds: A review. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 13, 685–707. doi: 10.1109/JSTARS.2020.2969119

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol. Plant* 13 (2), 187–2145. doi: 10.1016/j.molp.2020.01.008

Yuan, W., Wijewardane, N. K., Jenkins, S., Bai, G., Ge, Y., and Graef, G. L. (2019). Early prediction of soybean traits through color and texture features of canopy RGB imagery. *Sci. Rep.* 9 (1), 140895. doi: 10.1038/s41598-019-50480-x

Zhou, J., Fu, X., Zhou, S., Zhou, J., Ye, H., and Nguyen, H. T. (2019). Automated segmentation of soybean plants from 3D point cloud using machine learning. *Comput. Electron. Agric.* 162 (July), 143–153. doi: 10.1016/j.compag.2019.04.014

Zhou, Q.-Y., Park, J., and Koltun, V. (2018). Open3D: A modern library for 3D data processing. doi: 10.48550/arXiv.1801.09847

Zhu, B., Liu, F., Xie, Z., Guo, Y., Li, B., and Ma, Y. (2020a). Quantification of light interception within image-based 3-d reconstruction of sole and intercropped canopies over the entire growth season. *Ann. Bot.* 126 (4), 701–125. doi: 10.1093/aob/mcaa046

Zhu, R., Sun, K., Yan, Z., Yan, X., Yu, J., Shi, J., et al. (2020b). Analysing the phenotype development of soybean plants using low-cost 3D reconstruction. *Sci. Rep.* 10 (1). doi: 10.1038/s41598-020-63720-2

Zhu, F., Thapa, S., Gao, T., Ge, Y., Walia, H., and Yu, H. (2018). "3D reconstruction of plant leaves for high-throughput phenotyping," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA. 2018, 4285–4293. doi: 10.1109/BigData.2018.8622428