

## Article

# Genotyping Identification of Maize Based on Three-Dimensional Structural Phenotyping and Gaussian Fuzzy Clustering

Bo Xu <sup>1,2,3</sup>, Chunjiang Zhao <sup>1,2,3</sup>, Guijun Yang <sup>2,3</sup>, Yuan Zhang <sup>2,3</sup>, Changbin Liu <sup>2,3</sup>, Haikuan Feng <sup>2,3</sup>, Xiaodong Yang <sup>2,3</sup> and Hao Yang <sup>2,3,\*</sup>

<sup>1</sup> School of Chemistry and Bioengineering, University of Science and Technology Beijing, Beijing 100083, China; xub@nercita.org.cn (B.X.); zhaoj@nercita.org.cn (C.Z.)

<sup>2</sup> Key Laboratory of Quantitative Remote Sensing in Agriculture, Ministry of Agriculture and Rural Affairs, Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; yanggj@nercita.org.cn (G.Y.); zy8773582147@163.com (Y.Z.); liucb@nercita.org.cn (C.L.); fenghk@nercita.org.cn (H.F.); yangxd@nercita.org.cn (X.Y.)

<sup>3</sup> National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

\* Correspondence: yangh@nercita.org.cn

**Abstract:** The maize tassel represents one of the most pivotal organs dictating maize yield and quality. Investigating its phenotypic information constitutes an exceedingly crucial task within the realm of breeding work, given that an optimal tassel structure is fundamental for attaining high maize yields. High-throughput phenotyping technologies furnish significant tools to augment the efficiency of analyzing maize tassel phenotypic information. Towards this end, we engineered a fully automated multi-angle digital imaging apparatus dedicated to maize tassels. This device was employed to capture images of tassels from 1227 inbred maize lines falling under three genotype classifications (NSS, TST, and SS). By leveraging the 3D reconstruction algorithm SFM (Structure from Motion), we promptly obtained point clouds of the maize tassels. Subsequently, we harnessed the TreeQSM algorithm, which is custom-designed for extracting tree topological structures, to extract 11 archetypal structural phenotypic parameters of the maize tassels. These encompassed main spike diameter, crown height, main spike length, stem length, stem diameter, the number of branches, total branch length, average crown diameter, maximum crown diameter, convex hull volume, and crown area. Finally, we compared the GFC (Gaussian Fuzzy Clustering algorithm) used in this study with commonly used algorithms, such as RF (Random Forest), SVM (Support Vector Machine), and BPNN (BP Neural Network), as well as k-Means, HCM (Hierarchical), and FCM (Fuzzy C-Means). We then conducted a correlation analysis between the extracted phenotypic parameters of the maize tassel structure and the genotypes of the maize materials. The research results showed that the Gaussian Fuzzy Clustering algorithm was the optimal choice for clustering maize genotypes. Specifically, its classification accuracies for the Non-Stiff Stalk (NSS) genotype and the Tropical and Subtropical (TST) genotype reached 67.7% and 78.5%, respectively. Moreover, among the materials with different maize genotypes, the number of branches, the total branch length, and the main spike length were the three indicators with the highest variability, while the crown volume, the average crown diameter, and the crown area were the three indicators with the lowest variability. This not only provided an important reference for the in-depth exploration of the variability of the phenotypic parameters of maize tassels but also opened up a new approach for screening breeding materials.

Academic Editor: Georgios C. Koubouris

Received: 18 November 2024

Revised: 27 December 2024

Accepted: 30 December 2024

Published: 2 January 2025

**Citation:** Xu, B.; Zhao, C.; Yang, G.; Zhang, Y.; Liu, C.; Feng, H.; Yang, X.; Yang, H. Genotyping Identification of Maize Based on Three-Dimensional Structural Phenotyping and Gaussian Fuzzy Clustering. *Agriculture* **2025**, *15*, 85. <https://doi.org/10.3390/agriculture15010085>

**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/>).

**Keywords:** tassel; 3D phenotyping; TreeQSM; genotyping; clustering

---

## 1. Introduction

Maize is one of the world's most important food crops, and its yield is critical to ensure food security [1]. Genetic breeding techniques can help increase maize yields and ensure global food security. The accelerated localization of functional genes through high-throughput phenotypic information acquisition techniques has a huge market demand [2] and is key to developing improved varieties [3].

Studies have shown that the structure of maize tassels is directly related to pollen yield and pollination efficiency, which not only affects final kernel yield but also inbred line retention [4]. However, overly dense male tassel branching and oversized male tassel size do not promote final maize yield. On the contrary, on the one hand, it will reduce the photosynthetic efficiency of leaves due to the overly large shading area, and, on the other hand, it will consume a large amount of nutrients, affecting the formation and plumpness of grains [5,6]. Finding the right structure of maize tassels is important for breeding good maize varieties. In recent years, high-throughput phenotyping technologies have played an important role in the breeding work of many crops, and more and more genes related to maize tassel phenotypic traits have been successfully located [7]. High-throughput phenotyping technologies have been crucial in crop breeding. In combination with GWAS technology, more genes related to maize tassel traits like tassel length, branch length, branch number, stem diameter, main spike diameter, and stem length, which can be easily, cheaply, and accurately quantified, have been located successfully. This has made the maize tassel structure a convenient field for QTL study [8,9] and an important foundation for quickly screening excellent tassel traits [10,11].

However, for high-throughput phenotypic measurements of the maize tassels, the multitude of the characterization parameters and the short pollination period, during which the tassels stop growing and begin to degrade, make the effective time available for measuring structural traits in the tassels very limited. In addition, maize tassels are very easy to dehydrate after leaving the plant, resulting in tassels that are easy to break and even undergo large morphological changes, which is very unfavorable to the acquisition of high-throughput data on maize tassels, and indicators such as maximum crown diameter, crown height, average crown diameter, and crown width are also difficult to be measured manually. Therefore, it is particularly important to develop high-throughput measurement techniques for maize tassel.

Currently, a large number of image- and point cloud-based methods have been used to extract crop phenotypes, such as using images to extract the 3D structure of maize seedlings [12] and extracting the leaf area and plant height of maize [13–15], as well as extracting the number of branches, branch length, and the angle of the tassel [16,17], and systems have also been developed for this purpose [18–20]. However, for analyzing three-dimensional structural parameters, 2D images are unable to overcome the occlusion problem. Three-dimensional point cloud data offer significant advantages in extracting multiple structural parameters of a target object [21–23]. Currently, most point cloud data are generated by 3D scanning devices such as LiDAR (2D/3D), TOF cameras (time-of-flight cameras), and structured light scanners. Terrestrial laser scanning (TLS) can accurately characterize the 3D structure of trees with millimeter-level accuracy [24,25]. Lei and Suresh et al. [26] used terrestrial LiDAR data to extract the 3D structural information of maize plants, including leaf area and leaf angle. Paulus et al. [27] studied the acquisition and reconstruction of the 3D point cloud of barley using LiDAR with a measuring arm to extract the leaf area and plant height. Chaivivatrakul et al. [28] used a TOF camera to acquire

the point cloud data of maize seedlings to reconstruct a 3D model of maize seedlings, extracting information such as maximum/minimum stem diameter, stem height, leaf area, leaf length, and leaf angle. However, these methods often lack point cloud density or geometric accuracy when studying organ-scale objects such as maize tassel. Dong and Sheng et al. [4,29] constructed a 3D point cloud of a maize tassel using images and extracted metrics such as plant height, leaf area, main spike length, and branching number. However, there are problems of low automation, high cost, and low analysis efficiency. Xu et al. [30] made a significant development. They developed a system known as TIPS, short for three-D image phenotype system. This system utilizes the TreeQSM algorithm. With this algorithm, it reconstructs and analyzes the topology of a single tree. By doing so, it manages to extract the phenotypic parameters of the maize tassel structure that bears resemblance to the tree structure. This extraction process brings several solutions. It solves the problem of automatically acquiring maize tassel images. It also addresses the issues related to 3D point cloud generation and tassel topology reconstruction. Moreover, the system further evaluates certain aspects. It looks into the spatial variability of the point cloud. It also assesses the extracted tassel axial spindle lengths. Additionally, it examines the number of branches, the total branch lengths, and the branch angles. The evaluations lead to a conclusion. The results show that the system is both efficient and accurate. In another aspect, the system accomplishes a particular task. It resolves the tool problem for extracting phenotypic parameters of maize tassels. However, there is a drawback. There has been no analysis carried out on the genetic relationship between phenotype and genotyping.

Tassel clustering genotypes using tassel phenotypic parameters are efficient and easy to use, and they are an innovative aspect of this study. Since maize tassel structural phenotypes are largely controlled by genotypes, different genotypes exhibit significant differences in tassel morphology, providing feasibility for analyzing tassel structural phenotypic information to explore differences in phenotypic performance among four maize genotypes. Song et al. [31] obtained different gene categories after clustering different functional genes through the GFC algorithm, showing the distribution characteristics of gene expression data in the clustering space. Minton et al. [32] provided a step-by-step guide for conducting plant morphometric analysis using the Gaussian Mixture Model and established a new conceptual, statistical, and probabilistic framework for analyzing morphometric data in plant taxonomy. Yang et al. [33] extended the model-based Gaussian clustering method to the fuzzy model and put forward an unsupervised Gaussian clustering algorithm based on the fuzzy model, which solved the problems of the traditional Gaussian clustering algorithm being sensitive to initialization and requiring the number of clusters to be specified in advance. Fuchs et al. [34] compared the Gaussian Mixture Model with existing advanced mixed data clustering models on multiple commonly used datasets. The results showed that this model can achieve more accurate and stable clustering results when dealing with mixed data and is superior to traditional clustering methods. Yang et al. [35] used the GoldenGate assay kit (Illumina, San Diego, CA, USA) containing 1536 SNPs and an association mapping approach for 12 phenotypic traits to typify 527 lines and explored the relationship between genotypic and phenotypic variance. The relationship between genotypes and phenotypic variants was explored. Pan et al. [36] used hierarchical clustering to cluster the phenotypic parameters of 24 watermelon germplasm resources, revealing their diversity and variation patterns. Xie et al. [37] used UAV RGB images to cluster eight phenotypic parameters of 300 rapeseed varieties during flowering using the dynamic K-Means method, grouping the rapeseed varieties into four categories based on flowering characteristics. HAN et al. [38] studied phenotypic variation among different genotypes using dynamic changes in maize plant height and analyzed the maize genotypes using the GFC (Gaussian Fuzzy Clustering). Rincon, Mustafa,

and Shrestha et al. [39,40] conducted the clustering analysis of maize materials based on multiple phenotypic traits such as plant height, leaf length, width, leaf temperature, root density, and weight to characterize trait variability in different maize materials. However, studies on clustering genotypes using tassel phenotypic parameters are limited.

Existing research results have solved the problems of multi-angle image acquisition, 3D synthesis, topological reconstruction, and the automatic extraction of phenotypic parameters such as maize main spike length, branch number, branch length, and branch angle [30]. On this basis, this study further extracted seven phenotypic parameters such as main spike diameter, crown height, stem diameter, average crown diameter, maximum crown diameter, crown volume, and crown area from 1227 maize tassels. Additionally, after denoising the original point cloud data using Gaussian filtering and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithms, 1227 single-plant maize tassel point clouds were successfully extracted. Finally, genotypic clustering analysis was performed on these 1227 single-plant maize tassel point cloud data using the Gaussian Fuzzy Clustering (GFC) algorithm. The clustering results showed that the method of this study had a relatively good classification effect on Non-Stiff Stalk (NSS) and Tropical and Subtropical (TST) materials. The method could reveal the variability of multiple structural phenotypic parameters of the maize tassel. Additionally, compared with traditional commonly used machine learning algorithms, the clustering algorithm adopted in this study had a higher classification accuracy. This study innovatively used the three-dimensional structural phenotype of the maize tassel to cluster maize genotypes, providing a certain reference value for the screening of materials in the maize breeding process. Finally, after comparing the measured data, this study found that although the clustering accuracy of the measured data increased by 5% to 10%, the method adopted in this study was more efficient, more economical, and more practical.

## 2. Materials and Methods

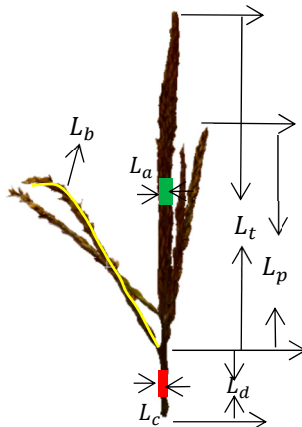
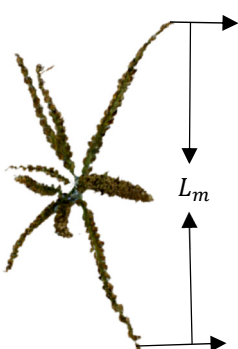
### 2.1. Study Area and Manual Measurement Data of Tassel

The study area is located at the National Precision Agriculture Research and Demonstration Base in Xiaotangshan, Changping District, Beijing (40°10'60" N, 116°26'30" E). In 2022, we collected a total of 409 samples, resulting in 1227 maize tassel images taken after pollination (each sample consists of three sets of images, with 72 images per set). Each sample was planted in two rows with a row spacing of 0.6 m and a plant spacing of 0.4 m. The genotypic types of these 409 sets of maize materials were based on the results analyzed by the Bayesian Markov Chain Monte Carlo (BMCMC) program using the STRUC-TURE software (version 2.3.4) [41,42]. These results classified the genotypes of the maize materials into a total of four subgroups: The Tropical and Subtropical (TST) maize germplasm resources are adapted to Tropical and Subtropical climates, possessing hard-core traits that enable them to grow well in hot and humid environments, and this genotype of germplasm resources has strong drought resistance. The Non-Stiff Stalk (NSS) and Stiff Stalk (SS) maize germplasm resources are the way that Dr. Duvick of Pioneer in the United States classifies maize hybrids according to the different requirements of the parent and the mother. The NSS is categorized as a replica, which is characterized by a small pollen load and a short pollen dispersal time, while the SS is classified as a parent, which is characterized by a high seed yield. Among these, there were 28 SS genotypes, 175 NSS genotypes, and 140 TST genotypes. The remaining 55 samples had genotyping probabilities below 0.60 and were classified as mixed subpopulations (MIXED).

The measured data in this study included the main spike length, branch length, branch number, spike diameter, stem diameter, stem length, crown diameter, and crown height of the tassel. The stem length refers to the portion from the lowest flower silk on the main stem of the tassel to the first node at the top of the maize stem. Maxcrown diam

and crown height refer to the maximum crown diameter and maximum crown height, respectively. A detailed measurement diagram of the phenotypic parameters is shown in Table 1. All lengths were measured in a straightened state, using a ruler with a resolution of 1 mm. The diameter of  $L_a$  is the average of the diameters of five equally spaced regions on the main spike, while  $L_d$  refers to the length from the starting point of the first branch of the tassel to the highest node of the maize plant below the tassel. Its diameter  $L_c$  is also the average of five equally spaced diameters.

**Table 1.** Descriptive statistics of the measured phenotypic parameters.

| Definition                                                                          | Name                | Max Value (cm) | Min Value (cm) | Mean Value (cm) | Variance | Note                |
|-------------------------------------------------------------------------------------|---------------------|----------------|----------------|-----------------|----------|---------------------|
|   | Main spike length   | 56.50          | 14.50          | 30.50           | 5.25     | $L_t$               |
|                                                                                     | Branch length       | 40.6           | 16.90          | 28.73           | 4.01     | $L_b$               |
|                                                                                     | Branch num          | 38.00          | 2.00           | 12.37           | 6.22     | Number of all $L_b$ |
|                                                                                     | Stalk diam          | 6.15           | 0.14           | 0.73            | 0.66     | $L_c$               |
|                                                                                     |                     |                |                |                 |          |                     |
|  | Main spike diameter | 10.00          | 0.35           | 1.00            | 1.03     | $L_a$               |
|                                                                                     | Stalk Length        | 50.00          | 6.00           | 19.44           | 5.55     | $L_d$               |
|                                                                                     | Maxcrown diam       | 38.50          | 1.80           | 10.37           | 6.70     | $L_m$               |
|                                                                                     | Crown height        | 36.30          | 3.00           | 18.32           | 4.50     | $L_p$               |

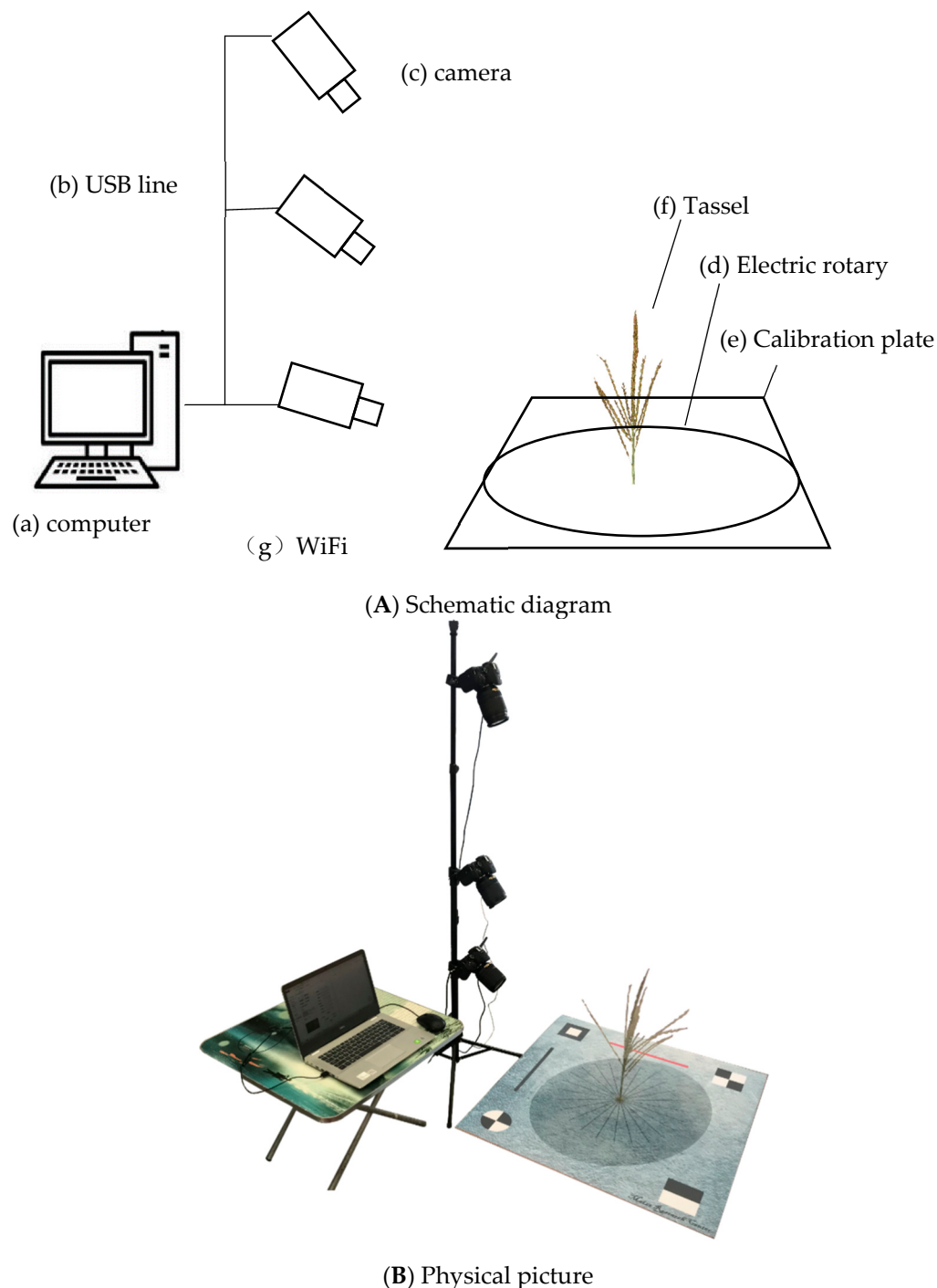
## 2.2. TIPS Development and Image Acquisition

The acquisition of image data is primarily automated through the TIPS system, which consists of four main components: hardware, data communication and transmission, data acquisition, and data processing. The hardware includes an electric turntable, a reference board, three digital cameras, and a computer. The data communication and transmission module consists of Wi-Fi wireless communication for the electric turntable and RS232-to-USB wired communication for the cameras. The data acquisition module includes parameter settings, image preview, control of camera shooting, and data classification and storage. The processing module encompasses 3D point cloud reconstruction based on image sequences, point cloud calibration, point cloud segmentation, and phenotypic information extraction.

Using the TIPS system's hardware module, data transfer, data collection, and data processing, we captured 360-degree images of the maize tassel. For each tassel, 72 images were taken, covering the upper, middle, and lower sections. The digital camera used was

a Nikon D5600 with an effective resolution of 24.16 megapixels, paired with a Nikon 18–140 mm f/3.5–5.6 G ED VR lens, supporting 7× zoom. The camera was set to a fixed focal length mode for the shooting. The electric turntable, controlled via wireless Wi-Fi with a precision of 0.01 degrees, has a diameter of 60 cm and is powered by a 24 V DC supply. The calibration board is a square with an 80 cm side length.

The data acquisition and storage processes were automatically completed using data acquisition control software developed in C# on the Visual Studio 2010 platform. Figure 1 shows the schematic and physical diagrams of the TIPS system structure.



**Figure 1.** System structure of TIPS.

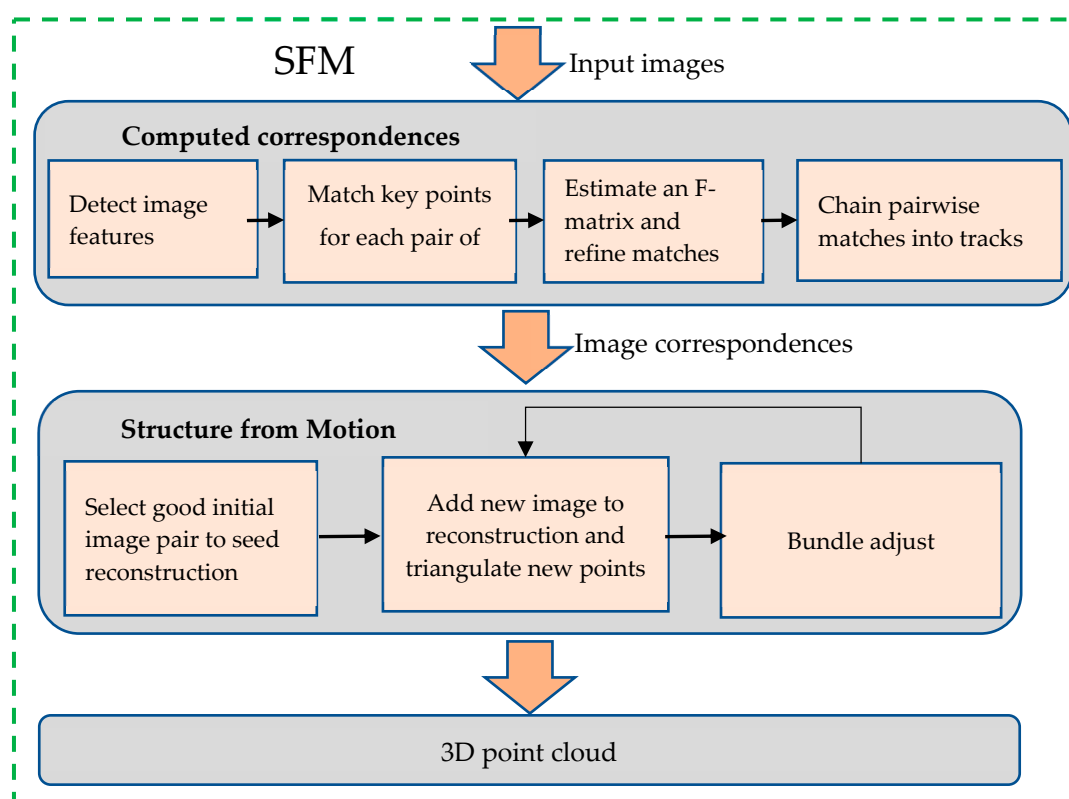
During the measurement process, the tassel is fixed at the center of the calibration board. The digital cameras and tripod positions are secured, and the focal length is

adjusted to ensure that the calibration board occupies more than 70% of the field of view. Additionally, the overlap of the tassel images in the upper, middle, and lower sections is also maintained at 70% or more.

The data acquisition software is then activated, setting the height and photo intervals for the three cameras before automatically collecting 360-degree image data of the tassel. A total of 1227 effective tassel images were obtained, resulting in 88,344 images in total. The entire process took 81.8 h.

### 2.3. Point Cloud Generation and Preprocessing

After image acquisition, the images need to undergo 3D point cloud reconstruction using the SFM (Structure from Motion) algorithm. As shown in Figure 2. The process of using the SFM algorithm to generate the 3D point cloud of the maize tassel is as follows. Firstly, the SFM algorithm extracts the focal length information from the input images. Subsequently, feature extraction algorithms such as SIFT are employed to extract image features. Then, the kd-tree model is utilized to calculate the Euclidean distance between the feature points of two images for performing feature point matching, with the aim of identifying the image pairs that possess the requisite number of feature point matches.



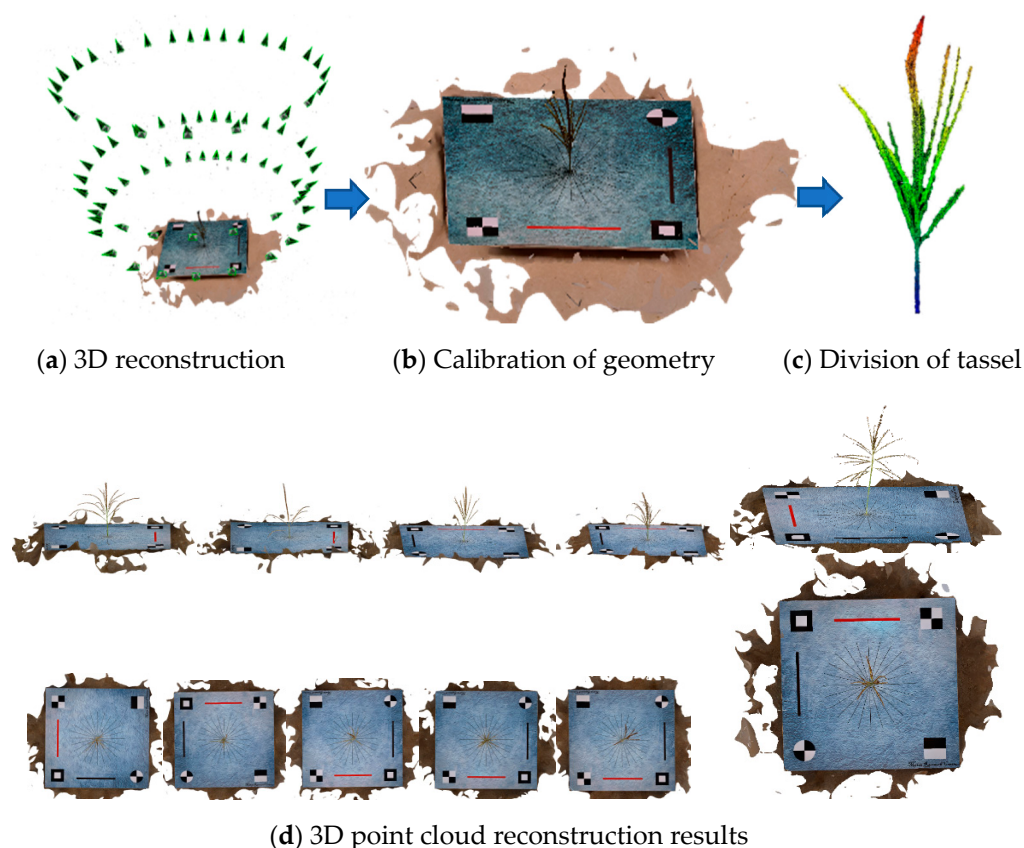
**Figure 2.** The process of SFM.

For each of these image matching pairs, the epipolar geometry is computed, and the F matrix is estimated and optimized through the RANSAC algorithm to enhance the quality of the matching pairs. If a feature point can be detected and matched within a matching pair and is continuously detectable, a matching point trajectory will be formed.

Next, the new images are incorporated into the SFM loop for bundle adjustment. This process continues until there are no more suitable images to be added. Finally, the sparse bundle adjustment was adopted in this study to generate the 3D point cloud of the maize tassel.

This algorithm is suitable for various types of scenes and objects, whether indoor or outdoor or static objects or dynamic scenes [43,44].

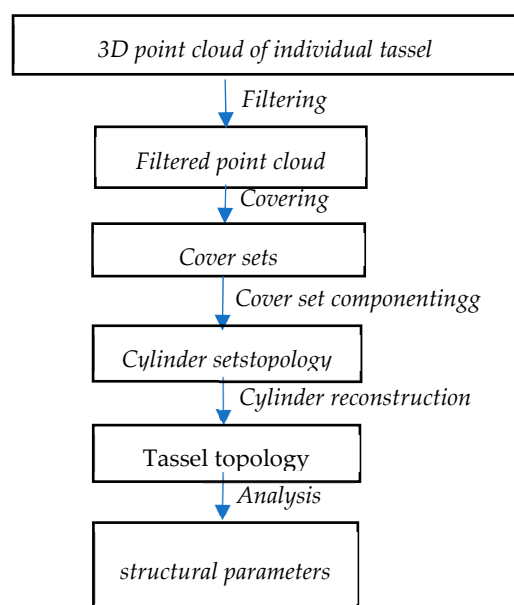
In this study, we used the domestic commercial software DJI Terra (Shenzhen DJI Technology Co., Ltd., Shenzhen, China, version 3.6.8) to batch reconstruct 3D point cloud data in .ply format. These data include not only the X, Y, and Z coordinates but also RGB color information. Figure 3 had showed the point cloud reconstruction process and some of the point cloud reconstruction results, with all 1227 samples successfully reconstructed.



**Figure 3.** Preprocessing process of image data.

After generating the point cloud, the spatial information of the 3D reconstruction does not agree with the physical values, and further calibration is required to accurately quantify the spike phenotype parameters. As shown in Figure 3b, the calibration was achieved by designing red and black calibration lines on the calibration board with a standard length of 25 cm, and the point cloud coordinates could be calibrated to the absolute coordinate system by the coefficient of variation between the predicted and actual values that was measured by constructing the ratio between the actual length of the calibration line (25 cm) and the values extracted by TreeQSM.

As shown in Figure 4, first, noise and isolated points in the point cloud are filtered out. Subsequently, the filtered point cloud is segmented into small sets that conform to the tassel surface, and the adjacent relationship of these coverage sets is defined. Next, sets not belonging to the tassel are removed to define tassel components. Tassel components refer to essentially independent parts or clusters in the point cloud, like a single branch, a group of branches, or even the entire tassel.



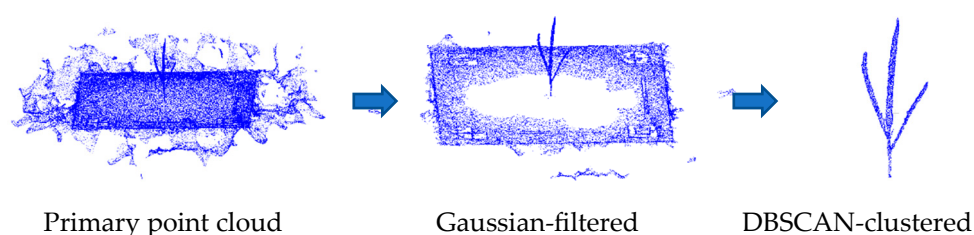
**Figure 4.** Flowchart of TreeQSM.

Following this, a surface growth model is employed during the segmentation of tassel components. By checking local connectivity, branches are identified. After that, each part is approximated as a sequence of cylindrical bodies, which may have varying radii, lengths, and directions.

Finally, based on the constructed cylindrical model, the tassel's topology and other feature parameters are calculated.

Since the extraction of point cloud parameters requires data containing only the object of interest, further filtering and clustering of the point cloud are necessary. As shown in Figure 5, the raw point cloud undergoes Gaussian filtering and DBSCAN clustering to obtain a point cloud containing only the tassel itself.

The Gaussian filtering process is performed directly in MATLAB 2017b using the `pcdenoise` (`noisyData`, "NumComponents", `numComponents`) function, with `numComponents` set to 0.1. After filtering, the DBSCAN (`eps` = 0.5, `min_samples` = 150) clustering function is used to extract the individual tassel point cloud. The process is illustrated in Figure 5.



**Figure 5.** Tassel division.

The preprocessing results indicate that out of the total 1227 tassel samples, 1152 were successfully isolated, achieving a success rate of 93.9%. The remaining samples were separated manually.

#### 2.4. Extraction of Phenotypic Parameters for Maize Tassel

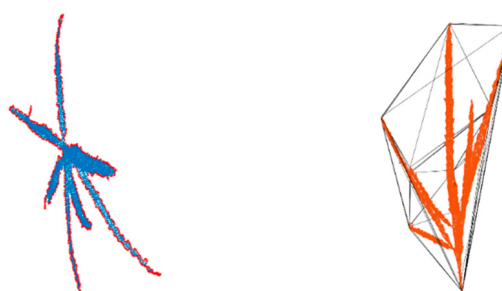
The phenotypic information extraction utilizes the TreeQSM algorithm and the Convex Hull algorithm. The metrics extracted by these methods are listed in Table 2. The TreeQSM algorithm has been widely applied for extracting the 3D structure of trees

[45,46] and has achieved excellent results. Currently, this algorithm is available in various software versions [47–50].

**Table 2.** The phenotypic indices extracted by different algorithms.

| Algorithms  | Extracted Indicators |
|-------------|----------------------|
| TreeQSM     | Lt                   |
|             | Lb(all)              |
|             | Branchnum            |
|             | Lc                   |
|             | La                   |
|             | Ld                   |
|             | Lm                   |
|             | Lp                   |
| Convex hull | Crowndiam(AVG)       |
|             | Crownarea            |
|             | CrownVolume          |

Previous studies have shown that the spatial variability along the x, y, and z axes of the maize tassel topology point clouds reconstructed using the TreeQSM algorithm and images acquired from this system is stable, demonstrating the high accuracy of the original point clouds. Additionally, the correlation coefficient between the predicted and measured values of multiple phenotypic parameters has an average absolute coefficient  $R^2 > 0.9$  [30]. It is necessary to remove the length of the stem while obtaining the length of the main spike. TreeQSM can record the coordinates of the lowest point during the reconstruction of the topological structure (usually the lowest point coordinates of the stem) and the coordinates of the first branch, as well as the category, length, and radius of all cylinder fitting segments. Subtract the coordinates of the lowest point from the coordinates of the first branch to obtain the length of the stem, and then obtain the length of the main spike. For the diameter of the main spike and the diameter of the stem, it is the average of all fitted cylinder diameters on the main spike and stem. This study used TreeQSM version 2.4.1. TreeQSM has five input parameters: PatchDiam1, PatchDiam2Min, PatchDiam2Max, BallRad1, and BallRad2. In this study, PatchDiam1, PatchDiam2Min, and PatchDiam2Max were adjusted to obtain the 3D topological structure of maize tassels and extract related phenotypic parameters [51]. Detailed parameter adjustment rules can be found in reference documents [52]. The point cloud spatial scale results extracted using the TreeQSM algorithm are non-physical values and require scale correction to ensure that the extracted phenotype parameter results are physical values with standard quantization units. As shown in Table 2, the convex hull method was used to extract the concave hull crown area (Figure 6a) and the convex hull volume (Figure 6b) of the tassel on the XOY projection plane. The crown area and volume were directly extracted using the boundary function in MATLAB (version 2017b), with the third parameter, representing the degree of looseness, set to 1.



(a) Area of concave projection      (b) Convex hull volume

**Figure 6.** Convex hull volume and convex hull crown area of tassel.

### 2.5. Methods of Cluster Analysis

The introduction mentions that the subjects of this study are maize materials from three genetic types (NSS, TST, and SS). The classification of these maize materials is based on genotype probabilities inferred using the STRUCTURE software to determine population structure and assign genotypes to subpopulations.

If the probability of any one genotype (NSS, TST, or SS) exceeds 60%, the material is classified into that genotype category. If none of the proportions exceed 60%, the material is classified into the MIXED category. Since the classification is based on probabilities, the study uses GFC to cluster the 1227 maize tassel samples based on the structural phenotypic parameters extracted using the TreeQSM model, aiming to identify the differences in tassel structure phenotypes among different maize genotypes.

This study uses the GFC model to perform clustering analysis on maize tassel phenotypic parameters. GFC is a probabilistic model that has been widely applied in various fields such as pattern recognition, computer vision, machine learning, data mining, and bioinformatics [53–55]. GFC combines the concepts of GFC. GFC allows data points to belong to multiple clusters with different degrees of membership rather than being hard-assigned to a single cluster. GFC uses Gaussian functions to define the membership function, resulting in fuzzy memberships. Specifically for this study, the clustering process and key variables are as follows and in Table 3:

- (1) Initialize the original data point matrix  $X$  and randomly generate the initial affiliation matrix  $U$ . Note that the initialized affiliation matrix must satisfy  $\sum_{k=1}^{11} u_{ik} = 1$ , where  $i$  is the row number, and  $k$  is the column number, from 1 to 11. Normalize the affiliation of each data point so that the sum of its affiliations to all clusters equals 1. Normalize the affiliation of each data point using the formula  $\frac{u_{ik}}{\sum_{j=1}^{11} u_{ij}}$  so that the sum of its affiliations to all clusters equals 1. Here, the normalization still results in the matrix  $U$ .

$$X = \begin{bmatrix} 7.5 & 26.0 & \dots & 16.9 & 254.9 \\ 8.0 & 34.3 & \dots & 81.7 & 1243.4 \\ \dots & \dots & \dots & \dots & \dots \\ 11.5 & 33.5 & \dots & 51.7 & 4615.8 \\ 11.6 & 26.0 & \dots & 48.5 & 1361.4 \end{bmatrix}$$

(Ld, Lt, Branchnum, Lm, Lp, Lc, La, Crownarea, Crowndiam (AVG), Lb (all), and CrownVolume)

$$U = \begin{bmatrix} 0.11 & 0.07 & \dots & 0.04 & 0.04 \\ 0.04 & 0.17 & \dots & 0.06 & 0.11 \\ \dots & \dots & \dots & \dots & \dots \\ 0.05 & 0.07 & \dots & 0.15 & 0.03 \\ 0.10 & 0.07 & \dots & 0.13 & 0.08 \end{bmatrix}$$

- (2) The next step is to calculate the clustering center  $v_k$  for each phenotypic parameter, where  $k$  is the number of features (here, 11), and the fuzzy coefficient  $m$  usually takes the value of 2.

$$v_k = \frac{\sum_{i=1}^{1127} u_{ik}^m x_i}{\sum_{i=1}^{1127} u_{ik}^m}$$

The cluster centroids are obtained by calculating as follows.

$$v_1 = \begin{bmatrix} 0.99 \\ 2.86 \\ 0.81 \\ 0.46 \\ 1.60 \\ 0.05 \\ 0.09 \\ 14.53 \\ 0.20 \\ 1.14 \\ 110.52 \end{bmatrix} \quad v_2 = \begin{bmatrix} 0.77 \\ 2.54 \\ 1.19 \\ 0.56 \\ 1.57 \\ 0.05 \\ 0.07 \\ 16.14 \\ 0.24 \\ 1.77 \\ 136.21 \end{bmatrix} \quad \dots \quad v_{10} = \begin{bmatrix} 1.04 \\ 2.91 \\ 1.04 \\ 0.88 \\ 1.34 \\ 0.06 \\ 0.10 \\ 16.84 \\ 0.38 \\ 3.97 \\ 305.34 \end{bmatrix} \quad v_{11} = \begin{bmatrix} 0.36 \\ 1.20 \\ 0.41 \\ 0.22 \\ 0.73 \\ 0.02 \\ 0.03 \\ 6.45 \\ 0.09 \\ 1.65 \\ 49.42 \end{bmatrix}$$

- (3) Compute the Euclidean distance  $d_{ik}$  from each data point  $x_i$  to each cluster center  $v_k$ . The calculation process and results are as follows:

$$d_{ik} = ||x_i - v_k||$$

$$d_{11} = \sqrt{(7.5 - 0.99)^2 + (26.0 - 2.86)^2 + \dots (254.9 - 110.52)^2} = 185.6$$

$$d_{21} = \sqrt{(8.0 - 0.77)^2 + (34.3 - 2.54)^2 + \dots (1243.4 - 136.21)^2} = 1257.4$$

.....

$$d_{1127-10} = \sqrt{(11.5 - 1.04)^2 + (33.5 - 2.91)^2 + \dots (46.15.8 - 305.34)^2} = 4385.3$$

$$d_{1127-11} = \sqrt{(11.6 - 0.36)^2 + (26.0 - 1.20)^2 + \dots (1361.4 - 49.42)^2} = 1142.5$$

- (4) Update the affiliation matrix based on the distance and fuzzy coefficients  $U$ , where each element in the  $U$  matrix is calculated as follows:

$$u_{ik} = \frac{1}{\sum_{j=1}^4 \left( \frac{d_{ik}}{d_{ij}} \right)^{\frac{2}{m-1}}}$$

where  $i$  and  $k$  denote the number of rows and columns, respectively,  $j$  denotes the number of classifications, and  $m$  is the fuzzy coefficient, which still takes the value of 2 here.

$$u_{11} = \frac{1}{1 + \left( \frac{185.6}{175.3} \right) + \left( \frac{185.6}{134.2} \right) + \left( \frac{185.6}{212.6} \right)} = 0.23$$

Use the new post  $U_{11} \dots U_{1127-11}$  to update the affiliation matrix  $U$ . Repeat processes (2) to (4) until the affiliation matrix  $U$  converges or reaches the set maximum number of iterations, i.e., 2000.

**Table 3.** Definitions of the main variables in GFC.

| Variable Name  | Symbol | Definition                                                                                                                                                 |
|----------------|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sample DataSet | $X$    | Consists of $d$ -dimensional sample points, that is, $X = \{x_1, x_2 \dots x_n\}$ , where $d$ is a $d$ -dimensional vector representing a sample point.    |
| Cluster Center | $v_k$  | Represents the center of the $k$ -th cluster and is a $d$ -dimensional vector used to represent the typical position of this cluster in the feature space. |

|                       |          |                                                                                                                                                                                             |
|-----------------------|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Membershi<br>p Matrix | $U$      | It is a $n \times c$ matrix, and the element $u_{ik}$ represents the membership degree of the sample point $x_i$ belonging to cluster $k$ , satisfying $0 \leq u_{ik} \leq 1$ for any $i$ . |
| Fuzziness<br>Index    | $d_{ik}$ | Each data point $x_i$ to each cluster center $v_k$ . When the distance is less than the set threshold, the point is assigned to the current category.                                       |

Since the genotype classification of maize tassel materials in this study is based on probabilities, the GFC algorithm is well suited for handling different genotypes of tassel materials.

This study used the Random Forest algorithm to analyze the importance of extracted maize tassel phenotype parameters, attempting to compare them with GFC classification results and verify the correctness of the results. Using the Random Forest model, the importance of the 11 extracted tassel phenotypic parameters was ranked. With a fixed random seed of 42, 70% of the data was randomly selected as the training set and 30% as the validation set.

### 2.6. Accuracy Evaluation

We use accuracy, precision, Recall, and F1 to judge the accuracy and effectiveness of the model clustering, calculated as follows:

$$\text{Accuracy} = (\text{Tp} + \text{Tn}) / (\text{Tp} + \text{Tn} + \text{Fp} + \text{Fn}) \quad (1)$$

$$\text{Precision} = \text{Tp} / (\text{Tp} + \text{Fp}) \quad (2)$$

$$\text{Recall} = \text{Tp} / (\text{Tp} + \text{Fn}) \quad (3)$$

$$\text{F1} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

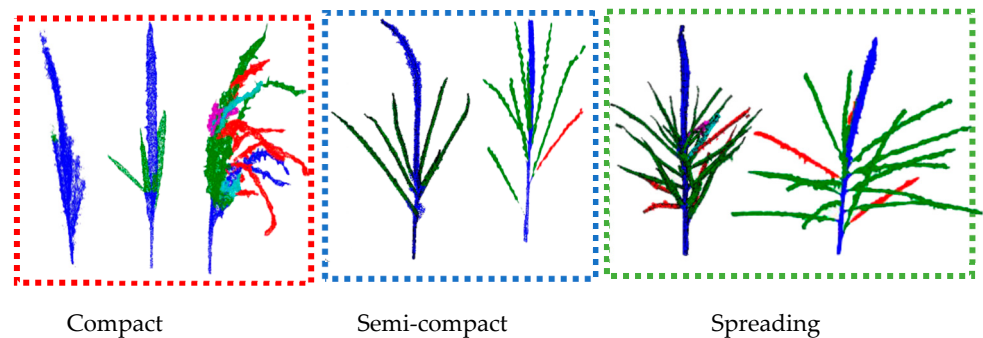
where Tp is predicted as 1 and is actually 1; Fp is predicted as 1 but is actually 0; Fn is predicted as 0 but is actually 1; Tn is predicted as 0 and is actually 0.

## 3. Results

### 3.1. Reconstruction Results of the 3D Topological Structure of the Tassel

In this study, the TreeQSM algorithm was optimized by adjusting three parameters (PatchDiam1, PatchDiam2Min, and PatchDiam2Max) to achieve the best reconstruction results. Initially, the parameters were set to 0.005, 0.0005, and 0.005, respectively, resulting in successful 3D topological reconstructions for 986 samples, with a success rate of 80.3%. By further adjusting the parameters, the number of successfully reconstructed tassels increased to 1194, yielding a final success rate of 97.3%. Unsuccessful reconstructions were mainly due to excessively compact tassel shapes, which made it challenging to correctly separate the main spike from the branches. Such data were excluded from subsequent statistical analysis.

Figure 7 shows the results of the TreeQSM reconstruction for some tassel topological structures. The results indicate that the TreeQSM model performs less effectively on tassels with a higher compactness. There are mainly two cases: one where the main stem is identified but cannot be distinctly separated from the branches and another where the main stem is completely failed to be identified. Both cases are considered unsatisfactory extraction results. In contrast, for tassels with lower compactness, the TreeQSM model effectively identifies the main spike and branches, providing the necessary conditions for the further extraction of tassel structural phenotype parameters.

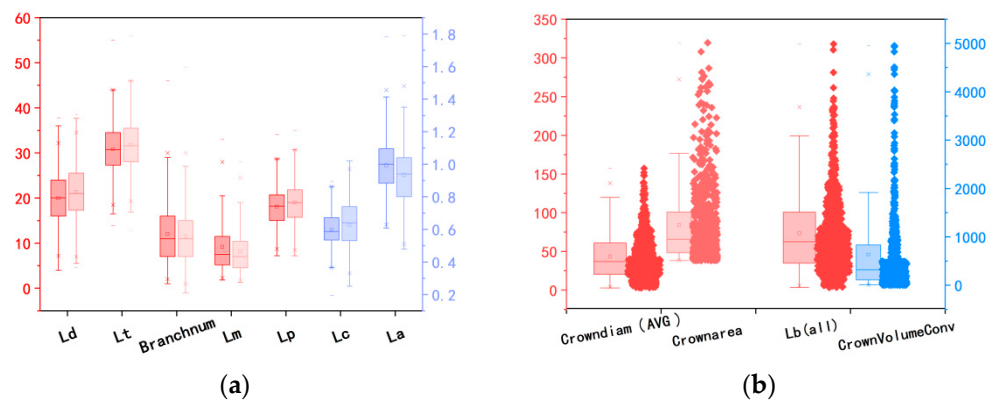


**Figure 7.** Results of the TreeQSM topology reconstruction.

### 3.2. Variability Analysis of Point Cloud and Extraction Results

Parameters extracted using the TreeQSM model are still in Euclidean distance and need to be corrected. The correction process and the accuracy of the results have been verified [30], and the absolute coefficient of variation  $P$  of the dataset in this study is between 34 and 39, but more than 84% of the samples have coefficients of variation between 36 and 38, which suggests that the spatial variability of the point clouds obtained under the same conditions is relatively small.

This study analyzed the data distribution of the 11 extracted parameters, as illustrated in the Figure 8. All phenotypic parameters, except for the number of branches, are measured in centimeters. The left side of each set of comparison data in Figure 6a is the measured value, and the right side is the model extracted value. The left four sets of data correspond to the scale labels on the left, while the right two sets of data correspond to the scale labels on the right. From the Figure 6, it is evident that the median, average, minimum, and maximum values of the stem length and main spike length are all larger than those measured. This is primarily because the TreeQSM model provides more detailed length calculations, whereas the measured data, although stretched straight, could not be fully straightened to avoid breaking the tassel. Additionally, the maximum crown diameter, crown height, stem diameter, and main spike diameter show considerable discrepancies between measured and extracted values. This is mainly due to the difficulty in accurately measuring the maximum crown diameter and crown height and the smaller values of stem and main spike diameters, which are prone to larger measurement errors. The average crown diameter, total branch length, crown area, and convex hull volume are all results extracted using TreeQSM. Notably, the average crown diameter is displayed as ten times its extracted value. The left vertical stem represents the red boxplot data scale, with units in cm for average crown diameter and total branch length and  $\text{cm}^2$  for crown area. The right vertical stem, in blue, represents the convex hull volume with units in  $\text{cm}^3$ . As shown in Figure 6, the average values for these four phenotypic parameters are significantly higher than the median values, indicating that more than 50% of the sample values are distributed below the average.



**Figure 8.** Data distribution.

Additionally, this study analyzed the variability of the 11 extracted parameters, as shown in Table 4. The results indicate that the extracted phenotypic parameters exhibit considerable variability, reflecting a non-concentrated data distribution and providing a more representative basis for the study's results.

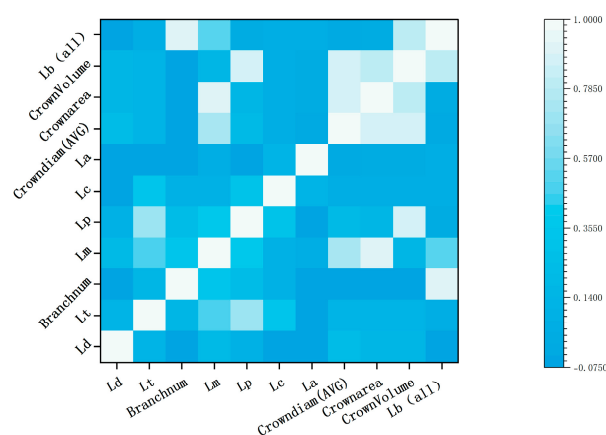
**Table 4.** Extracted phenotypic parameter variability results.

| Parameters      | Variability (CV) |
|-----------------|------------------|
| Ld              | 26.8%            |
| Lt              | 16.9%            |
| Branchnum       | 55.6%            |
| Lm              | 63.4%            |
| La              | 18.3%            |
| Crowndiam (AVG) | 64.6%            |
| Crownarea       | 58.2%            |
| Lb (all)        | 67.1%            |
| Lp              | 23.7%            |
| Lc              | 16.9%            |
| CrownVolume     | 82.9%            |

Note: The coefficient of variation (CV) is the ratio of the sample standard deviation to the mean value.

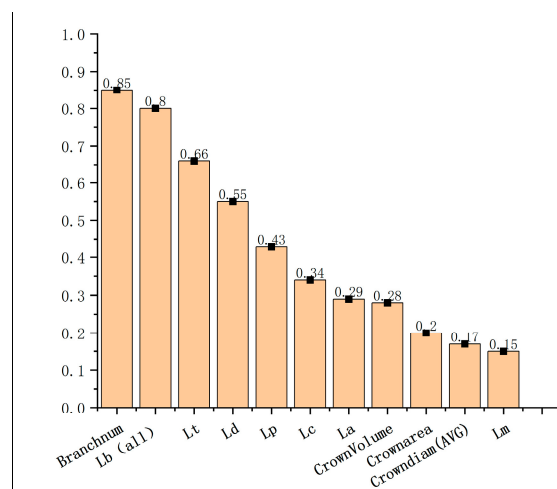
### 3.3. Analysis Results of Phenotypic Parameter Correlations and Importance

Using the Pearson correlation coefficient method, the correlations between the 11 phenotypic parameters of maize tassels extracted by TreeQSM were constructed. As shown in Figure 9, the highest correlations were found between the number of tassel branches and the total branch length, as well as between the crown area and the convex hull volume, both exceeding 0.9. The average crown diameter showed the next highest correlation with the crown area and convex hull volume, both above 0.8. In contrast, the correlations among main spike diameter, stem diameter, and stem length were relatively lower.



**Figure 9.** Correlation of phenotypic parameters.

The importance of 11 maize phenotypic parameters calculated using the Random Forest algorithm is shown in Figure 10. It can be observed that for the classification of maize genotypes, the top three most important phenotypic parameters are the number of branches (i.e., Branchnum), total branch length (i.e., Lb (all)), and main spike length (i.e., Lt). The least important parameters are the convex hull volume of the crown (i.e., CrownVolume), the average diameter of the crown (i.e., Crowndiam (AVG)), and the projected area of the crown (i.e., Crownarea).



**Figure 10.** Ranking of importance of phenotypic parameters.

### 3.4. Different Genotype Classification Results

This study performed probabilistic clustering using GFC on 10 out of the 11 extracted phenotypic parameters of maize tassels (excluding convex hull volume). The number of clusters was set to 4, and the number of iterations was set to 2000, calculating the probability of each sample belonging to each cluster. Each sample was assigned a probability of belonging to a particular cluster, with probabilities  $\geq 0.6$  used to classify the samples into NSS, SS, TST, or MIXED categories. The results are shown in Table 4. The columns represent the four clustering results of the GFC, while the rows indicate the number of samples for each genotype in the actual clustering results.

As shown in Table 5, the numbers in the table are the actual number of genotype samples in each category of classification results, and the order of the data volume of the four classification clusters is also consistent with the sample volume, which is NSS, TST, MIXED, and SS in descending order. In the third and the fourth clusters, the classification

accuracies of NSS and TST are 67.7% and 78.5%, respectively, and the clustering accuracies of NSS are lower than that of TST. One important reason for this is that the variability of phenotypic traits is greater in the material of the NSS genotype than in the TST. In addition, there is a MIXED category in the classification results, the probabilities of NSS and TST are close to each other and neither of them exceeds 0.6, which brings a certain degree of randomness to the distribution of NSS and TST. This also directly affects the ability of current classification methods to further improve the accuracy of NSS and TST classification.

**Table 5.** Results of GFC.

| Actual Results         |            |       |     |    |     |              |               |            |      |
|------------------------|------------|-------|-----|----|-----|--------------|---------------|------------|------|
|                        | Genotyping | MIXED | NSS | SS | TST | Accuracy (%) | Precision (%) | Recall (%) | F1   |
| Classification results | SS         | 10    | 76  | 21 | 26  | 85.34        | 15.79         | 25.00      | 0.19 |
|                        | MIXED      | 23    | 49  | 13 | 20  | 81.24        | 21.90         | 13.94      | 0.17 |
|                        | NSS        | 82    | 372 | 34 | 63  | 72.19        | <b>67.71</b>  | 70.86      | 0.69 |
|                        | TST        | 50    | 18  | 16 | 311 | 83.84        | <b>78.53</b>  | 74.05      | 0.76 |
|                        | Total      | 165   | 525 | 84 | 420 |              |               |            |      |

### 3.5. Comparison of Clustering Results for Different Phenotypic Parameters

In this study, the 11 extracted phenotypic parameters were randomly combined and then clustered. The parameter combinations are shown in Table 6, with a total of seven different combinations, labeled A to G. Each subsequent combination has one fewer parameter than the previous one. The GFC was again used for probabilistic clustering, with the samples divided into four clusters and classified based on probabilities  $\geq 0.6$ . Notably, combinations F and G, which had fewer parameters, were excluded because the GFC could no longer perform accurate clustering with these combinations.

**Table 6.** Different parameter combinations and clustering results.

| Group                  | Parameter Combinations                                                               |       |     |    |     |              |               |            |      |
|------------------------|--------------------------------------------------------------------------------------|-------|-----|----|-----|--------------|---------------|------------|------|
| A                      | Ld, Lt, Branchnum, Lm, Lp, Lc, La, Lb (all), Crowndiam (AVG), Crownarea, CrownVolume |       |     |    |     |              |               |            |      |
| B                      | Ld, Lt, Branchnum, Lm, Lp, Lc, La, Lb (all), Crowndiam (AVG)                         |       |     |    |     |              |               |            |      |
| C                      | Ld, Lt, Branchnum, Lm, Lp, Lc, La, Lb (all)                                          |       |     |    |     |              |               |            |      |
| D                      | Ld, Lt, Branchnum, Lm, Lp, Lc, La                                                    |       |     |    |     |              |               |            |      |
| E                      | Ld, Lt, Branchnum, Lm, Lp, Lc                                                        |       |     |    |     |              |               |            |      |
| F                      | Branchnum, Lm, Lp                                                                    |       |     |    |     |              |               |            |      |
| G                      | Lt, Lm, Lp                                                                           |       |     |    |     |              |               |            |      |
| (A)                    |                                                                                      |       |     |    |     |              |               |            |      |
| Actual Results         |                                                                                      |       |     |    |     |              |               |            |      |
|                        | Genotyping                                                                           | MIXED | NSS | SS | TST | Accuracy (%) | Precision (%) | Recall (%) | F1   |
| Classification results | SS                                                                                   | 3     | 63  | 22 | 28  | 86.93        | 18.97         | 26.19      | 0.22 |
|                        | MIXED                                                                                | 8     | 18  | 19 | 85  | 76.63        | 6.15          | 4.85       | 0.05 |
|                        | NSS                                                                                  | 148   | 305 | 30 | 23  | 64.74        | 60.28         | 58.10      | 0.59 |
|                        | TST                                                                                  | 6     | 139 | 13 | 284 | 75.38        | 64.25         | 67.62      | 0.66 |
|                        | Total                                                                                | 165   | 525 | 84 | 420 |              |               |            |      |
| (B)                    |                                                                                      |       |     |    |     |              |               |            |      |
| Actual Results         |                                                                                      |       |     |    |     |              |               |            |      |
|                        | Genotyping                                                                           | MIXED | NSS | SS | TST | Accuracy (%) | Precision (%) | Recall (%) | F1   |
| Classification         | SS                                                                                   | 11    | 51  | 0  | 29  | 85.34        | 0.00          | 0.00       |      |

|         |       |     |            |    |            |       |              |              |      |
|---------|-------|-----|------------|----|------------|-------|--------------|--------------|------|
| results | MIXED | 23  | 74         | 0  | 36         | 78.89 | 17.29        | 13.94        | 0.15 |
|         | NSS   | 99  | <b>271</b> | 64 | 57         | 60.30 | <b>55.19</b> | <b>51.62</b> | 0.53 |
|         | TST   | 32  | 129        | 20 | <b>288</b> | 73.79 | <b>61.41</b> | <b>68.57</b> | 0.65 |
|         | Total | 165 | 525        | 84 | 420        |       |              |              |      |

(C)

| Actual Results         |            |       |            |    |            |              |               |            |      |
|------------------------|------------|-------|------------|----|------------|--------------|---------------|------------|------|
|                        | Genotyping | MIXED | NSS        | SS | TST        | Accuracy (%) | Precision (%) | Recall (%) | F1   |
| Classification results | SS         | 27    | 54         | 0  | 20         | 84.51        | 0.00          | 0.00       |      |
|                        | MIXED      | 35    | 86         | 0  | 34         | 79.06        | 22.58         | 21.21      | 0.22 |
|                        | NSS        | 94    | <b>264</b> | 68 | 76         | 58.21        | <b>52.59</b>  | 50.29      | 0.51 |
|                        | TST        | 9     | 101        | 16 | <b>290</b> | 78.56        | <b>63.71</b>  | 69.05      | 0.69 |
|                        | Total      | 165   | 525        | 84 | 420        |              |               |            |      |

(D)

| Actual Results         |            |       |            |    |            |              |               |            |      |
|------------------------|------------|-------|------------|----|------------|--------------|---------------|------------|------|
|                        | Genotyping | MIXED | NSS        | SS | TST        | Accuracy (%) | Precision (%) | Recall (%) | F1   |
| Classification results | SS         | 27    | 92         | 15 | 34         | 81.41        | 8.93          | 17.86      | 0.12 |
|                        | MIXED      | 21    | 61         | 0  | 45         | 79.06        | 16.54         | 12.73      | 0.14 |
|                        | NSS        | 70    | <b>230</b> | 69 | 157        | 50.50        | <b>43.73</b>  | 43.81      | 0.44 |
|                        | TST        | 47    | 142        | 0  | <b>181</b> | 64.15        | <b>48.92</b>  | 43.10      | 0.46 |
|                        | Total      | 165   | 525        | 84 | 420        |              |               |            |      |

(E)

| Actual Results         |            |       |            |    |            |              |               |            |      |
|------------------------|------------|-------|------------|----|------------|--------------|---------------|------------|------|
|                        | Genotyping | MIXED | NSS        | SS | TST        | Accuracy (%) | Precision (%) | Recall (%) | F1   |
| Classification results | SS         | 33    | 86         | 0  | 30         | 80.49        | 0.00          | 0.00       |      |
|                        | MIXED      | 6     | 108        | 15 | 50         | 72.19        | 3.35          | 3.64       | 0.03 |
|                        | NSS        | 82    | <b>217</b> | 31 | 156        | 51.68        | <b>44.65</b>  | 41.33      | 0.43 |
|                        | TST        | 64    | 114        | 38 | <b>184</b> | 62.14        | <b>46.00</b>  | 43.81      | 0.45 |
|                        | Total      | 165   | 525        | 84 | 420        |              |               |            |      |

(F)

| Actual Results         |            |       |     |    |     |              |               |            |      |
|------------------------|------------|-------|-----|----|-----|--------------|---------------|------------|------|
|                        | Genotyping | MIXED | NSS | SS | TST | Accuracy (%) | Precision (%) | Recall (%) | F1   |
| Classification results | SS         | 43    | 104 | 0  | 49  | 76.55        | 0.00          | 0.00       |      |
|                        | MIXED      | 23    | 116 | 0  | 14  | 77.22        | 15.03         | 13.94      | 0.14 |
|                        | NSS        | 72    | 215 | 66 | 193 | 46.31        | <b>39.38</b>  | 40.95      | 0.40 |
|                        | TST        | 27    | 90  | 18 | 92  | 61.22        | <b>40.53</b>  | 21.90      | 0.28 |
|                        | Total      | 165   | 525 | 84 | 420 |              |               |            |      |

(G)

| Actual Results         |            |       |            |    |            |              |               |            |      |
|------------------------|------------|-------|------------|----|------------|--------------|---------------|------------|------|
|                        | Genotyping | MIXED | NSS        | SS | TST        | Accuracy (%) | Precision (%) | Recall (%) | F1   |
| Classification results | SS         | 11    | 126        | 8  | 86         | 74.96        | 3.46          | 9.52       | 0.05 |
|                        | MIXED      | 12    | 63         | 0  | 35         | 78.98        | 10.91         | 7.27       | 0.09 |
|                        | NSS        | 89    | <b>209</b> | 61 | 168        | 46.90        | <b>39.66</b>  | 39.81      | 0.40 |
|                        | TST        | 53    | 127        | 15 | <b>131</b> | 59.46        | <b>40.18</b>  | 31.19      | 0.35 |
|                        | Total      | 165   | 525        | 84 | 420        |              |               |            |      |

From the clustering results of different parameters (Table 6), it is evident that the precision for the SS and MIXED categories remains relatively low. The clustering accuracy

for the NSS and TST genotypes decreases progressively with the reduction in the number of parameters. In particular, a significant drop in accuracy is observed from combinations C to D, indicating that branch length contributes significantly to genotype classification and that its variability is greater. As the number of clustering parameters decreases, the accuracy further declines.

## 4. Discussion

### 4.1. Comparison of Results from Different Clustering Methods

The results obtained from different clustering methods vary. In this study, the clustering effects of three common machine learning methods, namely Support Vector Machine (SVM), Random Forest, and Backpropagation (BP) Neural Network, as well as three commonly used algorithms, namely, k-Means, HCM (Hierarchical) Clustering, and FCM (Fuzzy C-Means) Clustering, were compared. As concluded in previous research, using 10 phenotypic parameters to cluster into four categories can achieve the highest accuracy. Therefore, the same 10 phenotypic parameters were used in this study to compare the results. The sample data were randomly divided into 20 groups of training sets and validation sets at a ratio of 7:3. The clustering precision was calculated based on the proportion of correctly classified samples to the total number. The clustering results are shown in Table 7. When comparing these three machine learning algorithms, Random Forest has the highest accuracy on both the training set and the validation set, but its advantage is not significant. Its accuracy is between 50% and 60%, and the relatively higher accuracy on the training set is also due to the overfitting phenomenon. Judging from the accuracy results, the accuracies of all these three methods are not high, and the differences among them are not significant either. This is because all these three methods are supervised classification methods and are not suitable for the non-linear problem of clustering maize phenotypes and genotypes.

**Table 7.** The clustering results based on machine learning methods.

| Variety Number | RF                     |                             | SVM                    |                             | BPNN                   |                             |
|----------------|------------------------|-----------------------------|------------------------|-----------------------------|------------------------|-----------------------------|
|                | Train Set Accuracy (%) | Validation Set Accuracy (%) | Train Set Accuracy (%) | Validation Set Accuracy (%) | Train Set Accuracy (%) | Validation Set Accuracy (%) |
| 1              | 93.24                  | 46.48                       | 61.90                  | 57.7                        | 54.17                  | 53.52                       |
| 2              | 88.49                  | 53.52                       | 54.76                  | 49.30                       | 61.90                  | 53.52                       |
| 3              | 90.10                  | 54.93                       | 57.14                  | 50.70                       | 60.71                  | 56.34                       |
| 4              | 97.11                  | 50.70                       | 52.98                  | 47.89                       | 60.12                  | 46.48                       |
| 5              | 86.99                  | 49.30                       | 60.71                  | 49.30                       | 55.36                  | 49.30                       |
| 6              | 88.24                  | 57.75                       | 62.50                  | 50.7                        | 47.62                  | 57.75                       |
| 7              | 96.75                  | 56.34                       | 57.74                  | 42.25                       | 59.52                  | 43.66                       |
| 8              | 96.81                  | 57.75                       | 64.29                  | 39.40                       | 55.95                  | 56.34                       |
| 9              | 95.49                  | 57.75                       | 60.12                  | 46.48                       | 55.36                  | 60.56                       |
| 10             | 88.24                  | 59.15                       | 56.55                  | 53.52                       | 56.55                  | 52.11                       |
| 11             | 96.11                  | 52.11                       | 58.33                  | 45.07                       | 53.57                  | 45.07                       |
| 12             | 91.05                  | 54.93                       | 61.90                  | 52.11                       | 60.12                  | 46.48                       |
| 13             | 86.92                  | 52.11                       | 61.30                  | 45.07                       | 41.67                  | 42.25                       |
| 14             | 90.73                  | 54.93                       | 60.71                  | 50.70                       | 59.93                  | 52.64                       |
| 15             | 92.94                  | 47.89                       | 57.14                  | 49.30                       | 57.14                  | 49.30                       |
| 16             | 85.61                  | 47.89                       | 60.71                  | 39.44                       | 51.79                  | 52.11                       |
| 17             | 93.67                  | 50.71                       | 63.11                  | 52.10                       | 52.38                  | 53.52                       |
| 18             | 92.68                  | 52.11                       | 55.36                  | 45.07                       | 62.55                  | 52.11                       |
| 19             | 88.35                  | 49.30                       | 53.57                  | 56.34                       | 55.90                  | 38.03                       |

|                           |              |              |              |              |              |              |
|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 20                        | 92.94        | 50.71        | 64.29        | 57.75        | 56.55        | 43.66        |
| <b>Average</b>            | <b>91.73</b> | <b>52.82</b> | <b>59.26</b> | <b>49.01</b> | <b>55.94</b> | <b>50.24</b> |
| <b>Standard Deviation</b> | <b>3.67</b>  | <b>3.75</b>  | <b>3.46</b>  | <b>5.32</b>  | <b>5.03</b>  | <b>5.81</b>  |

By comparing the clustering results of the other three methods shown in Table 8, it can be found that their clustering precisions are all 6–13% lower than that of the GFC. The possible reason is that the GFC has the characteristics of fuzziness, probabilistic nature, and sensitivity to normally distributed data, while the classification rules for maize material genotypes used in this study are also based on probability, and the sample data also present the characteristics of approximate normal distribution.

**Table 8.** The clustering results based on the K-Means, HCM, and FCM.

| <b>Actual Results</b> |                   |              |            |           |            |                     |                      |                   |           |
|-----------------------|-------------------|--------------|------------|-----------|------------|---------------------|----------------------|-------------------|-----------|
|                       | <b>Genotyping</b> | <b>MIXED</b> | <b>NSS</b> | <b>SS</b> | <b>TST</b> | <b>Accuracy (%)</b> | <b>Precision (%)</b> | <b>Recall (%)</b> | <b>F1</b> |
| K-Means               | SS                | 11           | 24         | 0         | 16         | 88.72               | 0.00                 | 0.00              |           |
|                       | MIXED             | 65           | 109        | 15        | 68         | 75.53               | 25.25                | 39.34             | 0.31      |
|                       | NSS               | 72           | 229        | 46        | 184        | 49.95               | <b>43.16</b>         | 43.62             | 0.43      |
|                       | TST               | 18           | 163        | 23        | 152        | 60.48               | <b>42.72</b>         | 36.19             | 0.39      |
| HCM                   | SS                | 24           | 30         | 0         | 5          | 88.02               | 0.00                 | 0.00              |           |
|                       | MIXED             | 22           | 24         | 0         | 16         | 84.67               | 35.48                | 13.33             | 0.19      |
|                       | NSS               | 84           | 302        | 61        | 252        | 53.10               | <b>47.26</b>         | 57.52             | 0.52      |
|                       | TST               | 35           | 169        | 23        | 147        | 63.15               | <b>47.70</b>         | 49.29             | 0.48      |
| FCM                   | SS                | 3            | 42         | 0         | 11         | 88.27               | 0.00                 | 0.00              |           |
|                       | MIXED             | 54           | 109        | 15        | 89         | 72.86               | 20.22                | 32.73             | 0.25      |
|                       | NSS               | 76           | 302        | 46        | 68         | 65.41               | <b>61.38</b>         | 57.52             | 0.59      |
|                       | TST               | 32           | 72         | 23        | 252        | 75.29               | <b>66.49</b>         | 60.00             | 0.63      |
| Total                 |                   | 165          | 525        | 84        | 420        |                     |                      |                   |           |

#### 4.2. Clustering Results Under Different Clustering Modes

In the previous results section, this study clustered the data into four types. However, the “mixed” type does not belong to a definite genetic type of maize materials. This study explores the removal of the “mixed” category. The original MIXED samples were re-assigned to SS, NSS, or TST based on the highest proportion probability. The results are shown in Table 9. The redistributed sample numbers were NSS: 605, SS: 133, and TST: 456. The new samples were then subjected to GFC using the 10 parameters. The precision of classifying NSS was 55.7% in the first clustering result, TST genotype precision was 52.9% in the second clustering result, and SS precision was 18.9% in the third clustering result. Compared to the four-category classification, the precision significantly decreased.

**Table 9.** Clustering results of different classification methods.

| <b>Actual Results</b>  |                   |            |           |            |                     |                      |                   |           |
|------------------------|-------------------|------------|-----------|------------|---------------------|----------------------|-------------------|-----------|
|                        | <b>Genotyping</b> | <b>NSS</b> | <b>SS</b> | <b>TST</b> | <b>Accuracy (%)</b> | <b>Precision (%)</b> | <b>Recall (%)</b> | <b>F1</b> |
| Classification results | NSS               | <b>393</b> | 76        | 237        | <b>66.33</b>        | 29.92                | 282.14            | 0.54      |
|                        | TST               | 122        | 23        | <b>163</b> | <b>77.89</b>        | 35.57                | 73.94             | 0.48      |
|                        | SS                | 90         | <b>34</b> | 56         | 32.58               | 9.77                 | 6.48              | 0.08      |
|                        | Total             | 605        | 133       | 456        |                     |                      |                   |           |

#### 4.3. Comparison with Clustering Results of Measured Values

In this study, eight real-measured quantitative phenotypic traits other than maize tassel crown area, convex bunt volume, and average crown diameter (data that could not be measured realistically) were analyzed according to the clustering; the data samples were still the same as the 1194 groups used by the GFC algorithm, and the results are shown in Table 10. Here, it can be seen that the clustering accuracies of real-measured data were all improved compared to the accuracies of the automated clustering of this phenotyping platform, of which the clustering accuracies of the NSS genotypes were improved, NSS by 5.64% and TST by 10.86% (the classification accuracy of the system in this study was 51.62% and 68.57%), which indicates that the clustering effect of maize tassel genotypes using the methodology of this study is trustworthy and has some reference value.

**Table 10.** Comparison of clustering results between predicted and measured results.

| Data Type           | Actual Results |       |            |    |            | Data Type     | Actual Results |            |    |            |               |
|---------------------|----------------|-------|------------|----|------------|---------------|----------------|------------|----|------------|---------------|
|                     | Genotyping     | MIXED | NSS        | SS | TST        |               | MIXED          | NSS        | SS | TST        | Precision (%) |
| Classification data | SS             | 11    | 51         | 0  | 29         | Measured data | 13             | 43         | 0  | 19         | 0.00          |
|                     | MIXED          | 23    | 74         | 0  | 36         |               | 20             | 62         | 1  | 29         | 17.86         |
|                     | NSS            | 99    | <b>271</b> | 64 | 57         |               | 108            | <b>334</b> | 66 | 41         | <b>60.84</b>  |
|                     | TST            | 32    | 129        | 20 | <b>288</b> |               | 24             | 86         | 17 | <b>331</b> | <b>72.27</b>  |
|                     | Total          | 165   | 525        | 84 | 420        |               | 165            | 525        | 84 | 420        |               |

## 5. Conclusions

In this study, the TIPS system was used to achieve the automatic acquisition of maize male tassel images, and the SFM algorithm was used to achieve the synthesis from the image sequence to the 3D point cloud, while the Gaussian filter algorithm and the DBSCAN algorithm were used to achieve the de-noising of the original point cloud as well as the separation and extraction of the point cloud of the male tassel of a single plant. Based on the extracted point cloud, the TIPS system was used to implement a system specifically for the extraction of the tree structure using the algorithm TreeQSM for the reconstruction of the 3D point cloud of the corn tassel. The reconstruction of male ear topology, the extraction of multiple structural phenotypic parameters, and finally the analysis of the extracted structural phenotypic parameters with the genotypes of maize materials by the GFC clustering algorithm, which realizes the technical analysis of the whole set of processes from phenotypic parameter extraction to genotype classification, were explored in this study. The main contributions include the following:

- The system developed in this study has high efficiency and accuracy for extracting the structural phenotypes of maize tassel with high precision, especially the use of the combination of Gaussian filtering and DBSCAN algorithms to achieve the separation of the point cloud of the tassel of a single plant proved to be very effective.
- For the classification of genotypes of maize materials, the parameter importance was Branchnum > Lb (all) > Lt > Ld > Lp > Lc > La > Lm > Crownarea > Crowndiam (AVG) > CrownVolume, which revealed the magnitude of the variability of the tassel phenotypic parameters of the different genotypes of the maize materials.
- Compared with the traditional RF, SVM, and BPNN methods based on supervised classification, the GFC algorithm, an unsupervised classification method, separated NSS and TST maize genotypes more efficiently, with accuracies of 67.7% and 78.5%.

- (d) Comparing the clustering results of the measured data and the predicted data, although the clustering accuracy of the measured data is 5–10% higher than that of the predicted data, the method in this study has a higher economic and practical value.
- (e) Compared with the traditional process of classifying maize materials by using gene locus information, which is complicated, difficult to operate, and costly, this study has pioneered the first research on clustering maize genotypes using the phenotypic parameters of maize tassels and has proven that its results are more efficient, economical, easy to operate, and highly accurate. The fuzziness, probabilistic nature, and sensitivity to the normal distribution data of the Gaussian Fuzzy Clustering algorithm are in line with the classification rules for maize material genotypes and the characteristics of the sample data used in this study. Therefore, it has relatively good classification accuracy.
- (f) Phenotype is the result of gene action. The variability of the branch number, branch length, and main spike length, as well as the clustering accuracy results in this study, fully demonstrate that the heritability of tassel phenotypic traits in different maize genotypes is relatively strong.

In this study, we investigated the trait variation of different genotypes from the perspective of the three-dimensional structure and phenotype of maize tassel, which provides an important tool for the efficient screening of maize materials.

**Author Contributions:** Conceptualization, B.X. and G.Y.; methodology, B.X.; software, B.X.; validation, H.Y. and H.F.; investigation, B.X., H.Y. and H.F.; data curation, B.X. and C.L.; writing—original draft preparation, B.X., X.Y. and H.Y.; writing—review, editing, visualization, B.X., Y.Z., H.Y. and H.F.; supervision, C.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This article was funded by the National Key Research and Development Program of China (No. 2021YFD2000100, No. 2022YFF1003500, and No. 2021YFD1201601) and supported by the Collaborative Innovation Center for Modern Crop Production, which is co-sponsored by the Province and Ministry, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China; the earmarked fund was CARS-02, Special Fund for Construction of Scientific and Technological Innovation Ability of Beijing Academy of Agriculture and Forestry Sciences (KJCX20230434). We are also grateful to the anonymous reviewers and editors for their valuable comments and recommendations.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Gebbers, R.; Adamchuk, V.I. Precision agriculture and food security. *Science* **2010**, *327*, 828–831. <https://doi.org/10.1126/science.1183899>.
2. Tester, M.; Langridge, P. Breeding technologies to increase crop production in a changing world. *Science* **2010**, *327*, 818–822. <https://doi.org/10.1126/science.1183700>.
3. Shakoor, N.; Lee, S.; Mockler, T.C. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Curr. Opin. Plant Biol.* **2017**, *38*, 184–192. <https://doi.org/10.1016/j.pbi.2017.05.006>.
4. Han, D.; Yang, G.; Yang, H.; Qiu, C.; Chen, M.; Wen, W.; Niu, Q.; Yang, W. *Three Dimensional Information Extraction from Maize Tassel Based on Stereoscopic Vision*; Chinese Society of Agricultural Engineering: Beijing, China, 2018.
5. Geraldi, I.; Miranda Filho, J.; Vencovsky, R. Estimates of genetic parameters for tassel characters in maize (*Zea mays* L.) and breeding perspectives. *Maydica* **1985**, *30*, 1–14.
6. Liu, J.; Bai, Z.; Wu, W. Evolution of main ear traits of maize hybrids and their effects on breeding objectives. *Maize Sci.* **2004**, *12* (Suppl. S2), 3–4.

7. Xu, G.; Wang, X.; Huang, C.; Xu, D.; Li, D.; Tian, J.; Chen, Q.; Wang, C.; Liang, Y.; Wu, Y. Complex genetic architecture underlies maize tassel domestication. *New Phytol.* **2017**, *214*, 852–864. <https://doi.org/10.1111/nph.14400>.
8. Berke, T.; Rocheford, T. Quantitative trait loci for tassel traits in maize. *Crop Sci.* **1999**, *39*, 1439–1443. <https://doi.org/10.2135/cropsci1999.3951439x>.
9. Upadyayula, N.; Da Silva, H.; Bohn, M.O.; Rocheford, T. Genetic and QTL analysis of maize tassel and ear inflorescence architecture. *Theor. Appl. Genet.* **2006**, *112*, 592–606. <https://doi.org/10.1007/s00122-005-0133-x>.
10. Zhou, Y.; Kusmec, A.; Mirnezami, S.V.; Attigala, L.; Srinivasan, S.; Jubery, T.Z.; Schnable, J.C.; Salas-Fernandez, M.G.; Ganapathysubramanian, B.; Schnable, P.S. Identification and utilization of genetic determinants of trait measurement errors in image-based, high-throughput phenotyping. *Plant Cell* **2021**, *33*, 2562–2582. <https://doi.org/10.1093/plcell/koab134>.
11. Wang, X.; Li, J.; Han, L.; Liang, C.; Li, J.; Shang, X.; Miao, X.; Luo, Z.; Zhu, W.; Li, Z. QTG-Miner aids rapid dissection of the genetic base of tassel branch number in maize. *Nat. Commun.* **2023**, *14*, 5232. <https://doi.org/10.1038/s41467-023-41022-1>.
12. Xhu, B.; Liu, F.; Che, Y.; Hui, F.; Ma, Y. Three-dimensional quantification of intercropping crops in field by ground and aerial photography. In Proceedings of the 2018 6th International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications (PMA), Hefei, China, 4–8 November 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 8–12. <https://doi.org/10.1109/PMA.2018.8611616>.
13. Zhang, X.; Huang, C.; Wu, D.; Qiao, F.; Li, W.; Duan, L.; Wang, K.; Xiao, Y.; Chen, G.; Liu, Q. High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. *Plant Physiol.* **2017**, *173*, 1554–1564. <https://doi.org/10.1104/pp.16.01516>.
14. Das Choudhury, S.; Samal, A.; Awada, T. Leveraging image analysis for high-throughput plant phenotyping. *Front. Plant Sci.* **2019**, *10*, 508. <https://doi.org/10.3389/fpls.2019.00508>.
15. Das Choudhury, S.; Maturu, S.; Samal, A.; Stoerger, V.; Awada, T. Leveraging image analysis to compute 3D plant phenotypes based on voxel-grid plant reconstruction. *Front. Plant Sci.* **2020**, *11*, 521431. <https://doi.org/10.3389/fpls.2020.521431>.
16. Hartmann, A.; Czauderna, T.; Hoffmann, R.; Stein, N.; Schreiber, F. HTPheno: An image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinform.* **2011**, *12*, 148. <https://doi.org/10.1186/1471-2105-12-148>.
17. Gage, J.L.; Miller, N.D.; Spalding, E.P.; Kaeppler, S.M.; de Leon, N. TIPS: A system for automated image-based phenotyping of maize tassels. *Plant Methods* **2017**, *13*, 21. <https://doi.org/10.1186/s13007-017-0172-8>.
18. Klukas, C.; Chen, D.; Pape, J.-M. Integrated analysis platform: An open-source information system for high-throughput plant phenotyping. *Plant Physiol.* **2014**, *165*, 506–518. <https://doi.org/10.1104/pp.113.233932>.
19. Gibbs, J.A.; Pound, M.; French, A.P.; Wells, D.M.; Murchie, E.; Pridmore, T. Approaches to three-dimensional reconstruction of plant shoot topology and geometry. *Funct. Plant Biol.* **2016**, *44*, 62–75. <https://doi.org/10.1071/FP16167>.
20. Li, J.; Tang, L. Developing a low-cost 3d plant morphological traits characterization system. *Comput. Electron. Agric.* **2017**, *143*, 1–13. <https://doi.org/10.1016/j.compag.2017.09.025>.
21. Yang, W.; Guo, Z.; Huang, C.; Duan, L.; Chen, G.; Jiang, N.; Fang, W.; Feng, H.; Xie, W.; Lian, X. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* **2014**, *5*, 5087. <https://doi.org/10.1038/ncomms6087>.
22. Guo, W.; Fukatsu, T.; Ninomiya, S. Automated characterization of flowering dynamics in rice using field-acquired time-series rgb images. *Plant Methods* **2015**, *11*, 7. <https://doi.org/10.1186/s13007-015-0047-9>.
23. Lu, H.; Cao, Z.; Xiao, Y.; Fang, Z.; Zhu, Y.; Xian, K. Fine-grained maize tassel trait characterization with multi-view representations. *Comput. Electron. Agric.* **2015**, *118*, 143–158. <https://doi.org/10.1016/j.compag.2015.08.027>.
24. Kükenbrink, D.; Schneider, F.D.; Leiterer, R.; Schaepman, M.E.; Morsdorf, F. Quantification of hidden canopy volume of airborne laser scanning data using a voxel traversal algorithm. *Remote Sens. Environ.* **2017**, *194*, 424–436. <https://doi.org/10.1016/j.rse.2016.10.023>.
25. Kunz, M.; Hess, C.; Raunonen, P.; Bienert, A.; Hackenberg, J.; Maas, H.; Härdtle, W.; Fichtner, A.; Von Oheimb, G. Comparison of wood volume estimates of young trees from terrestrial laser scan data. *iFor. Biogeosci. For.* **2017**, *10*, 451–458. <https://doi.org/10.3832/for2151-010>.
26. Thapa, S.; Zhu, F.; Walia, H.; Yu, H.; Ge, Y. A novel LiDAR-based instrument for high-throughput, 3D measurement of morphological traits in maize and sorghum. *Sensors* **2018**, *18*, 1187. <https://doi.org/10.3390/s18041187>.
27. Paulus, S.; Behmann, J.; Mahlein, A.-K.; Plümer, L.; Kuhlmann, H. Low-cost 3d systems: Suitable tools for plant phenotyping. *Sensors* **2014**, *14*, 3001–3018. <https://doi.org/10.3390/s140203001>.
28. Chaivivatrakul, S.; Tang, L.; Dailey, M.N.; Nakarmi, A.D. Automatic morphological trait characterization for corn plants via 3d holographic reconstruction. *Comput. Electron. Agric.* **2014**, *109*, 109–123. <https://doi.org/10.1016/j.compag.2014.09.005>.

29. Wu, S.; Wen, W.; Gou, W.; Lu, X.; Zhang, W.; Zheng, C.; Xiang, Z.; Chen, L.; Guo, X. A miniaturized phenotyping platform for individual plants using multi-view stereo 3D reconstruction. *Front. Plant Sci.* **2022**, *13*, 897746. <https://doi.org/10.3389/fpls.2022.897746>.
30. Xu, B.; Wan, X.; Yang, H.; Feng, H.; Fu, Y.; Cen, H.; Wang, B.; Zhang, Z.; Li, S.; Zhao, C. TIPS: A three-dimensional phenotypic measurement system for individual maize tassel based on TreeQSM. *Comput. Electron. Agric.* **2023**, *212*, 108150. <https://doi.org/10.1016/j.compag.2023.108150>.
31. Song, D. Clustering Study of Gene Expression Data Based on Gaussian Mixture Model. Ph.D. Thesis, Jiangsu University, Zhenjiang, China, 2015.
32. Minton, R.L. Morphometrics Suggest Reduced Diversity in the Freshwater Snail Genus *Gyrotoma* Shuttleworth, 1845 (Gastropoda: Pleuroceridae). *Ann. Carnegie Mus.* **2021**, *87*, 79–88. <https://doi.org/10.2992/007.087.0103>.
33. Yang, M.-S.; Chang-Chien, S.-J.; Nataliani, Y. Unsupervised fuzzy model-based Gaussian clustering. *Inf. Sci.* **2019**, *481*, 1–23. <https://doi.org/10.1016/j.ins.2018.12.059>.
34. Fuchs, R.; Pommeret, D.; Viroli, C. Mixed Deep Gaussian Mixture Model: A clustering model for mixed datasets. *Adv. Data Anal. Classif.* **2022**, *16*, 31–53. <https://doi.org/10.1007/s11634-021-00466-3>.
35. Yang, X.; Gao, S.; Xu, S.; Zhang, Z.; Prasanna, B.M.; Li, L.; Li, J.; Yan, J. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* **2010**, *28*, 511–526. <https://doi.org/10.1007/s11032-010-9500-7>.
36. Pan, C.-X.; Hu, Y.; Ji, H.-B.; Li, Y.-M.; Chen, N.-L. Phenotypic diversity and clustering analysis of watermelon germplasm. *J. Plant Genet. Resour.* **2015**, *16*, 59–63.
37. Xie, Z.; Chen, S.; Gao, G.; Li, H.; Wu, X.; Meng, L.; Ma, Y. Evaluation of rapeseed flowering dynamics for different genotypes with UAV platform and machine learning algorithm. *Precis. Agric.* **2022**, *23*, 1688–1706. <https://doi.org/10.1007/s11119-022-09904-4>.
38. Han, L.; Yang, G.; Dai, H.; Yang, H.; Xu, B.; Feng, H.; Li, Z.; Yang, X. Fuzzy clustering of maize plant-height patterns using time series of UAV remote-sensing images and variety traits. *Front. Plant Sci.* **2019**, *10*, 926. <https://doi.org/10.3389/fpls.2019.00926>.
39. Rincón Sánchez, F.; Johnson, B.; Crossa, J.; Taba, S. Cluster analysis, an approach to sampling variability in maize accessions. *Agric. Food Sci.* **1996**, *41*, 4.
40. Shrestha, J. Cluster analysis of maize inbred lines. *J. Nepal Agric. Res. Counc.* **2016**, *2*, 33–36. <https://doi.org/10.3126/jnarc.v2i0.16119>.
41. Pritchard, J.K.; Stephens, M.; Rosenberg, N.A.; Donnelly, P. Association mapping in structured populations. *Am. J. Hum. Genet.* **2000**, *67*, 170–181. <https://doi.org/10.1086/302959>.
42. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **2003**, *164*, 1567–1587. <https://doi.org/10.1093/genetics/164.4.1567>.
43. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *32*, 1362–1376. <https://doi.org/10.1109/TPAMI.2009.161>.
44. Wu, C.; Zhan, S. Three-dimensional face modeling of SFM algorithm based on gabor feature. *J. Hefei Univ. Technol. Nat. Sci. Ed.* **2017**, *40*, 180–185. <https://doi.org/10.7666/d.Y2799194>.
45. Dassot, M.; Colin, A.; Santenoise, P.; Fournier, M.; Constant, T. Terrestrial laser scanning for measuring the solid wood volume, including branches, of adult standing trees in the forest environment. *Comput. Electron. Agric.* **2012**, *89*, 86–93. <https://doi.org/10.1016/j.compag.2012.08.005>.
46. Calders, K.; Newnham, G.; Burt, A.; Murphy, S.; Raumonen, P.; Herold, M.; Culvenor, D.; Avitabile, V.; Disney, M.; Armston, J. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods Ecol. Evol.* **2015**, *6*, 198–208. <https://doi.org/10.1111/2041-210X.12301>.
47. Raumonen, P.; Casella, E.; Calders, K.; Murphy, S.; Åkerblom, M.; Kaasalainen, M. Massive-scale tree modelling from TLS data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 189–196. <https://doi.org/10.5194/isprsannals-II-3-W4-189-2015>.
48. Raumonen, P.; Kaasalainen, M.; Åkerblom, M.; Kaasalainen, S.; Kaartinen, H.; Vastaranta, M.; Holopainen, M.; Disney, M.; Lewis, P. Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sens.* **2013**, *5*, 491–520. <https://doi.org/10.3390/rs5020491>.
49. Hackenberg, J.; Spiecker, H.; Calders, K.; Disney, M.; Raumonen, P. SimpleTree—An efficient open source tool to build tree models from TLS clouds. *Forests* **2015**, *6*, 4245–4294. <https://doi.org/10.3390/f6114245>.
50. Delagrangé, S.; Jauvin, C.; Rochon, P. PyeTree: A tool for reconstructing tree perennial tissues from point clouds. *Sensors* **2014**, *14*, 4271–4289. <https://doi.org/10.3390/s140304271>.

51. Moorthy, S.M.K.; Raunonen, P.; Van den Bulcke, J.; Calders, K.; Verbeeck, H. Terrestrial laser scanning for non-destructive estimates of liana stem biomass. *For. Ecol. Manag.* **2020**, *456*, 117751. <https://doi.org/10.1016/j.foreco.2019.117751>.
52. Hackenberg, J.; Morhart, C.; Sheppard, J.; Spiecker, H.; Disney, M. Highly accurate tree models derived from terrestrial laser scan data: A method description. *Forests* **2014**, *5*, 1069–1105. <https://doi.org/10.3390/f5051069>.
53. Wei, X. EM Algorithm and Its Initialization for Parameter Estimation of Gaussian Mixture Models. *Electron. Testing.* **2011**, *6*, 26–30. <https://doi.org/10.3969/j.issn.1000-8519.2011.06.008>.
54. Xing, C.; Zhao, Q.; Wang, X. Research on accelerated EM algorithm based on robust Gaussian mixture model. *Comput Appl Res* **2017**, *4*, 1042–1046.
55. Wu, T. Application of Gaussian Mixture Model Based on EM Algorithm in Iris Dataset. *Netw. Secur. Technol. Appl.* **2022**, *4*, 47–49.

**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.