Research on 3D Reconstruction Method of Fruit Trees Based on Camera Pose Recovery and Neural Radiation Field Theory

Huiyan Wang
Shandong University of Technology
Jianhang Wang
Shandong University of Technology
Binxiao Liu
Shandong University of Technology
Jinliang Gong
Shandong University of Technology
Yanfei Zhang
1392076@sina.com

Shandong University of Technology

Research Article

Keywords: Fruit Tree, 3D Reconstruction, Neural Radiation Field Theory, Camera Pose Recovery

Posted Date: June 10th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4469719/v1

Additional Declarations: No competing interests reported.

Research on 3D Reconstruction Method of Fruit Trees Based on Camera Pose Recovery and Neural Radiation Field Theory

Huiyan Wang[1], Jianhang Wang[1†], Binxiao Liu[2†], Jinliang Gong [2*†] , Yanfei Zhang[1*†]

[1]School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000， China

[2]School of Mechanical Engineering, Shandong University of Technology, Zibo 255000， China

*Corresponding authors. E-mails: 84374294@qq.com; 1392076@sina.com;
Contributing authors: 19862513603@163.com; 1837299354@qq.com;
2541403392@qq.com; †These authors contributed equally to this work.

Abstract

A method integrating camera pose recovery techniques with neural radiation field theory is proposed in this study to address issues such as detail loss and color distortion encountered by traditional stereoscopic vision-based 3D reconstruction techniques when dealing with fruit trees exhibiting high-frequency phenotypic details. The high cost of information acquisition devices equipped with image pose recording functionality necessitates a cost-effective approach for fruit tree information gathering while enhancing the resolution and detail capture capability of the resulting 3D models. To achieve this, a device and scheme for capturing multi-view image sequences of fruit trees are designed. Firstly, the target fruit tree is surrounded by a multi-angle video capture using the information acquisition platform, and the resulting video undergoes image enhancement and frame extraction to obtain a multi-view image sequence of the fruit tree. Subsequently, a motion recovery structure algorithm is employed for sparse reconstruction to recover image poses. Then, the image sequence with pose data is inputted into a multi-layer perceptron, utilizing ray casting for coarse and fine two-layer granularity sampling to calculate volume density and RGB information, thereby obtaining the neural radiation field 3D scene of the fruit tree. Finally, the 3D scene is converted into point clouds to derive a high-precision point cloud model of the fruit tree. Using this reconstruction method, a crabapple tree including multiple periods such as flowering, fruiting, leaf fall, and dormancy is reconstructed, capturing the neural radiation field scenes and point cloud models. Reconstruction results demonstrate that the 3D scenes of the neural radiation field in each period exhibit real-world level representation. The point cloud models derived from the 3D scenes achieve millimeter-level precision at the organ scale, with

tree structure accuracy exceeding 96% for multi-period point cloud models, averaging 97.79% accuracy across all periods. This reconstruction method exhibits robustness across various fruit tree periods and can meet the requirements for 3D reconstruction of fruit trees in most scenarios.

Key words: Fruit Tree, 3D Reconstruction, Neural Radiation Field Theory, Camera Pose Recovery

1 Introduction

In recent years, computer technologies such as cloud computing, big data, and the Internet of Things (IoT) have emerged as hot topics in academic research, with rapid advancements in each domain accelerating the transition of human society towards an era of intelligence(ZHOU et al 2012). The establishment of digital orchards using computer technologies can assist in the management and operation of fruit orchards, thereby liberating rural productivity and enhancing operational efficiency in fruit cultivation. Digital orchards represent a concrete manifestation of the concept of digital agriculture within the context of fruit orchard management, representing a new research direction in modern information technology applied to fruit cultivation(Wu et al 2019).

The digital fruit tree is an essential tool for establishing digital orchards(Zhou et al 2018), guiding unmanned agricultural machinery in tasks such as identification, localization(Mai et al 2015), path planning, visual obstacle avoidance(Zhao et al 2022), harvesting(Chen et al 2023), pruning(KOLMANIC S et al 2017,Zhang et al 2020), etc. Three-dimensional reconstruction of fruit trees constitutes the foundational and pivotal step in constructing digital fruit tree(NARVAEZ F Y et al 2017). Fruit trees possess numerous intricate detail features, complex spatial structures, and intertwined foliage, necessitating robust high-frequency detail representation capabilities in fruit tree three-dimensional reconstruction techniques to accurately preserve the phenotypic parameters of fruit trees. Algorithms for registering crown-level three-dimensional reconstructions of fruit trees using depth cameras have been proposed by Feng et al(Feng et al 2013)and Zhou et al(Zhou et al 2013). Chéné et al(Chéné Y et al 2012)utilized low-cost depth cameras to capture crown depth images of various plants, proposing a segmentation algorithm for plant depth images and conducting three-dimensional reconstructions of plant crowns. Adhikari et al(Adhikari B et al 2011)conducted data collection in outdoor orchard environments using the CamCube3.0 camera, establishing three-dimensional models of dormant apple trees, with this depth camera integrating a background light suppression technology suitable for outdoor environments. Xiong et al(Xiong et al 2019)conducted three-dimensional reconstructions of merged fruit tree point clouds using the Poisson surface reconstruction algorithm, further enhancing modeling accuracy of fruit trees. Scholars both domestically and abroad have made considerable efforts in enhancing the accuracy and detail representation capabilities of digital orchards. However, the aforementioned studies still exhibit shortcomings in representing high-frequency details of canopy foliage, thereby struggling to establish complex detail features in fruit tree models.

To enhance the representation capability of three-dimensional reconstruction techniques for high-frequency details, MILDENHALL et al(Mildenhall B et al 2021)proposed a method based on implicit neural rendering called Neural Radiance Fields (NeRF). This three-dimensional reconstruction approach employs neural networks as function approximators and utilizes implicit functions to represent scenes, rendering it more applicable to parts of scenes with high-frequency details such as fruit tree canopies, thereby improving the accuracy of fruit tree three-dimensional models and phenotype data. Motion recovery structure-multi-view stereopsis is a stereoscopic reconstruction technique highly suitable for fruit tree three-dimensional reconstruction(MILLER J et al 2015, GAO Q et al 2022), widely applied and with the NeRF theory showcasing significant potential in recent years by enhancing training speed and rendering quality(MARTIN-BRUALLA R et al 2021, MüLLER T et al 2022, TANCIK M et al 2023,BARRON J T et al 2021), particularly evident in fruit tree three-dimensional reconstruction.

The study devised equipment and a scheme for fruit tree image data acquisition, selecting a crabapple tree planted on campus as the reconstruction subject. Employing a combined approach of Neural Radiance Fields (NeRF) and Structure from Motion (SFM), termed SFM-NeRF, three-dimensional point cloud models were established for multiple periods throughout the tree's annual lifecycle. The superiority of SFM-NeRF relative to traditional reconstruction methods was assessed.

2 Materials and methods

A crabapple tree planted within the campus premises was selected as the reconstruction subject. Situated in a continental monsoon climate zone, the tree stands approximately 4 meters tall with a crown length of approximately 2.3 meters along its major axis and a width of 1.8 meters along its minor axis. The tree exhibits a complex branch structure, characterized by numerous small flowers and leaves, while the fruits are relatively sparse and often concealed amidst the branches and foliage, as illustrated in Figure 1.

Figure. 1 Reconstruction Subject

2.1 Fruit Tree Video Information Acquisition

In this study, a multi-angle image sequence of the crabapple tree was obtained by capturing videos and then extracting frames. The video acquisition device utilized was a homemade low-cost information collection platform. This platform primarily consists of a three-axis gimbal camera, a ground-based signal receiver, a Wi-Fi video transmission module, and antennas, as depicted in Figure 2. The three-axis gimbal stabilizer ensures the stability of the camera's horizontal and pitch angles, while the camera is equipped with digital stabilization functionality to ensure stable and undistorted images during platform movement. The information collection platform can be paired with a telescopic pole to increase the shooting height, with an adjustable height range of the telescopic pole ranging from 1.5m to 7m. Once connected to the remote controller, the gimbal camera's orientation can be controlled, and real-time video transmission from the collection platform to the remote controller's screen is enabled.

Figure. 2 Information Collection Equipment

During the information collection process, coordination between two individuals, labeled as A and B, is required. Individual A manually operates the collection platform, moving it in a spiral trajectory while gradually ascending, thereby circumnavigating the fruit tree, as illustrated in Figure 3. The spiral trajectory has radii and heights of 1.6m and 1.9m, respectively, with the endpoint coinciding with the starting point on a vertical line, completing one full revolution. Four sets of such spiral trajectories are executed, each originating from one of the four cardinal directions around the tree. Following the completion of the four sets of spiral trajectories, the collection platform is positioned on the telescopic pole, and further circumnavigation around the fruit tree at heights of 2.7m and 4m is conducted, as depicted in Figure 3. Individual B observes the video feed on the remote controller screen and controls the orientation of the gimbal camera during the movement of the collection platform by Individual A, ensuring that the center of the frame consistently faces a point on the tree trunk. This process facilitates the acquisition of videos from three different perspectives: top-down, horizontal, and bottom-up, thereby enabling the convergence of camera poses towards the center throughout the entire collection process. Upon completion of the aforementioned collection procedures, a total of four sets of spiral trajectory videos and two sets of circumnavigation videos are obtained, with a combined duration of approximately 150 seconds.

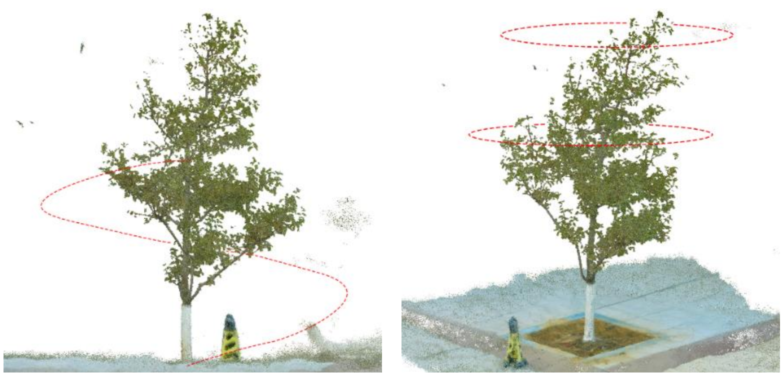

Figure. 3 Helical path and Circumferential path

2.2 Image Enhancement and Frame Extraction

Due to the limited performance of the gimbal camera, the color of the videos captured in 1.1 often deviates from what is perceived by the human eye, as illustrated in Figure 4. Particularly in bright daylight conditions, the captured videos may exhibit partially overexposed and partially underexposed frames. Uneven image exposure results in the loss of some fine details of the fruit tree. Contrast is a parameter that describes the brightness difference between different parts of a video frame. By reducing the video's contrast and adjusting its brightness, the uneven exposure issue can be addressed, thereby enhancing the robustness of the information collection process across different weather conditions. Contrast adjustment also affects the saturation of the video to some extent. Lowering the contrast diminishes the color difference in the image, resulting in a visually dim appearance. Therefore, it is necessary to appropriately increase color saturation to compensate for the decrease in contrast.

Figure. 4 The original image is shown on the left, the enhanced image is shown on the right

Using frame extraction, the processed videos obtained from the aforementioned procedure are converted into an image sequence. During the conversion from video to images, data loss is inevitable. To better preserve the original data of the videos, frame extraction is performed using an equal time interval approach. The six videos captured in 1.1 are concatenated into a single video with a duration of approximately 150 seconds. To ensure the effectiveness of the reconstruction, the repetition rate of the image sequence must be maintained at over 80%. Considering a time interval of approximately 0.25 seconds between adjacent views, the number of frames to be extracted is approximately 600 frames. By extracting one image every 0.25 seconds from this video, a total of 600 evenly extracted images are obtained from all video data.

2.3 Camera Pose Recovery via Sparse Reconstruction based on SFM

As the images obtained through video frame extraction do not contain pose data, camera pose recovery techniques are required to assign poses to the image sequence. In this study, the SFM method is employed for camera pose recovery, the basic principles of which are as follows.

The image sequence obtained through frame extraction exhibits a high degree of content overlap, with adjacent images repetitively recording certain features that are both distinct and morphologically stable. The Scale-Invariant Feature Transform (SIFT) algorithm is employed to detect and convert such parts of the image sequence into feature points. Subsequently, the detected feature points are used to establish correspondences between adjacent images, forming stereo pairs. Among these pairs, the stereo pair with the highest number of matching points and the best matching effect is selected as the initial stereo pair. Based on this stereo pair, a coordinate system is established to describe the poses of other images.

Subsequently, the Random Sample Consensus (RANSAC) algorithm is utilized to select matching points from the initial stereo pairs and to determine the optimal fundamental matrix, **A**. Using fundamental matrix **A** and the camera intrinsic matrices of images a and b, the essential matrix **E**, describing the relative poses between the two images, is computed. Singular Value Decomposition (SVD) of the essential matrix **E** yields the relative exterior matrix **F** of image b with respect to image a, where exterior matrix **F** represents the pose of image b in the coordinate system. Triangulation of the matching points using the exterior matrix **F** of the initial stereo pairs results in a sparse set of three-dimensional points distributed in the coordinate system. Two additional adjacent images are selected as a new stereo pair, and the steps of feature point extraction and matching are repeated to obtain matching points for the new stereo pair. By calculating the correspondence between the matching points of the new stereo pair and the sparse point set, the poses of the images in the new stereo pair are determined. Triangulation of the matching points of the new stereo pair is added to the sparse three-dimensional point set. This process is iteratively repeated by selecting new stereo pairs to expand the three-dimensional structure, resulting in the sparse three-dimensional point cloud of the fruit tree and the pose data of all images.

Upon completion of SFM sparse reconstruction, only the image pose data needs to be recorded for dense three-dimensional reconstruction using NeRF. The sparse three-dimensional point cloud of the fruit tree scene can be discarded to reduce data storage space.

Reprojection error refers to the positional discrepancy between the reprojected points from the sparse reconstructed three-dimensional point cloud onto the image sequence and the corresponding feature points. The quality of camera pose recovery can be assessed by computing the average reprojection error of all points in the three-dimensional sparse point cloud projected onto the respective image planes. Its formula is as follows.

$$R = \frac{\sum_{k=1}^{U}\sum_{j=1}^{V}\left\| \boldsymbol{K}_j \left[N_j \mid \eta_j \right] \boldsymbol{X}_k - x_{kj} \right\|_2}{U \times V} \tag{1}$$

In the equation, U represents the number of points in the sparse point cloud, V represents the total number of images, N_j denotes the intrinsic matrix of the j-th image, $[N_j|\eta_j]$ signifies the extrinsic matrix of the j-th image, X_k stands for the homogeneous coordinates of the k-th point, and X_{kj} represents the coordinates of the k-th point projected onto the j-th image.

2.4 Dense Reconstruction of Crabapple Tree Three-Dimensional Point Cloud Model based on NeRF

Training NeRF for three-dimensional reconstruction using image sequences with pose data yields the NeRF three-dimensional scene of the fruit tree. The specific principle is as follows:

The essence of the neural radiance field is a latent function that records volume density and color information. Initially, this scene is a discrete cloud of fog that cannot represent any details and requires iterative training to converge the latent function to obtain the final three-dimensional scene of the fruit tree. This function is approximated by a multilayer perceptron (MLP), which is a fully connected neural network. The implicit function approximated by MLP can be represented by the following formula.

$$\boldsymbol{F}(x, y, z, \theta, \varphi) \rightarrow (R, G, B, w) \tag{2}$$

Where $F(x,y,z,\theta,\varphi)$ represents the five-dimensional signal composed of the sampling point coordinates (x, y, z) and the camera ray direction (θ, φ), serving as the input vector for the multilayer perceptron (MLP); (R,G,B,w) denotes the four-dimensional signal consisting of the color (R,G,B) and volume density of the sampling point, serving as the output vector for the MLP. The light projection method is a sampling technique where a ray passing through the solid scene of the fruit tree is projected from the camera's optical center to a pixel on the image plane, and sampling is performed along this ray to obtain the sampling point. The specific principle is illustrated in Figure 5

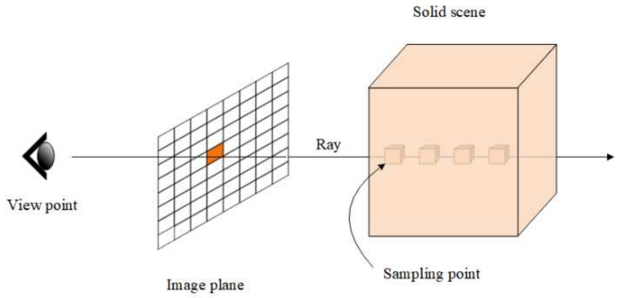

Figure. 5 light projection method

In order to achieve gradual convergence of the latent function during iteration, volume rendering of the color and volume density of the sampling points output by the MLP is necessary. Volume density is used as a weighting factor to sum the sampling points along the camera ray, and integration is performed to calculate the color of the pixels on the pixel plane intersected by the camera ray. The calculation formula is as follows.

$$\begin{cases} \boldsymbol{C}_c(\boldsymbol{r}) = \pi\int_{t_n}^{t_f} T(t)w(r(t))\boldsymbol{C}(\boldsymbol{r}(\boldsymbol{t}))dt \\ T(t) = exp(-\int_{t_n}^{t} w(r(s))ds) \end{cases} \tag{3}$$

Where $C_c(r)$ represents the pixel color; tn is the near point of the viewing frustum, and t_f is the far point of the viewing frustum. $T(t)$ denotes the cumulative transmittance coefficient; $w(r(r(t)))$ is the density of the sampling point at position $r(t)$; $w(r(r(s)))$ is the density of the sampling point at position $r(s)$; and $C(r(t))$ is the color of the sampling point at position $r(t)$.

To reduce computational complexity, alleviate the burden on computer hardware, and enhance rendering efficiency, hierarchical sampling with coarse-to-fine granularity is employed along the camera ray. Initially, coarse-grained sampling is performed to uniformly sample Nc points on the camera ray, which are then inputted into the MLP for volume density rendering, yielding the foundational distribution of volume density. Simultaneously, the coarse-grained rendering of pixel colors, $\hat{C}_c(r)$, is obtained. Subsequently, Nf fine-grained sampling points are extracted from positions with higher volume density, and these points are again inputted into the MLP for volume rendering to acquire the fine-grained rendering of volume density and pixel colors, $\hat{C}f(r)$. The formulas are as follows.

$$\begin{cases} \hat{\boldsymbol{C}}_{\boldsymbol{c}}(\boldsymbol{r}) = \sum_{g=1}^{N_c} w_g \boldsymbol{C}_g \\ \hat{\boldsymbol{C}}_f(\boldsymbol{r}) = \sum_{l=1}^{N_c+N_f} w_l \boldsymbol{C}_l \end{cases} \tag{4}$$

Where N_c and N_f represent the quantities of sampling points for coarse and fine granularity, respectively, w_g and w_l denote the cumulative volume densities of the coarse and fine granularity sampling points, and C_g and C_l represent the colors of the coarse and fine granularity sampling points, respectively.

After obtaining the pixel colors $\hat{C}_c(r)$ and $\hat{C}f(r)$ from coarse and fine granularity rendering, respectively, it is necessary to use a loss function to update the MLP's network weights in reverse, thereby gradually converging the scene to the final three-dimensional fruit tree scene. $\hat{C}_c(r)$ and $\hat{C}f(r)$ are compared with the colors of pixels intersected by the camera ray on the pixel plane to calculate the training loss. After traversing all pixels in an image, the loss function for that image can be obtained. The formula is as follows:

$$L = \sum_{r \in O} [\| \hat{\boldsymbol{C}}_c(\boldsymbol{r}) - \boldsymbol{C}(\boldsymbol{r}) \|_2^2 + \| \hat{\boldsymbol{C}}_f(\boldsymbol{r}) - \boldsymbol{C}(\boldsymbol{r}) \|_2^2] \tag{5}$$

Where Ĉc(r) and Ĉf(r) represent the pixel colors computed from coarse and fine granularity rendering, respectively, and C(r) denotes the color of the pixel along the ray direction. O represents all pixels in a view. After adjusting the network weights through the loss function, and after numerous iterations, continuously updating the volume density and pixel colors within the scene, the converged three-dimensional fruit tree scene can ultimately be obtained. The specific principle is illustrated in Figure 6

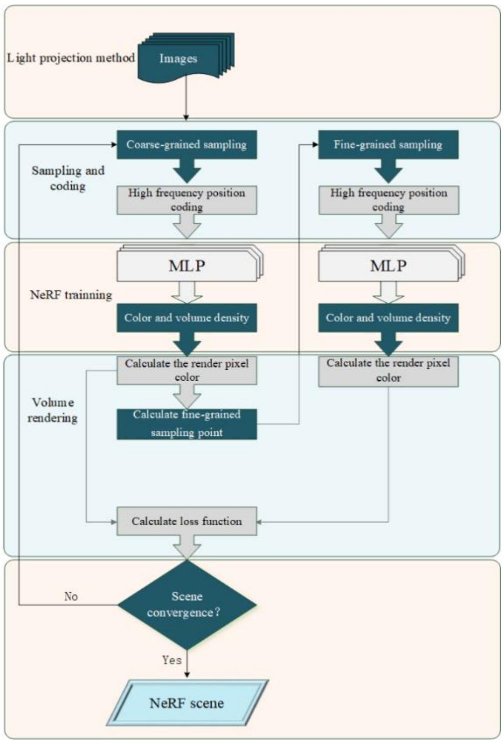

Figure. 6 The specific principle of NeRF

3 Results

This study conducted a total of 50 information collection sessions and 35 3D reconstructions of the tree from February 16, 2023, to March 20, 2024, covering the entire annual life cycle of the tree. Representative 3D reconstructions were selected during the flowering, fruiting, leaf fall, and dormant periods for evaluation of their representation effectiveness in terms of neural radiance field scenes, depth scenes, and point cloud models. A comparison was made between MVS and NeRF in terms of dimensional accuracy. The computer configuration used for training in this study was: CPU: AMD R5-7600X, GPU: NVIDIA GeForce RTX3080, Memory: 32GB.

3.1 Camera Pose Recovery Results

Using the SIFT algorithm to detect and match feature points in the image sequence, a total of 889,162 feature points were obtained, among which 135,067 feature points were mutually matched. The feature point detection results for image number 70 are illustrated in Figure 7, where the red points represent the detected feature points in the image. The feature point matching results between images number 151 and number 209 are depicted in Figure 7, where the endpoints of the green lines denote the mutually matched feature points between the two images. The SFM sparse reconstruction was employed to recover the pose data of the image sequence, yielding a total of 600 image poses with a successful recovery rate of 100%, as depicted in Figure 8. The recovered image poses are organized into four spiral and two ring patterns, highly consistent with the camera motion trajectory described in Section 1.1.

Figure .7 image matching

Figure .8 Illustration of camera poses

3.2 NeRF 3D Reconstruction Results

In each period of the image sequence, one real scene image is randomly selected and compared with the NeRF scene, point cloud model, and depth map, as shown in Figure Set 1. During the flowering period, the NeRF scene of the crabapple tree accurately reproduces the color and morphology of the flowers, consistent with the real scene, with minimal noise in the point cloud model. During the fruiting and leaf-falling periods, the NeRF scene of the crabapple tree restores the true colors of the branches and leaves, and the point cloud model exhibits distinct leaves and clear branches, visually reflecting changes in the number and color of crabapple leaves. During the dormant period, the NeRF scene and point cloud model of the crabapple tree vividly depict the tree's branch structure, providing guidance for pruning. The clear contour of the depth map depicts the distribution of flowers, the layering of branches, and accurately reflects the canopy structure of the crabapple tree, indicating NeRF's strong representation capability and adaptability to trees in various growth stages.

Figure Set 1 NeRF 3D Reconstruction Results

The comparison between the reconstruction results of NeRF and MVS is shown in Figure Set 2. At the macroscopic scale, the point cloud model reconstructed by NeRF exhibits fewer scattered noise points and almost no white noise, whereas the point cloud model reconstructed by MVS contains numerous scattered noise points, enveloping the entire tree and making it difficult to capture the tree's morphology. At the organ scale, the point cloud model reconstructed by NeRF accurately reproduces colors, with distinct shapes of flowers and leaves. In contrast, the point cloud model reconstructed by MVS struggles to discern any details, with flowers and leaves merging together. Due to the presence of a large number of white noise points, the colors in the point cloud model are severely distorted. When faced with the dormant period of the tree, characterized by relatively fewer details, MVS similarly fails to accurately capture the tree's branch structure. The branch edges are surrounded by white noise points, resulting in blurred edges and difficulty in recognition.

Figure Set 2 The comparison between the reconstruction results of NeRF and MVS

3.3 Analysis of Point Cloud Model Data

Due to the nature of the NeRF scene as a function implicitly representing pixel color and volume density, direct editing of the scene is not feasible. To assess the dimensional accuracy of the SFM-NeRF reconstruction method, it is necessary to convert the NeRF scene into a three-dimensional point cloud model and edit the point cloud model. Measurements revealed that the length of the rectangular tiles surrounding the apple tree on the ground was 150 cm. This data was used to perform scale recovery on the point cloud model to obtain the dimensional data of the apple tree.

Tree shape data (actual values) for the apple tree at various stages of its lifecycle were recorded during information collection and compared with the tree shape data (experimental values) recorded by the NeRF point cloud model to evaluate the dimensional accuracy of all point cloud models, as shown in Table 1.

Table 1 dimensional accuracy of all point cloud models

period\Tree data	height of tree (cm)	crown breadth (cm)	DBH (cm)	ground diameter (cm)
Actual values of flowering time	415.6	214.3	10.54	13.82
Experimental values of flowering time	422.2	228.2	10.42	14.20
error	1.59%	6.49%	-1.14%	2.75%
Actual values of full fruit stage	420.1	228.7	10.51	13.90
Experimental values of full fruit stage	412.0	235.9	10.66	14.10
error	-1.93%	3.15%	1.43%	1.44%
Actual values of leaf fall period	420.5	240.6	10.50	13.97
Experimental values of leaf fall period	412.9	233.5	10.27	14.34
error	-1.81%	-2.95%	-2.19%	2.65%
Actual values of dormant period	311.3	208.5	10.52	13.95
Experimental values of dormant period	321.8	211.7	10.48	14.03
error	3.37%	1.53%	0.38%	0.57%

The dimensional accuracy of the NeRF point cloud model reaches the centimeter level at the macroscopic scale and the millimeter level at the organ scale, with a dimensional accuracy generally exceeding 96%. The average errors of tree shape data during the flowering period, fruiting period, leaf-falling period, and dormant period are 2.99%, 1.99%, 2.40%, and 1.46%, respectively. The average error of dimensions for all periods is 2.21%. It is evident that the three-dimensional point cloud models reconstructed by NeRF exhibit high dimensional accuracy across various stages.

To assess the advantage of SFM-NeRF over MVS in point cloud accuracy, eight reference points were randomly placed at various branches of the apple tree. SFM-NeRF and MVS were respectively used to reconstruct the apple tree point cloud models with these reference points. The accuracy of SFM-NeRF and MVS point cloud models was evaluated based on the consistency of tree shape data and branch dimensions. The positions of reference points are illustrated in Figure 9, and the dimensional statistics at each reference point and the tree shape dimensions at this time are summarized in Table 2.

Figure.9 reference points

Table 2 dimensional statistics at each reference point and the tree shape dimensions

	Tree data (cm)			diameter of reference points on tree trunk (cm)							
	height of tree	crown breadth	ground diameter	1	2	3	4	5	6	7	8
actual values	330.2	232.5	14.1	10.11	7.41	7.13	2.10	2.89	2.80	2.76	3.07
Experimental value of Nerf	321.6	224.7	14.5	10.44	7.22	7.26	2.28	2.92	2.60	2.61	2.98
Deviation value of Nerf	8.6	7.8	0.4	0.33	0.19	0.13	0.18	0.03	0.20	0.15	0.09
Error value of Nerf	2.60%	3.35%	2.83%	3.16%	2.56%	1.82%	8.57%	1.04%	7.14%	5.43%	2.93%
Experimental value of MVS	345.6	234.1	13.5	10.90	7.35	7.62	2.48	3.50	3.37	2.95	2.40
Deviation value of MVS	15.4	1.6	0.6	0.79	0.06	0.49	0.38	0.61	0.57	0.19	0.67
Error value of MVS	4.66%	0.49%	4.26%	7.81%	0.81%	6.87%	18.10 %	21.1%	20.4%	6.88%	21.8%

The SFM-NeRF reconstructed point cloud model exhibits an average dimensional error of 3.77%, an average tree shape data error of 2.93%, and an average branch dimension error of 4.08%. In contrast, the MVS reconstructed point cloud model shows an average error of 10.29%, an average tree shape data error of 3.14%, and an average branch dimension error of 12.97%. SFM-NeRF demonstrates a reduction in average error of 63.36% compared to MVS for the apple tree point cloud model, 28.19% for tree shape data, and 60.35% for branch dimensions, indicating significant improvements in dimensional consistency.

Further analysis of the table reveals that the difference in accuracy between MVS and SFM-NeRF point cloud models is relatively small for tree shape data but significantly larger at the marked branch points. Specifically, the dimensional errors at the main trunk positions 1-3 for the SFM-NeRF point cloud model are comparable to those at the branch positions 4-8, generally below 4%. In contrast, while the MVS point cloud model exhibits relatively small dimensional errors at the main trunk positions 1-3, the errors at the branch positions 4-8 are typically around 20%, leading to an overall decrease in accuracy. These findings indicate that MVS struggles to accurately capture data on apple tree branches at the crown level, whereas SFM-NeRF consistently records complex details of branch data. In terms of dimensional consistency, the proposed SFM-NeRF 3D reconstruction method demonstrates enhanced capability for representing high-frequency details.

4.Conclusion

The study proposes a three-dimensional reconstruction approach that integrates motion recovery structure algorithm with the theory of neural radiance fields (NeRF), and designs an information acquisition device and scheme for individual fruit trees. The information acquisition platform achieves image stabilization and camera pose stabilization, with visual feedback capability, enhancing the efficiency of image acquisition. Utilizing the platform, 600 multi-view images of the fruit tree are obtained by frame extraction at equidistant time intervals from videos. The SIFT algorithm detects feature points in the image sequence, and the camera poses are recovered using MVS sparse reconstruction, achieving a 100% success rate in pose recovery, highly consistent with the actual camera motion trajectory. The low-cost information acquisition device and scheme facilitate the acquisition of multi-view image sequences and camera poses for fruit tree reconstruction, lowering the threshold for three-dimensional reconstruction.

A cherry tree grown in a continental monsoon climate region is selected as the reconstruction target, and multiple point cloud models are reconstructed for various growth stages including flowering, fruiting, leaf fall, and dormancy. The reconstruction results demonstrate that SFM-NeRF exhibits excellent high-frequency detail representation capabilities for fruit trees at different growth stages. The NeRF scenes for the cherry tree in each growth period exhibit highly consistent representation with reality, accurately capturing macro morphology, local organs, leaf color, and other phenotypic information. The point cloud models for the cherry tree in different growth periods exhibit few noise points, color restoration, clear tree contours, sharp leaf edges, and clear crown outlines in the depth map, accurately reflecting the crown structure of the cherry tree. The size accuracy of the point cloud models reaches centimeter-level precision at the macroscopic scale and millimeter-level precision at the organ scale. The size accuracy of the point cloud models across multiple periods generally exceeds 96%, with an average accuracy of 97.79% for tree structure data across all periods. The reconstructed cherry tree model with labeled points achieves an average precision of 97.07% for tree structure data, representing a 28.19% reduction in error compared to traditional MVS reconstruction methods. The size precision of trunk markers generally exceeds 96%, with an average size precision of 95.92%, representing a 60.35% reduction in size deviation compared to traditional MVS reconstruction methods and an overall error reduction of 63.36%. This approach provides a reference for addressing the challenges of capturing high-frequency phenotypic details with traditional reconstruction methods. The reconstruction method demonstrates good robustness for fruit trees across all growth stages, widely applicable in growth monitoring, fruit picking, pruning, and other industries.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their critical comments and suggestions for improving the manuscript.

FUNDING

This work was funded by the Top Talents Program for One Case One Discussion of Shandong Province, the Key Research and Development Program of Shandong Province (Major Innovative Project in Science and Technology) (2020CXGC010804), Shandong Provincial Natural Science Foundation (ZR2021MC026).

Reference

[1] ZHOU Guomin. Research and Development of Digital Orchard in China J. AGRICULTURE NETWORK INFORMATION,2012(01):10-12.

[2] Wu Wenbin，Shi Yun，Duan Yulin et al. The precise management of orchard production driven by the remote sensing big data with the SAGI J. China Agricultural Informatics,2019,31(04):1-9. DOI：http://dx.doi.org/10.12105/j.issn.1672-0423.20190401

[3] Zhou Guomin，Qiu Yun，Fan Jingchao et al. Research progress and prospect of digital orchard techniques J. China Agricultural Informatics,2018,30(01):10-16. DOI：http://dx.doi.org/10.12105/j.issn.1672-0423.20180102

[4] Mai Chunyan，Zheng Lihua，Sun Hong et al. research on 3D Ｒeconstruction of Fruit Tree and Fruit recognition and Location Method Based on ＲGB-D Camera J. Transactions of the Chinese Society for Agricultural Machinery，2015，46(S1)：35-40. DOI：http://dx.doi.org/10.6041/j.issn.1000-1298.2015.S0.006

[5]Zhao Jinze. System Research on Robot Apple Picking Path Planning Based on Deep Reinforcement Learing D. Zibo:Shandong University of Technology，2022.

[6]Chen Qing，Yin Chengkai，Guo Ziliang et al. Current status and future development of the key technologies for apple picking robots J. Transactions of the Chinese Society of Agricultural Engineering，2023,38(4)：1-15. DOI：http://www.tcsae.org/cn/article/doi/10.11975/j.issn.1002-6819.202209041

[7] KOLMANIC S，TOJNKO S，UNUK et al. The computer-aided teaching of apple tree pruning and training J. Computer Applications in Engineering Education, 2017, 25(4):568-577.DOI: 10.1002/cae.21821

[8] ZHANG C，YANG G，JIANG et al. Apple tree branch information extraction from terrestrial laser scanning and backpack-lidar J. Remote Sensing, 2020, 12(21): 3592. DOI: 10.3390/rs12213592

[9] NARVAEZ F Y，REINA G，TORRES-TORRITI M et al. A survey of ranging and imaging techniques for precision agriculture phenotyping[J]. IEEE-ASME Transactions on Mechatronics,2017, 22(6): 2428-2439.DOI: 10.1109/tmech.2017.2760866

[10]Feng Juan，Liu Gang，Wang Shenwei et al. Multi-source Images Registration for Harvesting Robot to Recognize Fruits［J］. Transactions of the Chinese Society for Agricultural Machinery，2013，44(3) : 197-203．DOI：http://dx.doi.org/10.6041/j.issn.1000-1298.2013.03.036

[11] Zhou Wei，Feng Juan，Liu Gang et al. Application of image registration technology in apple harvest robot［J］. Transactions of the Chinese Society of Agricultural Engineering，2013，29(11) : 20-26．DOI：http://www.tcsae.org/cn/article/doi/10.3969/j.issn.1002-6819.2013.11.003

[12] Chéné Y，Rousseau D，Lucidarme P et al． On the use of depth camera for 3D phenotyping of entire plants［J］． Computers and Electronics in Agriculture，2012，82: 122 -127．DOI：10.1016/j.compag.2011.12.007

[13]Adhikari B，Karkee M. 3D reconstruction of apple trees for mechanical pruning[C]//2011 Louisville, Kentucky, August 7-10, 2011. American Society of Agricultural and Biological Engineers, 2011: 1. DOI：10.13031/2013.38139

[14]XIONG Longye，WANG Zhuo，HE Yu , Application of fruit tree reconstruction and fruit recognition methods in picking scenes［J］． Transducer and Microsystem Technologies，2019，38(8)153-156．DOI：10．13873 /J．1000—9787(2019) 08—0153—04

[15]Mildenhall B，Srinivasan P P，Tancik M et al. Nerf: Representing scenes as neural radiance fields for view synthesis[J]. Communications of the ACM, 2021, 65(1): 99-106. DOI : 10.1145/3503250

[16] MARTIN-BRUALLA R，RADWAN N，SAJJADI M S et al. Nerf in the wild: Neural radiance fields for unconstrained photo collections[C]//; proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Virtual: IEEE, 2021: 7210-7219., F, 2021 [C]. DOI：https://doi.org/10.1109/CVPR46437.2021.00713

[17] MüLLER T，EVANS A，SCHIED C et al. Instant neural graphics primitives with a multiresolution hash encoding [J]. ACM Transactions on Graphics (ToG), 2022, 41(4): 1-15. DOI : 10.1145/3528223.3530127

[18] TANCIK M，WEBER E，NG E et al. Nerfstudio: A modular framework for neural radiance field development [EB/OLJ]. (2023-7-23) [2023-7-27], DOI : https://dl.acm.org/doi/fullHtml/10.1145/3588432.3591516

[19] BARRON J T，MILDENHALL B，TANCIK M et al. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields[C]//; proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision(ICCV). Virtual, : IEEE, 2021: 5855-5864.F, 2021 [C]. DOI：https://doi.org/10.1109/ICCV48922.2021.00580

[20]MILLER J，MORGENROTH J，GOMEZ C. 3D modelling of individual Tree using a handheld camera: Accuracy of height, diameter and volume estimates [J]. Urban Forestry & Urban Greening, 2015, 14(4): 932-940. DOI：10.1016/j.ufug.2015.09.001

[21]GAO Q，KAN J. Automatic Fforest DBH Mmeasurement Bbased on Sstructure from Mmotion Pphotogrammetry[J]. Remote Sensing, 2022, 14(9): 2064. DOI：10.3390/rs14092064