Three-Dimensional Reconstruction Method of Nanguo Pear Fruit Based on SfM and Improved Neural Radiance Field

Yuhang Wang, Qinghong Wu*

***Abstract*—This study presents SfM-NGP, an efficient 3D reconstruction framework for Nanguo pears that addresses the limitations of conventional methods in computational efficiency and equipment cost through the integration of improved Neural Radiance Field (NeRF) and Structure from Motion (SfM) techniques. By developing an optical flow-based adaptive frame extraction strategy for optimal keyframe selection and implementing Instant-NGP with multi-resolution hash encoding to reduce the MLP network to four layers (64 neurons/layer), the proposed method achieves significant performance improvements, as demonstrated in experiments using Liaoning Anshan Nanguo pears captured with a consumer smartphone (Xiaomi 13, 4K video). SfM-NGP completes single-fruit reconstruction in just 2.57 minutes on an RTX 3060 GPU, showing 92%, 99.3%, and 65.7% faster processing compared to SfM-MVS, SfM-NeRF, and 3DGS respectively, while delivering superior reconstruction quality with a peak PSNR of 25.96 dB, 15%, 7.8%, and 3% improvements over benchmarks and the lowest standard deviation in fruit diameter measurements. The framework enables high-fidelity morphology and texture reconstruction under natural lighting conditions, providing a cost-effective solution for agricultural applications, including fruit phenotyping, quality inspection, and digital orchard management.**

***Index Terms*—Three-Dimensional Reconstruction, Neural radiance field, Motion recovery structure, Instant-NGP, Nanguo Pear**

I. Introduction

THE Nanguo pear is primarily cultivated in the Anshan, Haicheng, Xiuyan, and Liaoyang districts of Liaoning Province, China. This variety is renowned for its delectable, fruity flavor and nutritional value, and it is a notable source of trace elements such as zinc, iron, potassium, and calcium. The following nutrients have been shown to be higher in crude protein, crude fiber, crude fat, lysine, and soluble sugar content compared to other fruits: antioxidant, blood pressure, heat, cool, moisten the lung, eliminate phlegm, clear heat, detoxification, and other effects. This has led to a high level of popularity among the general public. This substance has many benefits, including antioxidants, blood pressure, cooling, moistening the lungs, eliminating phlegm, clearing heat, and detoxification effects[1]. These properties have contributed to its popularity among the public. Presently, the primary grading methods employed for Nanguo Pear are manual and mechanical. The efficiency and precision of artificial grading are comparatively low. Mechanical grading, a technique employed to surmount the limitations of artificial grading, has been demonstrated to enhance grading efficiency and quality. However, mechanical grading is more susceptible to damage, which can affect the appearance of Nanguo Pear. The present situation with regard to the testing and grading of Nanguo pears is as follows: such activities are, for the most part, carried out in factories and large-scale collection organizations. As a result, farmers and enterprises maintaining limited stocks of Nanguo pears face significant operational challenges.

Presently, three-dimensional reconstruction technology offers significant advantages in various aspects of human production and life. In the domain of industrial prediction[2] and agriculture, the implementation of a three-dimensional model of the Nanguo pear fruit has garnered significant attention. This approach facilitates expeditious detection and enables precise measurement of the fruit's dimensions, morphology, hue, and other attributes. Consequently, it fosters scientific and technological advancements within the Nanguo pear industry. Moreover, the integration of the three-dimensional model of Nanguo Pear with information technology enhances its adaptability to the relevant production process and facilitates researchers' intuitive observation of model data. This development holds significant potential in fields such as quality control, packaging design, marketing, and scientific research.

Presently, the field of domestic research on Nanguo pear-related three-dimensional reconstruction remains in its nascent stages. The majority of extant research in this area is based on the reconstruction of common fruits or plants. HAO et al. proposed a three-dimensional reconstruction method that employs deformable convolution and Laplace pyramid residuals to process multi-view data[3]. TANG et al. employed a three-dimensional reconstruction method, utilizing the time-difference technique in conjunction with visible light imaging, to extract the plant traits of maize seedlings[4]. ZHOU et al. performed a three-dimensional reconstruction of plants based on ground-based LiDAR for characterization measurements[5]. PAN et al. developed an innovative three-dimensional reconstruction technique integrating LiDAR sensors, professional-grade cameras, and agricultural robots, which achieves high-precision crop reconstruction[6]. However, these methods have problems with high equipment costs, harsh environmental conditions, and poor reconstruction efficiency, which makes it difficult to be widely used in large-scale production.

Structure from motion-multi view stereo (SfM-MVS) is

Manuscript received April 23, 2025; revised July 11, 2025.

Yuhang Wang is a postgraduate student of School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, Liaoning, 114051, China (e-mail: 921292314@qq.com).

Qinghong Wu is a professor of School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, Liaoning, 114051, China (Corresponding author to provide e-mail:aswqh@163.com).

a prevalent solution in 3D reconstruction. The operation of Structure from Motion (SfM) is predicated on the principle of extracting feature points from images captured from multiple viewpoints. These feature points are then used to reconstruct the 3D trajectory of the camera and the 3D structure of the scene through the corresponding links between these feature points in different images[7].

The unique advantage of SfM over traditional stereo vision methods is that it can build high-precision 3D models from unstructured, randomly acquired images. Because of this feature, SfM has been popularized in many fields of practical application scenarios. Hu et al. used the SFM algorithm to obtain 3D point cloud data of wolfberry plants for 3D reconstruction[8].HE et al. carried out a 3D reconstruction of soybean plants by SfM-MVS[9].GAO et al. applied SfM-MVS to the measurement of tree trunk diameter in the field forest inventory and achieved a more satisfactory measurement effect[10].

Neural radiance fields (NeRF) is a 3D reconstruction method based on implicit neural rendering[11]. It can reconstruct a three-bit model with high accuracy by inputting 2D images. In recent years, many scholars have successively proposed various schemes to optimize NeRF.LI et al. skipped empty regions and terminated occluded regions early to reduce the number of sampling points per ray to improve the training speed[12].SMITT et al. proposed a PAg-NeRF to achieve panoramic 3D representation of crops using an RGB-D (red-green-blue-depth) depth camera[13]. SAEED et al. proposed PeanutNeRF for the 3D reconstruction of peanut plants and obtained results with high accuracy[14]. Muhammad et al. utilized NeRFacto for the 3D reconstruction of plants in the field[15]. Yang et al. combined NeRF with the RandLA-Net network to realize semantic segmentation of fruit 3D point clouds[16]. Although the improved NeRF model reaps obviously results in reconstruction quality, its reconstruction efficiency is not satisfactory. A large amount of time needs to be invested in the model training and rendering phases, which seriously hinders the large-scale application of this technique. Moreover, current research mainly focuses on the 3D reconstruction of crops in the laboratory environment. Once out of the laboratory, the expensive acquisition equipment, complex acquisition process, and long reconstruction cycle all make the feasibility of this technology in practical applications questionable. Therefore, there is an urgent need for a new method to realize the reconstruction of Nanguo Pear in a general-purpose environment.

MULLER et al. proposed an instant neural graphics primitives (Instant-NGP) method based on multi-resolution hash coding to improve the efficiency of Nerf reconstruction. Instant-NGP mainly replaces spatial location coding with multi-resolution hash coding[17], which reduces network Instant-ngp mainly replaces spatial location coding with multi-resolution hash coding, thus reducing the network training time and computational resources greatly improving the computational efficiency of neural networks without sacrificing the reconstruction quality. And it has achieved better results in practical applications[18, 19] . Therefore, in this paper, we choose SfM and Instant-NGP as the 3D reconstruction method for Nanguo Pear fruits, later referred to as SfM-NGP.

II. Conclusion

A. Material Preparation and Methods

The Nanguo Pear produced in Haicheng, Anshan, Liaoning province, was purchased in Tmall supermarket, and a total of 85 pieces of Nanguo Pear were purchased and graded as large, medium, and small fruits, and 60 pieces of them were randomly selected as experimental materials.

In the data acquisition, this study adopts the SfM method to recover the camera position, which does not need to rely on expensive equipment, but through the matching and motion estimation of the feature points in the Nanguo Pear fruit images, recover the camera position and output the internal and external parameters of the camera. The SfM technique identifies the key feature points from the images of multiple angles, establishes the correspondence between the feature points of different viewpoints, and finally calculates the optimization algorithms such as beam by identifying key feature points from multi-angle images, SfM technology establishes the correspondence between feature points of different viewing angles and finally computes the camera's shooting position attitude and the 3D coordinates of the scene through optimization algorithms such as beam leveling. The output camera position and shooting camera parameters are inputted into Instant-NGP for training to complete the 3D reconstruction of the Nanguo Pear multi-view image. The specific process is shown in Figure 1.

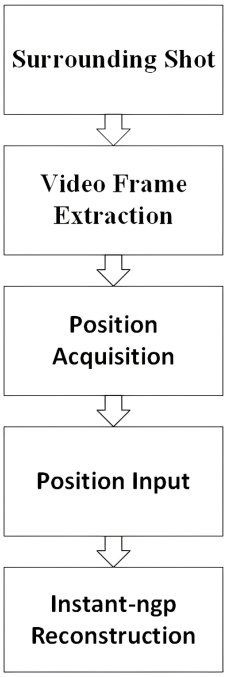


Fig. 1. 3D Reconstruction Flowchart

B. Data Acquisition

The method proposed in this study is suitable for fruit farmers or enterprises to collect data at any time. Using commercially available smartphones (Xiaomi 13 with Hyper OS 2.0), the system enables high-quality data acquisition without specialized equipment. All Nanguo pear specimens were imaged under consistent indoor natural lighting conditions following a standardized protocol. Each fruit was

placed on a 20-cm diameter black circular platform and recorded using 4K video resolution while maintaining a constant angular velocity during 60 seconds of continuous orbital capture. The smartphone lens was carefully centered on the specimen throughout the filming process to ensure data consistency. This approach successfully generated a representative dataset of 60 individual pears, providing a reliable foundation for subsequent analytical investigations in fruit morphology research.

Since it is difficult to control the device to shoot at a uniform speed all the time in the actual shooting, there may be changes in the speed of the device movement, which leads to the easy omission of keyframes in the video frame extraction, and affects the position recovery and the final reconstruction effect. Therefore, this study proposes a dynamic frame extraction method with an adaptive strategy. The frame extraction is adjusted according to speed during shooting. The optical flow method is commonly used to calculate the motion vectors of pixels between consecutive frames based on the assumptions of constant brightness and small motion. The frame-drawing frequency is increased in high-speed motion scenes and decreased in static scenes.

Firstly, Farneback dense optical flow is used to compute the motion of each pixel in the whole image to obtain the motion vector field for consecutive frames. The motion intensity is determined by counting the average value of the optical flow amplitude.Next, the number of pixels that differ between neighboring frames is calculated sequentially, and the standard deviation σ of the number of all pixels, as well as the average value θ, are found. A threshold for the number of different pixels is set as $2\sigma+\theta$, and when the difference pixels exceed the threshold, the current frame is marked as a keyframe. In the process of frame extraction, JPG is selected as the output format, and a four-digit code is used to name the file (e.g., “0001.jpg”, “0002.jpg”, etc.), which ensures the consistency of the file sequence and the convenience of searching for a single image. The file sequence is coherent, and searching for a single image is convenient. All extracted images were stored in a specific folder for subsequent experiments.

The multi-view 2D image sequence obtained by the dynamic frame extraction method based on an adaptive strategy is applied to COLMAP for incremental SfM to reconstruct the sparse point cloud of the extracted multi-view image. In this process, feature extraction and matching are carried out by the PINHOLE (pinhole) camera model, which finally realizes the accurate calibration of the internal and external parameters of the camera and the recovery of the camera position.

First, the feature points in the image are extracted and matched to determine the position and attitude of the object in the image. In addition, they are used to recover the position of the point in 3D space. The scale-invariant feature transform (SIFT)[20] algorithm detects potential pairs of scale- and rotation-invariant feature points in all images, which are collections of pixel points with unique characteristics that can be recognized from different viewpoints. For each identified feature point, its descriptor is computed. A descriptor is a set of visual information vectors containing the region around a feature point. The SIFT algorithm uses a data structure called an ”octree” to represent the descriptors. By comparing the descriptors of features extracted from different images, we can find the correspondence between them. A pair of feature points that satisfy a certain condition is recorded as a matching point. The nearest neighbor matching algorithm of the KD tree[21] (K-Dimensional Tree) is used to calculate the Euclidean distance of the matched feature points to realize fast stereo matching of the key points. Based on the matched feature point pairs, the essence matrix between the two images is calculated. This matrix reflects the relative position and pose information between the two images. Triangulate feature points based on the essential matrix to generate 3D spatial points. Triangulation involves using two matched image coordinates, image poses, and the camera’s intrinsic matrix to determine the 3D coordinates of the matched points. Then, perform bundle adjustment (BA) optimization on all generated 3D points and estimated poses. Minimize the reprojection error to remove points with excessive errors. The camera parameters of each image are estimated sequentially to generate a sparse point cloud. Finally, the camera poses and sparse point clouds of all generated images are output. The specific process is shown in Figure 2.

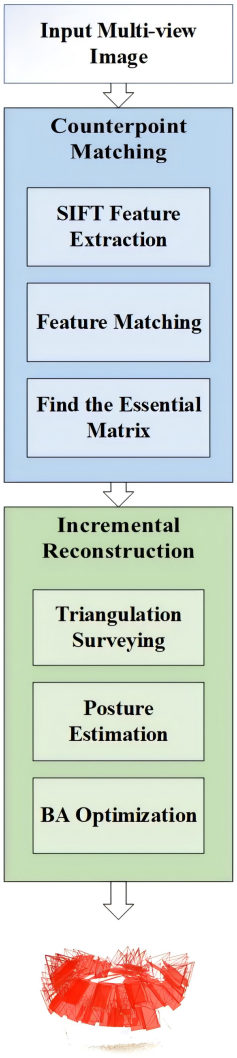


Fig. 2. SfM position acquisition flowchart

C. Nerf Introduction

NeRF (Neural Radiance Fields) is a deep learning-based 3D scene implicit representation and rendering technology. The core idea is to implicitly model continuous 5D radiation fields F_θ through a multilayer perceptron (MLP)[22]. A new perspective synthesis is achieved by differentiable body rendering, which transforms 2D images into 3D scenes. Given the 3D coordinates (x, y, z) and the observation direction (θ, ϕ) as inputs, it outputs the color c=(r, g, b) of the point and the bulk density σ. As shown in Equation (1).

$$\mathrm{F}_\theta(x, y, z, \theta, \Phi) \rightarrow (r, g, b, \sigma) \tag{1}$$

The input 3D coordinates and observation direction information are first mapped to the high-dimensional space through position encoding to enhance the high-frequency capture capability and avoid the poor performance of rendering in representing high-frequency variations in color and geometry due to the bias of the deep network learned low-frequency functions. The specific encoding is shown in Equation (2).

$$\gamma(p) = \begin{pmatrix} \sin(2^0\pi \mathrm{p}), \cos(2^0\pi \mathrm{p}), \ldots, \\ \sin(2^{\mathrm{L}-1}\pi \mathrm{p}), \cos(2^{\mathrm{L}-1}\pi \mathrm{p}) \end{pmatrix} \tag{2}$$

In the equation, p represents input information, and L denotes a high-dimensional space. In the experiment, L = 10 was selected for the three-dimensional coordinates (x, y, z), the final 60-dimensional vector γ(x)60 is obtained, and L = 4 was selected for the observation directions (θ, ϕ), the final 24-dimensional vector γ(d)24 is obtained.γ(x) represents the encoded position information, γ(d) represents the encoded direction information. The encoded information enters the MLP network, where the MLP network first uses eight fully connected layers to process the three-dimensional positions (x, y, z), outputting voxel density σ and a 256-dimensional feature vector; then, the 256-dimensional feature vector obtained above is combined with the two-dimensional orientation angle (θ, ϕ), followed by processing through one fully connected layer containing 128 neurons, outputting color c = (r, g, b). NeRF uses volume rendering, performing discrete sampling through camera rays and calculating pixel colors through integration to complete three-dimensional reconstruction. As shown in Equation(3), (4):

$$r(t) = o + td \tag{3}$$

$$\hat{\mathrm{C}}_{(\mathrm{r})} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{T_i}\alpha_\mathrm{i}\mathrm{c_i} \tag{4}$$

Included among these, α_i indication of transparency. T_i is cumulative transmittance. δ_i is the neighboring sample spacing.As shown in Equation (5),(6). To improve the rendering efficiency, a hierarchical body sampling strategy is used: the first stage uniformly samples points N_c=64 to train the coarse network, and the second stage optimizes the details based on the weight distribution probability density function to sample points N_f=128.Nerf specific process is shown in Figure 3.

$$\alpha_i = 1 - \exp(-\sigma_i\delta_i) \tag{5}$$

$$T_i = \prod_{j-1}^{i-1}(1 - \alpha_i) \tag{6}$$

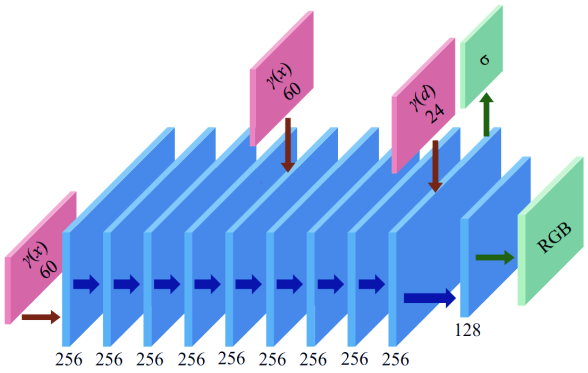
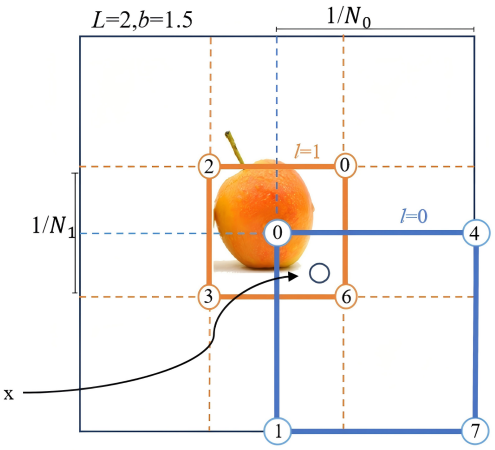


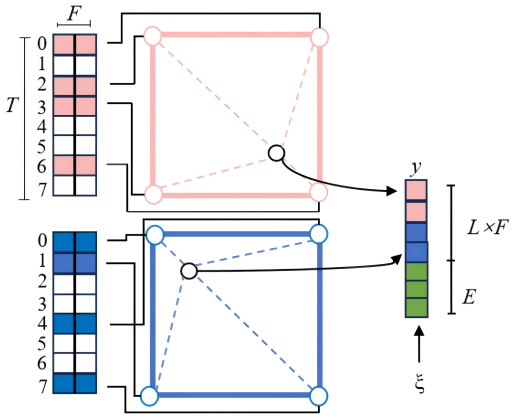
Fig. 3. NeRF Flowchart

D. Instant-ngp Optimization for Nerf

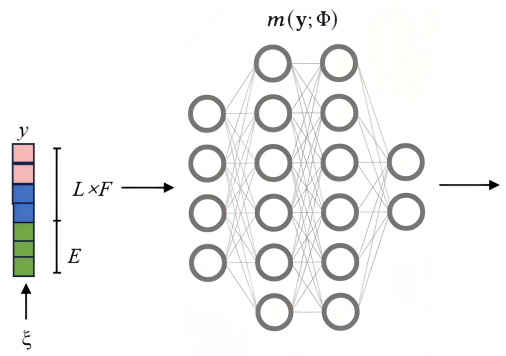
Nerf is unsatisfactory in training speed and efficiency due to its complex coding computation and multilayer MLP network structure. In order to solve this problem, instant-ngp proposes a way to achieve an accurate and efficient representation of the 3D space through multilayer Hash coding by using Multi-Resolution Hash Encoding instead of positional coding. The Multi-Resolution Hash Encoding reconstruction flowchart is shown in Figure 4.



a.Voxel vertex hash mapping



b.Hash table generation for different resolutions



c.neural network connection

Fig. 4. Instant-NGP flowchart

First, for a certain sample point x in the image, two different resolution edges are selected, their surrounding voxels are searched separately around the sample point x, and hashing is implemented on the vertices of these grids. Second, a hash table T is maintained for the different resolution grids respectively, and given that the input data are all two-dimensional images, the feature dimension F = 2. All the generated grid vertices are indexed and searched in their corresponding hash tables, respectively, to obtain the corresponding F-dimensional feature vector. The acquired feature vectors are linearly interpolated according to the relative positions of x in the different resolution grids. This is to achieve accurate matching of the input coordinates. Subsequently, the individual feature vectors, after linear interpolation of points x at each resolution, are connected to other parameters, such as camera parameters. Finally, the results are fed into a fully connected network to complete the coding training and output the color values c = (R, G, B) of the sampled points with the volume density σ.

The improvements to Instant-NGP primarily focus on the encoding strategy for the positional information of sampling points x. The positional information of sampling points x is decomposed into L layers of hash encoding, with each layer of hash encoding corresponding to image information representation at different resolution levels. Within each hash encoding layer, the x coordinate is determined by calculating the weighted average of the feature values of its surrounding encoded points, which serves as its feature value within that layer. Subsequently, the feature values obtained from all layers are integrated to form an independent L-layer structure. Each layer contains at most T F-dimensional feature vectors, where T is the capacity of the corresponding hash table. The feature vectors are stored at grid vertices. Typical values for the hyperparameters are shown in Table I.

TABLE I
HASH ENCODING PARAMETER TABLE

Parameter symbol	Parameter name	typical value
L	Storey	16
T	Hash table size	$2^{14} \sim 2^{24}$
F	Feature Dimension	2
N_{min}	Coarsest resolution	16
N_{max}	Finest resolution	$512 \sim 524288$

The size of the hash resolution of each layer is N_{min} as well as N_{max}, while the specific layer resolution of each layer needs to be calculated by the exponent of the scale factor b, as shown in Equations (7),(8).

$$N_l := \left\lfloor N_{min} \cdot b^{l-1} \right\rfloor (l = 1.2.3...L) \tag{7}$$

$$b := \exp\left\{ \frac{\ln N_{max} - \ln N_{min}}{L-1} \right\} \tag{8}$$

The choice of should match the best detail in the training data. Since the number of L levels is large, the growth factor is typically small. b denotes the ratio of the number of grid points in one direction between two adjacent resolutions and typically takes values in the range [1.26, 2] in typical parameters.

To encode the x of a three-dimensional space into a one-dimensional hash space, a hash mapping function is required, using the space hash function form shown in Equation (9).

$$h_{(x)} = \left\{ \begin{matrix} l \\ \oplus \\ i=1 \end{matrix} \quad x_i \pi_i \right\} mod(T) \tag{9}$$

In the equation, l denotes the dimension of the spatial position x of a voxel point, $\oplus$ denotes the exclusive OR operation, denotes a larger prime number, and mod denotes the modulo operation. This new encoding method enables Instant-NGP to build a lighter network, reducing the large network originally used by NeRF (8 layers with 256 neurons per layer) to a smaller network with only 4 layers and 64 neurons per layer. The number of neurons is significantly reduced compared to the original NeRF, making training more efficient and faster.

III. Results and Analysis of Nanguo Pear Reconstruction

A. Camera Position Recovery Effect

In this study, video data of 60 Nanguo Pear fruits were captured using a consumer-grade smartphone and numbered from 1 to 60. One hundred high-quality multiview Nanguo Pear fruit images were obtained after each fruit's video data was processed by a dynamic frame extraction method with an adaptive strategy. Among these, taking Fruit 1 as an example, a total of 104 images were obtained after processing. Using the SfM algorithm, the camera pose of all images was successfully restored, achieving a success rate of 100%.The original pixel resolution of the Fruit 1 video was 2160×3840. The pixels of the images obtained through frame extraction were also 2160×3840, with no surface damage. The reprojection error was 1.1928, demonstrating the high accuracy of pose recovery. The average number of feature points observed in each image is about 4708.2, which indicates that the image sequence obtained by the dynamic frame extraction method with adaptive strategy has a balanced distribution of feature points; the average trajectory length of 3D points obtained is 5.10, which indicates that more than 5.1 pictures recover each 3D point; the stability and reliability of the dynamic frame extraction method with adaptive strategy for 3D reconstruction are ensured.

Figure 5 shows the camera pose and sparse point cloud results after the sparse reconstruction of Fruit 1 using SfM. The tetrahedron represents the camera shooting direction, with the axis of the tetrahedron pointing vertically downward

from the vertices to the base. The sparse point cloud effect shown in the figure is satisfactory, indicating that the adaptive dynamic frame rate reduction method used in this study effectively addresses the challenge of data collection using a handheld smartphone in indoor environments, enabling the acquisition of complete and effective image sequences even with low-cost data collection devices in indoor settings.

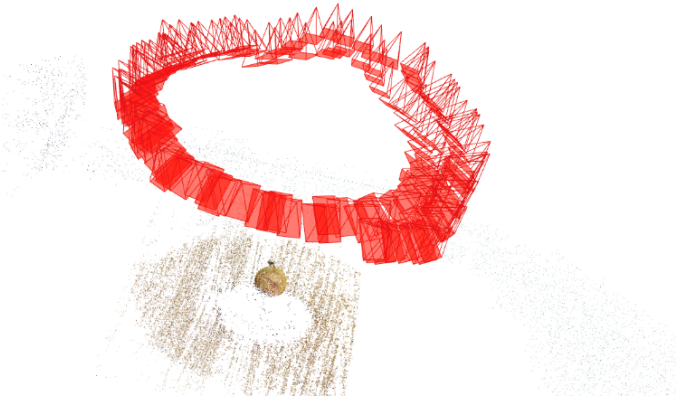


Fig. 5. Camera Position and Sparse Point Cloud

B. Nanguo Pear reconstruction results

The camera position data obtained by SfM reconstruction were used to reconstruct the Nanguo Pear fruit in three dimensions using Instant-NGP. The hardware configuration of the equipment used is AMD Ryzen 7 5800H CPU, NVIDIA GeForce RTX 3060 GPU, 16GB RAM, and 12GB video memory. The peak signal-to-noise ratio parameter of the real scene is mainly used in evaluating the modeling performance of Instant-ngp, and the formula for calculating the PSNR is shown in Equations (10),(11).

$$PSNR = 10 \times \log\left[\frac{(2^n - 1)^2}{MSE}\right] \tag{10}$$

$$MSE = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(X(i,j) - Y(i,j))^2 \tag{11}$$

In the equation, MSE is the mean square error between the original image and the processed image; H and W are the height and width of the image, respectively; n is the number of bits per pixel; PSNR is measured in dB, with higher values indicating less distortion.Figure 6 illustrates the Loss and PSNR curves with training time.

During the reconstruction process, the PSNR metric was used as the primary evaluation criterion for reconstruction quality. The PSNR value reached 23 dB within 20 seconds of training, exceeded 24 dB in the subsequent 1 minute, and peaked at 25.96 dB. This meets the high-quality compression requirements for general images, demonstrating the high-quality performance of Instant-NGP in terms of reconstruction quality.

In terms of reconstruction efficiency, the real-time performance of the Instant-NGP reconstruction process is as follows: within the first 20 seconds of training, the Instant-NGP model gradually eliminates unknown elements in the scene

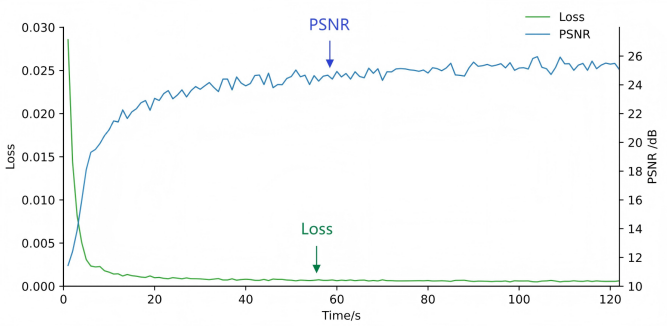
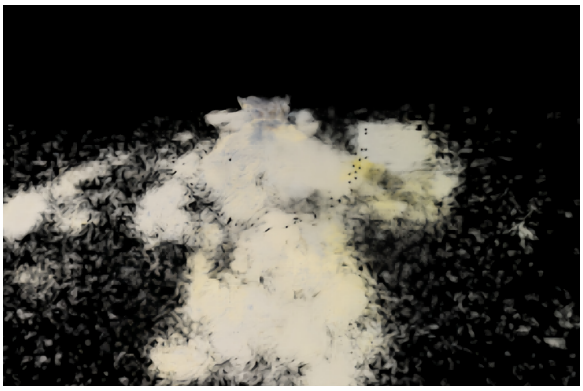


Fig. 6. Curve of Loss and PSNR with training time

by utilizing multi-view image information and camera pose data to obtain modeling results. After 2 minutes of training, the model completes the entire reconstruction, as shown in Figure 7. This demonstrates the high-quality performance of Instant-NGP in terms of reconstruction efficiency.



a.Training 1s

b.Training 20s

As shown in Figure 8 to compare it with the original image, through the fruit shape, surface texture, and fruit stalks for comparative analysis, the overall morphology of Instant-NGP reconstructed fruits has a very small gap with the real image, and maintains a complete and accurate morphology; the reconstructed surface of the fruits does not have any distortion or missing phenomena and demonstrates a high degree of accuracy.

C. Analysis of Nanguo Pear fruit reconstruction results

To verify the generality and accuracy of the SfM-NGP method proposed in this paper for the 3D reconstruction of

c.Training 2min

Fig. 7. Fruit reconstruction results

a.Original image

b.Reconstruction results

Fig. 8. Comparison of fruit reconstruction results with real images

Nanguo Pear, 60 fruits were reconstructed under the same hardware conditions using SfM-NGP, SfM-MVS, SfM-Nerf, and 3DGS methods, and the modeling model was taken as a screenshot by the pixel method of estimation of the fruit diameter, the calculation of the average error and the comparative analysis of the reconstruction accuracy, and the standard deviation was used as a mathematical tool for accuracy evaluation. The standard deviation was used as a mathematical tool to evaluate accuracy. This method can effectively quantify the performance differences in model accuracy and efficiency between different reconstruction methods, as shown in Equation(12).

$$\alpha = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{12}$$

In the equation, N is the number of measurements, x_i is the length of each measurement, μ is the fruit diameter, and α is the standard deviation.

Fruit diameter reconstruction time, peak signal-to-noise ratio PSNR, fruit diameter standard deviation, and video memory consumption were used as the basis for determination. Table II demonstrates the comparison results.

TABLE II
COMPARISON OF SFM-NGP AND OTHER METHODS RESULTS

Methods	time/min	PSNR/dB	Standard deviation	VRAM/GB
SfM-NGP	2.57	24.32	0.5	6.02
SfM-MVS	35.85	21.08	1.3	5.86
SfM-Nerf	370.72	22.56	1.1	11.74
3DGS	4.26	23.67	0.9	8.93

First, the average reconstruction time of SfM-NGP is significantly shorter than that of traditional SfM-MVS, SfM-Nerf, and 3DGS, reducing it by 92%, 99.3%, and 65.7%, respectively, demonstrating its extremely high efficiency in 3D reconstruction tasks. Additionally, in terms of reconstruction quality, the average peak signal-to-noise ratio (PSNR) of SfM-NGP is significantly higher than that of the other three 3D reconstruction methods. After training, the PSNR stabilizes above 25 dB, representing improvements of 15%, 7.8%, and 3%, respectively. The standard deviation of the fruit diameter is smaller than that of the remaining three methods. It proves that SfM-NGM can present clearer and more accurate reconstruction results in terms of image quality and 3D scene characterization. Meanwhile, although SfM-NGP consumes more computational resources than SfM-MVS, it is much smaller than SfM-Nerf and 3DGS, which indicates that it can realize high-precision reconstruction by using less computational resources.

Figure 9 shows a comparison of the modeling graphics. In terms of overall shape, SfM-NGP presents a more complete and clear image, accurately reconstructing the shape of the fruit. However, the reconstruction results of SfM-MVS and SfM-Nerf contain a large amount of noise, with the fruit shape being relatively uneven, the contours blurred, and the surface texture rough. The details of the fruit cannot be perfectly restored. Although the 3DGS method achieves a certain level of clarity, it exhibits numerous point cloud holes on the surface, and its real-world performance and clarity still fall short of the method proposed in this study.

In summary, the proposed SfM-NGP method demonstrates significant advantages over traditional SfM-MVS, SfM-Nerf, and 3DGS in terms of reconstruction time and reconstruction quality, making it suitable for the three-dimensional reconstruction of South Asian pear fruits.

IV. Comparison with other literature methods

Previously, three-dimensional reconstruction of fruits often relied on high-precision equipment or required a significant amount of time. Guan et al.[23] used three Kinect v2 sensors to construct a corn data collection system, which achieved high-precision reconstruction of corn but was too complex and expensive for widespread application. Pan et al.[6] used the traditional SfM-MVS method with cameras to reconstruct trees, which had low equipment costs, were time-consuming, and could not achieve real-time reconstruction. LI et al.[24] used RGB-D depth cameras to reconstruct greenhouse sweet pepper plants, achieving a balance between equipment cost, reconstruction quality, and time, but their overall performance still fell short of the method proposed in this paper.

a.SfM-MVS modeling results

b.SfM-Nerf modeling results

c.3DGS modeling results

d.SfM-NGP modeling results

Fig. 9. Comparison of modeling results

V. Conclusion

This study primarily addresses the rapid three-dimensional reconstruction of Nanguo pear fruits. An adaptive dynamic frame selection method is proposed and applied to the SfM-NGP method, which combines SfM and Instant-NGP, enabling rapid three-dimensional reconstruction of Nanguo pears under indoor conditions. The adaptive dynamic frame selection method avoids the loss of important frame rates caused by changes in device movement speed during actual shooting. The method performs excellently in sparse reconstruction in SfM, successfully recovering the camera poses of all images. Each image contains 4,708.2 feature points, with a reprojection error of 1.1928 and an average trajectory length of 5.10 for the 3D points. This provides a practical and convenient data acquisition method for the 3D reconstruction of Nanguo pears and other types of fruits, with broad application prospects. We also propose an SfM-NGP method, which only requires a smartphone compared to traditional methods. In the 3D reconstruction phase, compared to traditional SfM-MVS, SfM-Nerf, and 3DGS methods, the time required was reduced by 92%, 99.3%, and 65.7%, respectively. For the PSNR metric, after training, it stabilized above 25 dB, improving by 15%, 7.8%, and 3%, respectively. The standard deviation of fruit diameter is minimal, validating the high reliability of this method and laying a solid foundation for future fruit surface parameter extraction.

References

[1] T. Shi and S. Shi, "Green fruit and vegetable storage and processing technology," *Shenyang: Liaoning Science and Technology Press*, p. 315, 2015.

[2] J. Zhao, Y. Guan, and Z. Chen, "Numerical modelling and optimization of 3D surface roughness forecasting in milling," *Engineering Letters*, vol. 31, no. 4, pp. 1343–1347, 2023.

[3] Z. Hao, Z. Zhang, H. Li, B. Xu, X. Zhang, M. Xu, and W. Wang, "Multi-view 3D reconstruction based on

deformable convolution and laplace pyramid residuals," *IAENG International Journal of Computer Science*, vol. 51, no. 7, pp. 896–905, 2024.

[4] J. Li and L. Tang, "Developing a low-cost 3D plant morphological traits characterization system," *Computers and Electronics in Agriculture*, vol. 143, pp. 1–13, 2017.

[5] Z. Zhou, B. Chen, G. Zheng, B. Wu, X. Miao, D. Yang, and C. Xu, "Plant phenotypic measurement of solidago canadensis using terrestrial LiDAR scanning," *Chinese Journal of Ecology*, vol. 39, no. 1, pp. 308–314, 2020.

[6] Y. Pan, K. Hu, T. Liu, C. Chen, and H. Kang, "Phenorobot: An auto-digital modelling system for in-situ phenotyping in the field," *arXiv preprint arXiv:2402.09685*, 2024.

[7] J. Pu, F. Song, and Q. Leng, "Study on key technologies of 3D reconstruction based on sfm algorithm," *Electronic Technology*, vol. 50, no. 6, pp. 36–37, 2021.

[8] X. Hu, Z. Yang, L. Wang, P. Zhang, and Y. Meng, "Measurement of lycium barbarum phenotypic parameters based on 3D reconstruction," *Journal of Agricultural Sciences*, vol. 44, no. 02, pp. 31–38, 2023.

[9] W. He, Z. Ye, M. Li, Y. Yan, W. Lu, and G. Xing, "Extraction of soybean plant trait parameters based on sfm-mvs algorithm combined with GRNN," *Frontiers in Plant Science*, vol. 14, no. 7, p. 1181322, 2023.

[10] Q. Gao and J. Kan, "Automatic forest DBH measurement based on structure from motion photogrammetry," *Remote Sensing*, vol. 14, no. 9, p. 2064, 2022.

[11] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[12] R. Li, M. Tancik, and A. Kanazawa, "Nerfacc: A general nerf acceleration toolbox," *arXiv preprint arXiv:2210.04847*, 2022.

[13] C. Smitt, M. Halstead, P. Zimmer, T. Läbe, E. Guclu, C. Stachniss, and C. McCool, "Pag-nerf: Towards fast and efficient end-to-end panoptic 3D representations for agricultural robotics," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 907–914, 2023.

[14] F. Saeed, J. Sun, P. Ozias-Akins, Y. J. Chu, and C. C. Li, "Peanutnerf: 3D radiance field for peanuts," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 6254–6263.

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

[16] D. Yang, J. Guo, J. Yang, Y. Zhang, and Y. Lan, "3D reconstruction and semantic segmentation of fruit trees based on nerf and improved randla-net," *Journal of South China Agricultural University*, 2025.

[17] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM transactions on graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.

[18] J. Gong, B. Liu, P. Wei, H. Wang, Y. Zhang, and Y. Lan, "Three-demensional reconstruction of reconstructing fruit tree images using camera pose recovery and neural radiance fields theory," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 39, no. 22, 2023.

[19] B. Wang and X. Shen, "3D reconstruction of plant based on nerf," in *2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*. IEEE, 2024, pp. 345–348.

[20] D. G. Low, "Distinctive image features from scale-invariant keypoints," *Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[21] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *Journal of the ACM (JACM)*, vol. 45, no. 6, pp. 891–923, 1998.

[22] J. Gan, L. Chen, Q. Shen, X. Cao, Y. Yao *et al.*, "Advances in differentiable rendering based on 3D gaussian splatting," *Laser & Optoelectronics Progress*, vol. 61, no. 16, pp. 1 611 010–1 611 010, 2024.

[23] H. Guan, X. Zhang, X. Ma, Z. Zhuo, and H. Deng, "Recognition and phenotypic detection of maize stem and leaf at seedling stage based on 3D reconstruction technique," *Optics & Laser Technology*, vol. 187, p. 112787, 2025.

[24] B. Li, Q. Wu, J. Wu, M. Zhang, and H. Li, "Optimization of crop 3D point cloud reconstruction strategy based on the multiview automatic imaging system," *Transactions of the Chinese society of agricultural engineering*, vol. 39, no. 9, pp. 161–171, 2023.