Three-dimensional reconstruction of potato plants based on improved neural radiation field theory

Shunkang Ling
School of Automation Science and Engineering
Xi'an Jiaotong University
Xi'an, China
lingsk0109@163.com 0009-0000-8466-1583

Nianyi Wang
School of Software Engineering
Xi'an Jiaotong University
Xi'an, China
ny.wang@stu.xjtu.edu.cn 0000-0002-1351-4452

Laiyi Fu*
School of Automation Science and Engineering
Xi'an Jiaotong University
Xi'an, China
laiyifu@xjtu.edu.cn 0000-0001-9086-3982

***Abstract*—Intelligent and reliable access to high-quality crop phenotypic big data, which can be achieved by three-dimensional (3D) reconstruction of crop plants, will greatly promote the research in digital breeding. However, efficient achievement of satisfactory accuracy in 3D construction has remained challenging. To address this, we developed a dataset preprocessing algorithm (referred to as Laplacian) as well as an efficient and effective 3D reconstruction algorithm (referred to as NGP). As a case study, the Laplacian algorithm was tested on video datasets sampled from potato plants. Results showed that 3D reconstruction algorithms including NGP as well as classical algorithm Colmap delivered higher PSNR (1.08%-1.62% increase) and SSIM (0.34%-0.74% increase) on optimized datasets than raw datasets, implying the effectiveness of the Laplacian algorithm. Comparing NGP and Colmap on the same (raw and optimized) datasets showed that NGP delivered significantly better in PSNR and SSIM. In the meanwhile, the tests showed that in terms of reconstruction speed, NGP reconstructs 10.2 times faster than Colmap when the number of iterations was set as 20000. The strategy proposed in this study provides a novel solution for achieving efficient, low-cost and high-precision 3D reconstruction of field crops, which paves the path for intelligent and reliable phenotyping.**

Keywords—potatoes, neural radiation field, 3D reconstruction, phenotype information

I. Introduction

Crop phenotypic information refers to the physical, physiological and biochemical traits that can reflect the structural and functional characteristics of crop cells, tissues, organs, plants and populations, which are essentially the temporal three-dimensional expression of crop gene maps, geographic differentiation, and intergenerational evolution patterns [1]. Whether it is for crop functional gene analysis or variety improvement, a large amount of phenotypic data is needed to support; these data not only help to clarify the correlation between phenotypic traits and different genes, but also intuitively reflect the growth of crops for variety screening [2]. Due to the dynamic changes and complex spatial structure of crops in the process of growth and development, the traditional two-dimensional image analysis can not obtain enough phenotypic information [3]. And 3D data reconstruction technology is an effective method to digitally perceive crop phenotypes [4].

With the development of multi-source image processing technology, 3D reconstruction based on stereo vision provides an efficient, accurate and low-cost method for crop 3D information acquisition [5]. Structure from motion(SFM) can quickly acquire crop information based on 2D image sequences taken from multiple viewpoints. However, for target objects with complex structures such as plants, the sparse point cloud obtained based on SFM method contains less 3D information. By combining SFM with multi-view stereo (MVS), researchers can generate a dense 3D point cloud containing more information, which contains the advantages of both dense point cloud and colour texture information [6]. It has been successfully applied to measure surface defects, leaf morphology, population structure and other crop phenotypes [7-8]. However, this method requires the acquisition of a large number of high-overlap images, and the dense reconstruction stage consumes a long time, making it difficult to efficiently obtain massive phenotypic information, which to a certain extent restricts the application of phenotyping technology in the breeding management of facility agriculture [9].

Neural radiance fields (NeRF) is a deep learning model for 3D implicit space to learn and represent 3D scenes by multi layer perceptrons (MLPs). NeRF has become an important branch direction in 3D reconstruction research due to its advanced visual quality and view synthesis effect, and is active in the research of computer vision [10]. Currently, NeRF research hotspots mainly focus on the improvement of increasing the training and inference speed [11], large scene reconstruction [12], reconstruction quality [13], scene editable [14], and dynamic scene reconstruction [15], etc. NeRF combines the characteristics of deep learning and traditional 3D vision, and has the advantages of end-to-end synthesis route and high synthesis quality [16]. It has great potential in 3D reconstruction and phenotypic information acquisition of field crop potato.

In order to comprehensively collect large-scale crop phenotypic information, accuracy and efficiency become key indicators. Firstly, accuracy requires comprehensive and high-quality crop image information to be collected in a short period of time. In this study, video data are used to augment the data volume, which is interspersed with a large amount of interfering and low-quality data. For this reason, we use the Laplacian operator to filter the quality of the dataset, evaluate the photo quality by analysing the edge and texture features in the image, and then select high-quality images suitable for 3D reconstruction. Secondly, to meet the requirement of efficient reconstruction, we improve the Instant-NGP based NeRF algorithm by using a multi-scale hash table to store the weights, which significantly improves the training and rendering speed of the NeRF model.

2024 China Automation Congress (CAC) | 979-8-3503-6860-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/CAC63892.2024.10865480

II. Test Materials and Data Acquisition

The test objects were potatoes grown in indoor incubators, and four key varieties of plants, namely, Longshu No. 6, Longshu No. 7, Longshu No. 8 and Longshu No. 14, were selected for fine cultivation research. Two plants of each variety were transplanted into standard pots with a soil depth of 25 cm, and cultivated in a plant incubator at the School of Automation Science and Engineering, Xi'an Jiaotong University (34° 24′ N,108° 97′ E) at a constant temperature (22° C), humidity (65%), and light (8140lx for 16h), as shown in Fig. 1(a). The indoor multi-view images were sampled on 6 June 2024, when the potatoes were at the 39-day growth stage of potato set.

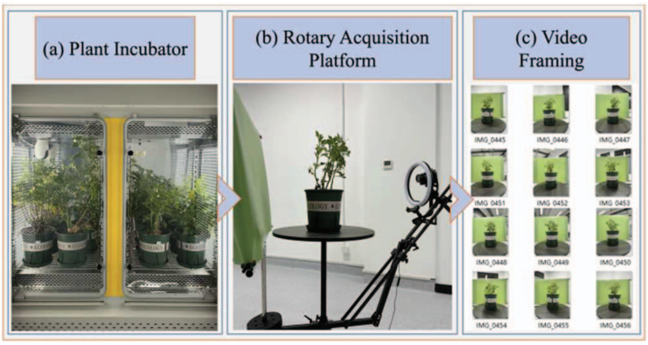

Fig. 1. Schematic diagram of data acquisition.

A self-constructed multi-view rotary acquisition platform suitable for crop phenotyping was used, as shown in Fig. 1(b). In order to avoid the misalignment of leaves caused by plant shaking, which leads to missing reconstruction, the image acquisition scheme of fixed table and rotating camera is adopted. The main body of the phenotyping platform consists of a fixed table, a multi-directional adjustable camera slide, a background plate, a fill light, etc., and the sampling position can be set accurately. The sampling equipment is a digital SLR camera (Canon M5) with a 15-45mm focal length variable lens. To ensure sufficient depth of field and image quality, the aperture is set at 6.0, ISO 100, and the exposure time is automatically matched in aperture priority mode.

III. 3D Reconstruction and Phenotypic Measurement Methods

A. *Image preprocessing algorithm based on Laplacian operator*

Laplacian operator is a widely used second-order differential operator in image processing, which is mainly used in the fields of edge detection, image segmentation, feature extraction, etc [17]. Since the Laplacian edge detection operator is a differential edge detection operator based on second-order derivatives, which is independent of the direction, only the size, so the computation is small, and its expression is (1):

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \qquad (1)$$

In this paper, edge detection is performed on surround images using Laplacian operator, which identifies the edges by calculating the second order derivatives of the pixel values in the image, i.e., where the pixel values change the most in the image, and outputs values that represent the clarity of the image. According to the degree of blurring that occurs in the overall image, the appropriate value is taken as the lower limit of blur detection, so that the number of images screened out is greater than 60% of the total number to ensure the completeness of the reconstruction. Specific implementation process: firstly, compress the image to be tested to a single-channel grey-scale map; secondly, convolve with the Laplacian operator kernel and calculate the output variance to get the blurring degree score. When the variance value is larger, the image is clearer, and vice versa. In this paper, this algorithm achieves the acquisition of high-quality datasets, and the preprocessing flowchart is shown in Fig. 2.

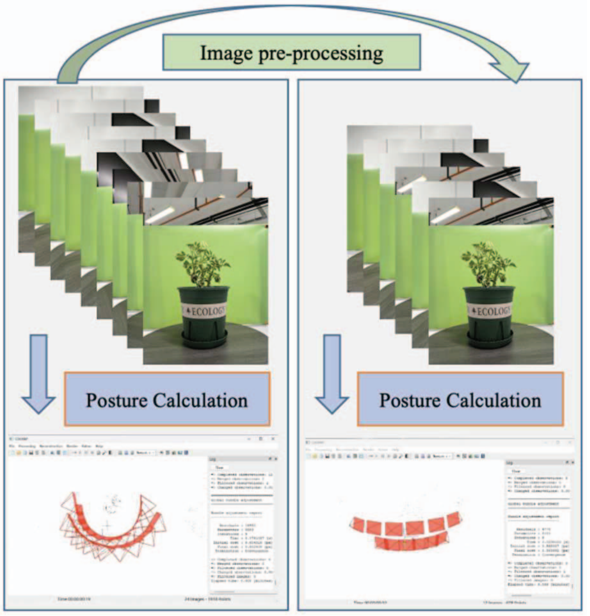

Fig. 2. Schematic diagram of dataset preprocessing.

B. *Nerf reconstruction algorithm*

NeRF is a representation of the scene as a radiation field approximated by a neural network [18], which is also modelled as an implicit scene representation, with the baseline approximated using an MLP. The working process is divided into two main parts, which are scene representation and voxel rendering.

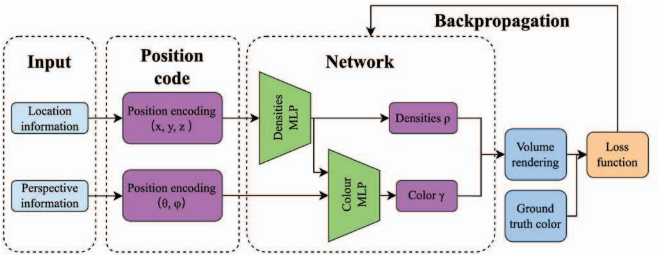

Fig. 3. Schematic diagram of NeRF structure.

The schematic is shown in Fig. 3, where the neural radiation field uses a fully connected layer deep network for scene representation, querying 5-dimensional coordinates along the emitted rays to synthesise the view and projecting the output colour and density into the image using stereo rendering techniques. The scene representation of the neural radiation field is iteratively optimised by calculating the difference between the rendered and real images and comparing them using a loss function. The colour on the rendered image can be represented in integral form as (2):

$$C(r) = \sum_{i=1}^{N} T_i [1 - \exp(-\sigma_i \delta_i)]\, c_i \qquad (2)$$

Where σ_i is the volume density at sampling point i, δ_i is the distance between sampling point i and the neighbouring sampling point, and T_i is the cumulative projection rate. Where δ_i, T_i is calculated as in (3)and(4):

$$\delta_i = t_{i+1} - t_i \qquad (3)$$

$$T_i = \exp(- \sum_{j=1}^{i-1} \sigma_i \delta_j) \qquad (4)$$

C. Instant-NGP acceleration algorithm

Instant Neural Graphics Primitives (Instant-NGP) [19] is a technique for accelerating neural graphic models as shown in Fig. 4. There is a difference from NeRF:

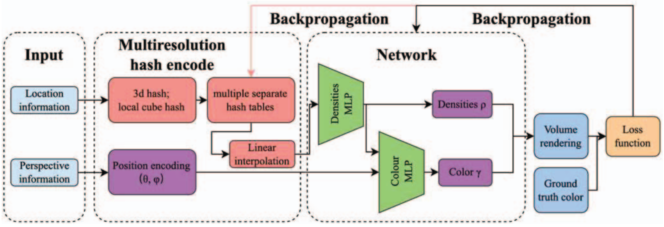

Fig. 4. Schematic diagram of Instant-NGP structure.

- The new Multiresolution hash encoding structure is used to replace the positional encoding; in the hash encoding session, the positional information is used to calculate and locate the cube of the sampling point in the volume gard; then the 8 vertices of the cube are used as inputs and mapped to the 8 indices on the hash table by the 3d hash algorithm. Using these 8 indexes, the hidden vector corresponding to each resolution level in the hash table, i.e., the density feature, can be obtained. The density feature dimension is 8*L (resolution level) *F (feature encoding dimension). In this paper, L is 16 and F is 2.
- In the linear interpolation session, the attributes of the sampled points at multiple resolutions are obtained by interpolating the attributes of the grid's 8 vertices stored on multiplle separate hash tables, and the final output dimension is L*F=32.
- Instant-NGP uses an MLP with two hidden layers, the first MLP takes the hash encoding of the sampling point location as input and serves as a decoder and adaptive output to adjust the correct feature value in case of hash conflict; the second MLP is similar to the original MLP of the NeRF, and both are used to encode viewpoint-related appearance features.

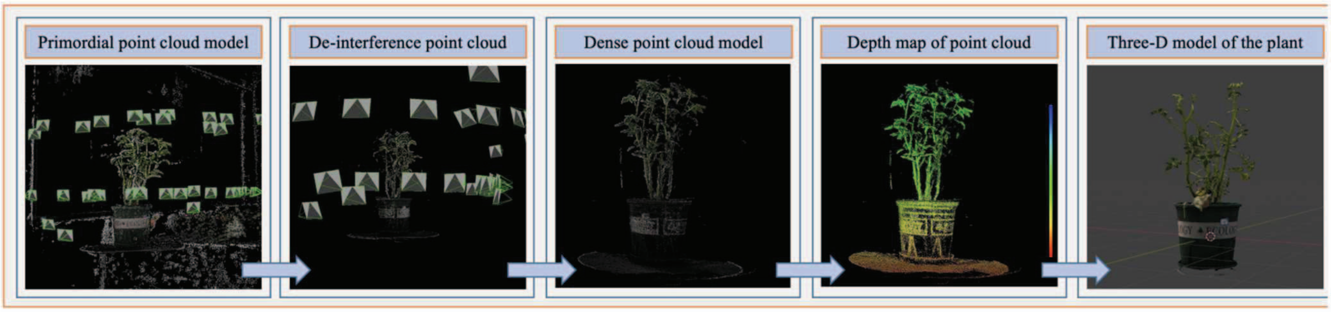
Primordial point cloud model
De-interference point cloud
Dense point cloud model
Depth map of point cloud
Three-D model of the plant

Fig. 5. Schematic diagram of reconstruction process.

The biggest difference of the NeRF-based Instant-NGP network is that a parameterised voxel grid is chosen as the scene representation. This allows the model to learn so that the parameters saved in the voxel become the shape of the scene density.The reason why NeRF is slow is because of the large scale network required for high reconstruction quality requirements. This leads to a large amount of time for each pass of the network for the sampled points. The grid interpolation method used in this paper, on the other hand, achieves high-precision scene building by increasing the density of the voxel, and although this will lead to high memory usage, the reconstruction speed will also be improved, which meets the needs of large-scale fast sampling in this paper.

IV. Test Results and Analyses

A. Experimental design

The improved NeRF algorithm used in this paper is based on Instant-NGP [19], which improves the multi-resolution hash coding and achieves the acceleration of training and rendering process. In addition, in order to validate the image preprocessing algorithm based on Laplacian operator proposed in this paper, the performance of each 3D reconstruction algorithm on different datasets will be verified separately, including reconstruction speed, reconstruction quality and effect.

The experimental environment is Windows 11, CPU is 12th Gen Intel Core i7-12700F, and GPU is NVIDIA GeForce RTX 4080. the comparison algorithm chosen is the classical 3D reconstruction algorithm Colmap reconstruction algorithm [20], which is based on SfM and MVS algorithms were developed.

B. Reconstruction process

The 3D reconstruction algorithm used in this study, the reconstruction process is shown in Fig. 5. Its specific process is:

- On the basis of the optimised dataset, the positional attitude of the camera is estimated using the structure from motion based algorithm. The method adopts incremental reconstruction, selects disordered images for feature matching and optimises the matching based on geometrical conditions, recovers the sparse structure of the point cloud by triangulation and performs relative pose estimation, then adjusts the data structure by bundle adjustment (BA), and then outputs the camera's internal references and positional poses.
- An implicit representation of the 3D scene is obtained based on the NeRF model. The machine position and

internal parameters are input into the NeRF to generate the 3D scene. Since the NeRF model is an implicit representation of the 3D scene, it is not a direct generation of an explicit representation of the 3D point cloud or mesh model. Therefore, in this study, the marching cubes (MC) algorithm is used to extract the mesh point cloud from the 3D scene obtained by the NeRF model to achieve the construction of the 3D model.

- The original 3D scene contains the potato plant model we are interested in as well as the background, which is an interference term. The sparse points are used to construct the boundary of the sampling space, i.e., the potato plant boundary, so as to eliminate the background interference and obtain the target dense point cloud model. And based on this, we can further obtain the depth point cloud map and high fidelity 3D model of the target object.

C. *Speed and quality of reconstruction*

In order to quantitatively verify the reconstruction speed of different algorithms, the session uses preprocessed optimised datasets. Traditional Colmap-based 3D reconstruction needs to calculate the depth information of the image, which often takes a long time. The Colmap-based dense point cloud reconstruction needs to complete the operations of camera position estimation, feature point extraction, image de-distortion, image photometric sideloading and depth computation, which is more time-consuming up to 63.226 min. while the Instant-NGP based NeRF reconstruction greatly shortens the 3D reconstruction time after using the multi-resolution hash position coding. Since the reconstruction effect of NeRF depends on the number of iterations, the effect graphs under different iterations are shown in Fig. 6. It can be seen from the effect diagram that the model reconstruction effect is basically qualified after the iteration number reaches 10000; when it reaches 20000 times, the model reaches a relatively perfect result. Therefore, this paper records the number of iterations until 20000 times.

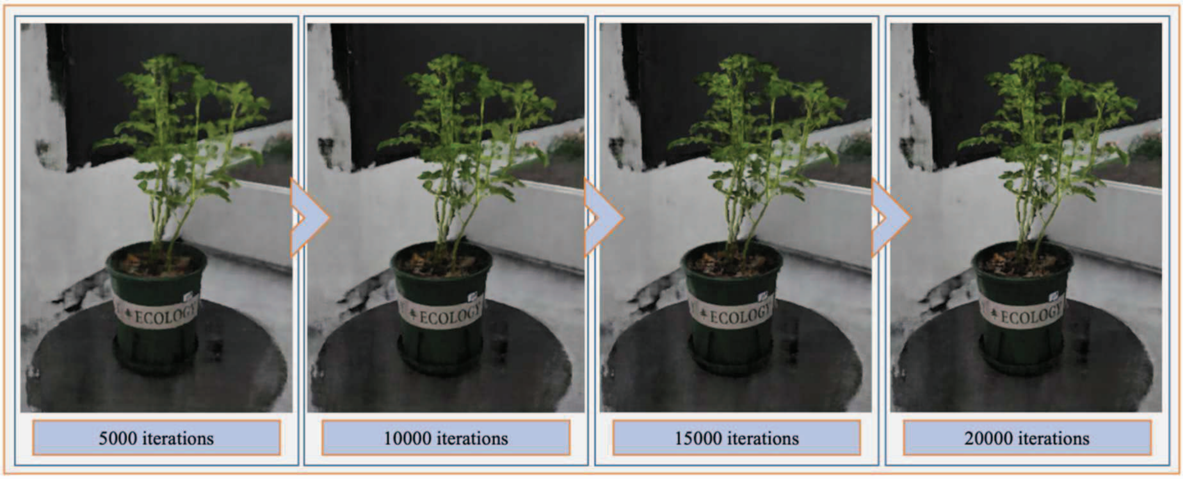

Fig. 6. Effect of different number of iterations.

Different iteration numbers and their corresponding times are shown in Table 1. From the table, it can be seen that the reconstruction speed is faster when the number of iterations is 100-5000; the reconstruction speed slows down when the number of iterations is greater than 5000. Calculations show that the Instant-NGP based NeRF algorithm used in this paper takes 9.98% of the time of the traditional Colmap algorithm in terms of reconstruction speed.

TABLE I. TIME OF TIME CONSUMPTION OF DIFFERENT 3D RECONSTRUCTION METHOD

Iterations /times	100	200	500	1000	2000
Time /mins	0.03	0.05	0.15	0.29	0.61
Iterations /times	3000	5000	10000	15000	20000
Time /mins	0.91	1.55	3.14	4.61	6.31

D. *Reconstruction effects and analyses*

In order to quantify the quality of the dataset preprocessing algorithm proposed in this paper for the NeRF and Colmap reconstruction models, we choose two commonly used evaluation metrics, Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [21]. The original dataset and the dataset filtered by the quality of Laplacian operator are used in the dataset respectively, and the number of NeRF iterations is chosen to be 20000 times, and the evaluation results are shown in the Table 2.

TABLE II. EVALUATION OF EXPERIMENTAL RECONSTRUCTION MODEL

Evaluation indicators	**PSNR**	**SSIM**
NGP-Original data	25.27	0.8257
NGP-Optimising data	25.68	0.8318
Colmap-Original data	13.87	0.6158
Colmap-Optimising data	14.02	0.6179

From the table, it can be seen that after using the optimised dataset, the PSNR and SSIM of both NeRF and Colmap algorithms are better than the original dataset: the NeRF algorithm improves the PSNR and SSIM by 1.62% and 0.74%, respectively; and the Colmap algorithm improves the PSNR and SSIM by 1.08% and 0.34%, respectively. The reconstruction effect of NeRF algorithm is also significantly better than that of Colmap algorithm.

In addition, it can be seen from the table that the Nap reconstruction results outperform Colmap in both PSNR and SSIM metrics, regardless of the dataset used.

In order to more intuitively see the improvement of the reconstruction effect by the preprocessed dataset. The reconstruction effect of Instant-NGP based NeRF algorithm on different is shown in Fig. 7. From the point cloud form, it can be intuitively seen that: at flower pot A, where some textures are similar, there is a missing reconstruction using the original dataset; at branch B, where some details are complicated and prone to shaking and dragging, there is a missing situation; and at C, where there is a background interference, there is a background interference situation.

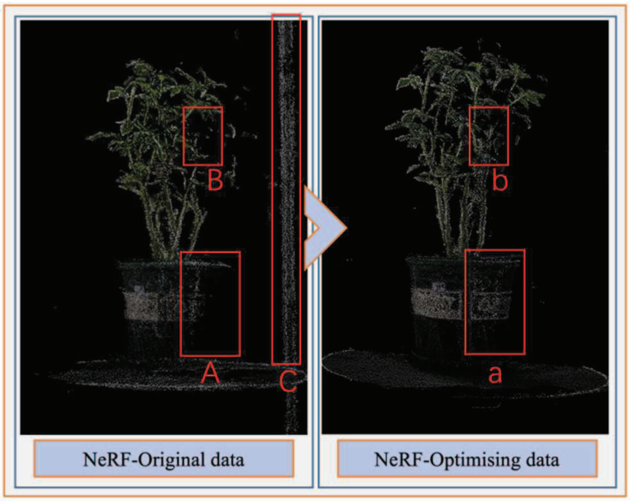

Fig. 7. Reconstruction effect of different datasets.

REFERENCES

[1] Zhao C J. Big data of plant phenomics and its research progress [J]. Journal of Agricultural Big Data, 2019, 1(2): 5−18.

[2] Fiorani F, Schurr U. Future scenarios for plant phenotyping[J]. Annual review of plant biology, 2013, 64(1): 267-291.

[3] Gibbs J A, Pound M, French A P, et al. Plant phenotyping: an active vision cell for three-dimensional plant shoot reconstruction[J]. Plant physiology, 2018, 178(2): 524-534.

[4] Zhao C, Lu S, Guo X, et al. Advances in research of digital plant: 3D digitization of plant morphological structure[J]. Scientia Agricultura Sinica, 2015, 48(17): 3415-3428.

[5] Pound M P, French A P, Murchie E H, et al. Automated recovery of three-dimensional models of plant shoots from multiple color images[J]. Plant physiology, 2014, 166(4): 1688-1698.

[6] Wu S, Wen W, Wang Y, et al. MVS-Pheno: a portable and low-cost phenotyping platform for maize shoots using multiview stereo 3D reconstruction[J]. Plant Phenomics, 2020.

[7] Xiao S, Chai H, Shao K, et al. Image-based dynamic quantification of aboveground structure of sugar beet in field[J]. Remote Sensing, 2020, 12(2): 269.

[8] Liu F, Hu P, Zheng B, et al. A field-based high-throughput method for acquiring canopy architecture using unmanned aerial vehicle images[J]. Agricultural and Forest Meteorology, 2021, 296: 108231.

[9] Li Y, Liu J, Zhang B, et al. Three-dimensional reconstruction and phenotype measurement of maize seedlings based on multi-view image sequences[J]. Frontiers in plant science, 2022, 13: 974339.

[10] 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.

[11] Fridovich-Keil S, Yu A, Tancik M, et al. Plenoxels: Radiance fields without neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5501-5510.

[12] Rematas K, Liu A, Srinivasan P P, et al. Urban radiance fields[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 12932-12942.

[13] Barron J T, Mildenhall B, Verbin D, et al. Mip-nerf 360: Unbounded anti-aliased neural radiance fields[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5470-5479.

[14] Yuan Y J, Sun Y T, Lai Y K, et al. Nerf-editing: geometry editing of neural radiance fields[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 18353-18364.

[15] Li T, Slavcheva M, Zollhoefer M, et al. Neural 3d video synthesis from multi-view video[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5521-5531.

[16] Zhu F, Guo S, Song L, et al. Deep review and analysis of recent nerfs[J]. APSIPA Transactions on Signal and Information Processing, 2023, 12(1).

[17] Bansal R, Raj G, Choudhury T. Blur image detection using Laplacian operator and Open-CV[C]//2016 International Conference System Modeling & Advancement in Research Trends (SMART). IEEE, 2016: 63-67.

[18] 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.

[19] 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.

[20] Condorelli F, Rinaudo F, Salvadore F, et al. A comparison between 3D reconstruction using nerf neural networks and mvs algorithms on cultural heritage images[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2021, 43: 565-570.

[21] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.