Research Article

Research on Outdoor Garden Scene Reconstruction Based on PMVS Algorithm

Wei He

School of Design and Art, Xijing University, Xi'an, Shaanxi 710123, China

Correspondence should be addressed to Wei He; 20070044@xijing.edu.cn

Received 29 September 2021; Revised 18 October 2021; Accepted 22 October 2021; Published 24 November 2021

Academic Editor: Le Sun

Copyright © 2021 Wei He. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The three-dimensional reconstruction of outdoor landscape is of great significance for the construction of digital city. With the rapid development of big data and Internet of things technology, when using the traditional image-based 3D reconstruction method to restore the 3D information of objects in the image, there will be a large number of redundant points in the point cloud and the density of the point cloud is insufficient. Based on the analysis of the existing three-dimensional reconstruction technology, combined with the characteristics of outdoor garden scene, this paper gives the detection and extraction methods of relevant feature points and adopts feature matching and repairing the holes generated by point cloud meshing. By adopting the candidate strategy of feature points and adding the mesh subdivision processing method, an improved PMVS algorithm is proposed and the problem of sparse point cloud in 3D reconstruction is solved. Experimental results show that the proposed method not only effectively realizes the three-dimensional reconstruction of outdoor garden scene, but also improves the execution efficiency of the algorithm on the premise of ensuring the reconstruction effect.

1. Introduction

With the rapid development of big data and Internet of things technology, geographic information system and digital city have also developed rapidly [1, 2]. The application of 3D scene modeling technology in digital city construction engineering is becoming more and more important. 3D reconstruction has always been a research hotspot in computer vision and other related fields [3–5]. It is an important means for people to obtain the three-dimensional information of objects, such as the traditional geometric modeling technology, which requires a good professional level and a large workload. It can also be obtained by 3D laser scanning technology [6]. This method is greatly affected by the outdoor environment and can not obtain the target texture information. In contrast, the 3D reconstruction method based on image feature points does not have many limitations of the above modeling methods [7, 8]. It only needs to input the image, has low cost, and does not need other special prior information. Through excellent and advanced algorithms, the 3D information of objects and scenes in the image can be recovered. Not only is the required

equipment simpler and has less restrictions on the scene, but also an accurate and realistic model can be obtained.

Obtaining the three-dimensional model of the target object from the two-dimensional image has always been a research hotspot in the field of computer vision. After years of efforts by scholars at home and abroad, remarkable results have been achieved. People summarized and proposed a relatively complete theoretical system of computer vision and began the research on multiview stereo matching [9]. Then many important research results emerged one after another. In 1992, the research team of Carnegie Mellon University proposed a reconstruction method based on optical flow method [10]. Debevec et al. of Berkeley University implemented the famous building reconstruction system Façade [11]. The system needs to obtain the approximate model of the camera and building object in advance, then reproject the model on the image to reduce the error, and finally reconstruct the three-dimensional model of the building within the specified error allowable range. Pollefeys et al. obtained sequence images by photographing the same scene from different perspectives and reconstructed the three-dimensional model of the object in the

scene based on Structure from Motion, i.e., SFM [12]. Michael Goesele et al. proposed an adaptive MVS (Multi-view Stereo) algorithm considering the influence of image light variation and noise [13]. Based on such algorithms, Furukawa et al. further proposed PMVS (Patch Multiview Stereo) based patch dense reconstruction method [14].

However, it is difficult to obtain 3D model and establish 3D scene. One of the traditional methods to obtain 3D models is through geometric modeling technology, such as solid modeling, implicit surface, etc. Nevertheless, not only does geometric modeling technology require a good technical level and a lot of work, but it is also very difficult for it to construct complex models and scenes [15, 16]. In addition, 3D laser scanning technology can efficiently obtain the point cloud of model and scene, and the accuracy is also relatively high. However, this method usually can not obtain the texture information of the target, and this method is greatly affected by the environment, especially in the outdoor open scene [17]. Compared with these methods, image-based 3D reconstruction technology needs simpler equipment and can obtain accurate and realistic models. At the same time, the development of image acquisition technology makes the data sources required for 3D reconstruction very rich, which further reduces the cost of reconstruction [18]. Therefore, in recent decades, image-based 3D reconstruction technology has been an important research topic in the field of computer vision. At the same time, with the development of virtual reality technology in recent years, this 3D reconstruction technology is becoming more and more important for the demand of large-scale scenes and models.

Although the image-based 3D reconstruction process has simple equipment and low cost, when recovering the 3D information about the object through the image, there will be a large number of redundant points in the point cloud, the density of the point cloud is not enough, the model fidelity is not high, and sometimes holes will even appear, resulting in a large difference between the final reconstruction result and the actual situation [19]. In addition, 3D reconstruction consumes a lot of time, cost, and efficiency in the process of image feature matching. In view of these shortcomings of the existing research, based on the traditional PMVS 3D reconstruction theory, this paper optimizes the image matching process, reduces the time cost of feature matching, improves the 3D reconstruction process, and solves the shortcomings of the existing methods. Finally, the optimized and improved 3D reconstruction process is integrated into the 3D reconstruction system of outdoor garden scene design.

2. Related Works

3D reconstruction is the process of capturing the shape and appearance of real objects and expressing the geometric information of 3D objects into a data model that can be stored and processed by computer, so as to carry out further research work. 3D reconstruction based on stereo vision is to recover the 3D information of the scene by processing the input image sequence and using the principle of stereo vision [20]. As shown in Figure 1, the reconstruction algorithm

FIGURE 1: Schematic diagram of reconstruction method based on stereo vision.

based on stereo vision generally obtains two images from different perspectives of the same spatial scene by the camera and processes the images.

Three-dimensional reconstruction is a difficult problem that scholars at home and abroad have been studying. In order to realize the three-dimensional reconstruction of an object, it is necessary to determine the three-dimensional contour of the reconstructed object and understand the spatial coordinates of the points on the contour, but the problem is how to obtain the missing depth information of the two-dimensional image. There are many implementation methods. According to the different methods of obtaining depth, it can be roughly divided into active mode and passive mode. The active method of a given depth map uses a rangefinder to mechanically or radioactively interfere with the reconstructed object in order to obtain a depth map, such as structured light method, moire fringe method, laser rangefinder, and other active sensing technologies [21]. However, these methods need professional instruments and are expensive, which can not be popularized among the general public. The passive method of 3D reconstruction does not have any contact with the reconstructed object and only needs to use relevant sensors to obtain its geometric information. Generally, a sensor refers to a light sensing sensor sensitive to visible light, and the input of the method is a set of digital images (one, two, or more) or video. In this case, we call it image-based reconstruction, and the output is a three-dimensional model. Compared with the active method, the passive method can be applied to a wider range of situations. In this paper, 3D reconstruction based on image sequence is studied.

3D reconstruction based on image sequence does not need a priori knowledge, but a series of images taken from different angles for the target scene or object, as well as computer vision related algorithms and mathematical calculation inference [22]. As shown in Figure 2, the process is divided into four parts: feature point extraction, sparse reconstruction, dense reconstruction, and surface reconstruction. This paper focuses on the first three parts.

Sparse reconstruction can be roughly divided into two types according to different image methods. The first is the incremental method, which first calculates the parameters of two or three images and then adds new images from the sequence for reconstruction [23]. The image initially selected by this method has a great impact on the later work, and the

FIGURE 2: Construction process of 3D reconstruction model based on multiview images.

error transmission can not ensure the accuracy. The second is a nonincremental method, which uses the method of decomposing the measurement matrix [24]. All pictures are matched in pairs to obtain the relative position parameters to determine the global position. This method is more accurate and efficient than the previous method, so it is one of the most widely used methods at present, and it is also the method used in this experiment. Dense reconstruction is to match more point pairs and obtain as many and uniform point cloud data as possible. Point cloud data is like bricks for building models. The quality and quantity of bricks are directly related to the quality of the whole building. However, the current methods have their own advantages and disadvantages and can not meet all the needs. For example, the voxel based dense reconstruction algorithm generates point cloud rules, which is convenient for subsequent surface reconstruction, but it is difficult to be suitable for large-scale scene reconstruction. The point cloud generated by patch based dense reconstruction algorithm has high precision and uniform distribution and has been widely used for its good reconstruction effect. However, this method also has its disadvantages. The cost of high precision is that it takes a long time, and there are holes in weak texture areas. Therefore, the research on how to obtain high-quality and large number of dense point clouds is of great significance and practical value.

3. Method

3.1. Overview of PMVS Algorithm. The extraction and matching of feature points is to extract stable feature points in the image, match these feature points, connect the same feature points, track the feature points, and then calculate the position relationship between images by using the relevant theory of multiview geometry. Therefore, obtaining a large number of stable feature points is an essential and important step for subsequent 3D reconstruction. In order to obtain a large number of robust feature points, this paper first describes the SIFT algorithm with invariance of scale, rotation, and illumination. The algorithm is stable and is widely used in computer vision fields such as 3D reconstruction and target tracking, as shown in Figure 3.

SIFT algorithm is robust to scaling, rotation, and translation, but affine image extraction and matching is not accurate. ASIFT algorithm can match affine images well by simulating the rotation of the optical axis of the camera and can achieve complete affine invariance.

After extracting the feature points, the feature vector is obtained. Next, the features are matched to obtain the one-to-one correspondence between pixels in the images. Feature point matching usually includes violence matching algorithm, nearest neighbor matching algorithm, and so on.

Brute force matcher (BF) is the simplest feature matching method. The violent matching method is simple to operate, but the amount of calculation is huge. If there are many images or the image resolution is large, the time and cost of violent matching are also huge. And the distance threshold is not easy to determine; too large will produce a lot of false matches, and too small is easy to miss a lot of correct matches. Therefore, it is difficult to meet the practical application.

The nearest neighbor matching algorithm [25] in the matching algorithm is similar to the k -nearest neighbor algorithm in retrieval classification. The k -nearest neighbor algorithm needs to give a query point and a positive integer k and then find the k data closest to the query point from the data set. The nearest neighbor query is its special case; that is, the k value is 1.

The nearest neighbor (NN) algorithm in feature matching was proposed by Muja and Lowe [25], and its specific algorithm process is as follows:

- (1) For the feature vector v in image 1, find the two vectors v_1 and v_2 closest to the feature vector in image 2, and $v_1 < v_2$ in the distance.
- (2) If $\text{distance}(v, v_1) / \text{distance}(v, v_2) < m$, v_1 is the match of v ; otherwise v does not match in image 2. The threshold value m is generally 0.6. The definition of distance is not unique. Euclidean distance or included angle distance can be used.

The definition of feature point matching refers to the process of establishing correlation between different image data sets. In the field of computer vision, this process is usually called stereo image pair problem. Image matching is the process of establishing the relationship between the feature points extracted from the original image and then estimating the three-dimensional position coordinates of the feature points through the projection model. In the image space, a depth map will be formed (assign a relative depth value to each image pixel). In the object space, the point cloud model of the object is usually formed. The matching in this paper is based on ASIFT features. Compared with the general SIFT features, it has better scene adaptability and richer extracted feature points, which can meet the needs of this paper.

In this paper, feature point matching is based on ASIFT feature, and the feature is represented by the descriptor vector of feature points. Generally, the spatial Euclidean distance is used to measure the distance between vectors, that is, the similarity. The similarity of descriptors can be expressed by the Euclidean distance between two feature vectors. However, the matching algorithm of this method will have some wrong matching. On the one hand, many of the feature points detected in the image can not find the

FIGURE 3: SIFT feature description.

correct matching points, because the feature points may be extracted from the background or from nontarget areas. On the other hand, due to the high-dimensional characteristics of the descriptor, the nearest point may not be the point close to the descriptor. Therefore, the effect of setting a threshold value of the global nearest matching point distance is not good, and there will be a lot of mismatches. An improved method is the way of setting a nearest neighbor and next nearest neighbor ratio as mentioned by Lowe in this paper.

Because it is easy to duplicate texture and color in the image, false matching often occurs in the process of feature matching. In the process of 3D reconstruction, stable and accurate feature points are needed. The occurrence of false matching will cause serious errors in the calculation results and have a serious impact on the reconstruction results. Therefore, removing mismatch is an important step in the process of feature matching.

Patch based multiview stereo vision (PMVS) is one of the most widely used algorithms with good reconstruction effect. The algorithm outputs dense rectangular blocks covering the visible surface in the input image. The algorithm does not need any initialization of visual shell, bounding box, and other information, and it will automatically detect obstacles and discard outliers. The key to its performance is to strictly abide by the local photometric consistency and global visibility constraints. Stereo vision is realized through a series of programs of matching, expansion, and filtering. Starting from a sparse set of matching key points, the seed patch is repeatedly extended to the nearby grid, and the visibility constraint is used to filter out the wrong matching.

PMVS algorithm shall ensure that there is at least one patch projection in each grid $c_i(x, y)$ of each picture. The implementation process of PMVS includes matching, expansion, and filtering. Starting from the sparse matching key point set, repeat the expansion, and then filter out the wrong matching through visibility constraints.

The flow of PMVS algorithm can be described according to its implementation steps. In the first step, sparse matching points are obtained after feature matching and wrong matching is eliminated. Then, the second and third steps

FIGURE 4: Schematic diagram of point cloud model reconstruction based on PMVS algorithm.

need to be iterated n times (generally $n = 3$). The flow chart is shown in Figure 4 as follows.

3.1.1. Initial Feature Matching. As the first step of the algorithm, Harris and Dog operators are used to detect the corner and speckle features in each image. In order to ensure the uniform coverage of patches, we divide the grid on each image. The cell size is $A * B \text{ Pixel}^2$ (note the difference from the previous one), and it is used as the distribution cell of corners and spots. There should be at least N local maximum points in each grid. Usually, we make $B = 32$ pixels, $N = 4$ in the experiment. After extracting these features from each image, multiview matching is performed to generate a sparse set of patches, which are then stored in the grid covering each image cell $d(i, j)$. Each image in the sequence picture is used as the reference image $E(p)$ in turn. In the remaining pictures, the pictures with an angle between the main optical

axis and $E(p)$ of less than 60 degrees form the image set $H(p)$. Then match the photos in $E(p)$ and $H(p)$.

3.1.2. Patch Extension. Since the initial matching has only a sparse set of patches, expansion is very important to generate sufficiently dense patches. In this step, the sparse patches generated in the previous step are iteratively diffused to their fields as seed patches until they can cover the visible surface in the scene. For a given patch p , the neighborhood grid set of the image block satisfying the expansion condition is recorded as $d(p)$:

$$d(p) = \{d_i(x', y') | p \in Q_i(x, y), |x - x'| + |y - y'| = 1\}. \quad (1)$$

When two patches p and p' are stored in adjacent grid cells $d_i(x, y)$ and $d_i(x', y')$ of the same image H in $S(p)$, their tangent planes are very close:

$$|(d(p) - d(p')) \cdot n(p)| + |(d(p) - d(p')) \cdot n(p')| < 2\rho_2. \quad (2)$$

That is, when p and p' meet the above conditions, they can be considered adjacent, similar to p_1 , where the value of p_2 is determined by the depth of $d(p)$ and $d(p')$.

3.1.3. Patch Filtering. Filtering is to remove some wrong patches generated in patch expansion as much as possible, so as to further enhance visibility consistency and eliminate wrong matching, so as to improve the accuracy of reconstruction. Most of the filters removed by patch filtering follow the principle of visual consistency.

The first thing to filter is the patches outside the reconstruction target, such as obstacles. Suppose that patch p_0 is an external patch, and $V(p)$ is a set of patches obscured by p_0 , which can also be said to be a discontinuous set of patches (discontinuity is that p and p' are not adjacent). When p_0 satisfies the following relationship, it is removed as an outlier:

$$|N(p_0)|\overline{M}(p_0) < \sum_{p_j \in U} \overline{M}(p_j). \quad (3)$$

As can be seen from the above formula, when p_0 is an abnormal value, the values of $M(p_0)$ and $|N(p_0)|$ will be very small, so generally p_0 will be filtered out.

3.2. Improvement of PMVS Algorithm. Although the reconstruction effect of PMVS algorithm is good, the spatial complexity and time complexity of PMVs algorithm are also very large with the increase of the number of pictures and the improvement of image resolution, especially in the steps of feature point matching and patch diffusion. Considering the complexity of outdoor garden scene, in order to better meet the application environment of dense point cloud, this paper improves the PMVS method.

First, we need to improve the candidate selection strategy. In the original PMVS algorithm, the feature points of initial feature matching are extracted by Harris

operator and Dog operator. However, the feature points extracted in this way are not comprehensive in the sparse texture region. The PMVS method proposed by Furukawa and Ponce has been proved by experiments on various data sets; the experimental objects include objects with fine surface details, outdoor scenes, and scenes with moving obstacles in different positions in multiple static images [14]. However, when the surface of the reconstructed target object is uneven, rough, and unsmooth or the photo is deformed due to elevation, PMVS algorithm may have wrong matching when selecting seed candidate points, resulting in wrong details. In the initial feature matching, each feature point on the reference image looks for candidate matching points on other images, and each pair of candidate matching points intersect to obtain model points. So far, we can see that the original method will cause feature point matching error, increase matching time, and affect the final reconstruction effect. Therefore, this paper proposes to change the projection distance to linear distance when selecting matching candidate points, which can take into account a variety of situations, the selected candidate points will be more reliable, and the reconstruction effect is better than the original in theory.

Some improvements are also needed in patch expansion. When optimizing its center and normal vector, we use the $E(p)$ and $N(p)$ information of patch p . However, when the referenced photos are distorted, the patch can not be well adapted to the actual reconstruction of the object surface due to the deviation of the normal direction. To solve this problem, this paper proposes a method to optimize and correct the normal direction of the patch by using the geometric information of the local patch, so that the patch can better reconstruct the object surface and improve the accuracy. The improved method adds a patch normal vector correction process to the traditional PMVS reconstruction method. The basic steps and algorithm flow chart of the improved algorithm are given below, as shown in Figure 5.

The improved method uses as much local geometric information as possible to optimize and correct the normal vector of the new patch. It is only for this region and will not affect other regions. In this way, the patches in the neighborhood affect and interact with each other, and the position is more accurate, which makes the reconstructed surface more smooth and is conducive to reducing the transmission of error information, so as to reconstruct more and better patches.

It can be seen from the improvement that the improved algorithm is more suitable for the application scenario of this paper and improves the execution efficiency of the algorithm. Compared with the original algorithm based on patch reconstruction, the improved algorithm uses feature extraction and matching to obtain richer matching information and eliminates the wrong matching points and points with large reprojection errors, which ensures the accuracy of the initial patch set and improves the accuracy of the reconstruction model to a great extent. Moreover, due to the elimination of the wrong matching points, the efficiency of search in the diffusion process is improved and the amount of calculation in the diffusion process is reduced.

FIGURE 5: Schematic diagram of point cloud model reconstruction based on improved PMVS algorithm.

4. Experiment and Analysis

Due to the complexity of 3D reconstruction process, the intermediate stage not only needs the results of the previous stage as input, but also provides input for the next stage. Often, a whole set of process is difficult to complete in one framework. If you use multiple platforms, you not only need to be familiar with the new platform, but also need to coordinate the input and output formats between various platforms. The system implemented in this paper is developed based on Visual Studio 2015. The class libraries needed are concentrated in the open source OpenMVS framework, including OpenCV, Ceres, GLOG, Eigen, OpenMVS, and other class libraries. The program implemented by the system provides some methods in the class library of the collection, some can be directly applied, and some need to be modified accordingly. The SIFT algorithm process is modified. Before threshold screening, bilinear pixel interpolation is added, and then the modified method is used to extract feature points. The dense reconstruction algorithm implemented in OpenMVS should also be appropriately modified according to the principles described above. In the whole process, a series of files will be generated, including feature point files, camera parameters, matching results, sparse point clouds, and dense point clouds.

Firstly, the function of the system is defined: recover and reconstruct its three-dimensional dense point cloud model from a series of images taken from different perspectives of the same object. Design function modules: data input and output, feature point extraction, camera calibration, feature point matching, sparse reconstruction, dense reconstruction, etc. The work flow chart of the system is shown in Figure 6.

To evaluate the proposed method, we used a data set downloaded from the latest technology (called TB-Roses [26]). TB-Roses dataset is composed of 400 images of rose bushes recorded in a real garden with a resolution of $640 * 480$ pixels². It is designed to test the segmentation and rendering algorithm of rose branches in horticultural robot applications. The image is provided with the ground truth value to mark the branch of segmentation. As shown in Table 1, the characteristic information of the data set used is displayed, including type, number of samples, resolution,

brightness, and type of ground truth contained. The maximum and minimum average image brightness values for the dataset are also included in the table (in the range $[0, 255]$). The relevant eigenvalues better reflect how the outdoor dataset has more variable lighting.

Therefore, we can evaluate the method in this paper through this data set. Firstly, taking TB-Roses as the experimental object, four common segmentation methods such as DeepLabv3 [27], U-Net [28], FCSN [29], and SegNet [30] are evaluated, and their different super-parameters are analyzed. Secondly, the postprocessing effect of parallax calculation and combination of segmented image and parallax image is evaluated. Next, different skeletonization methods for detecting branches are compared. Finally, the accuracy of 3D reconstruction is evaluated.

Three evaluation indexes commonly used for such tasks are selected to evaluate the performance of the proposed method, which are Precision, Recall, and Recall, respectively, defined as follows:

$$\text{Precision} = \frac{T_1}{T_1 + T_2},$$

$$\text{Recall} = \frac{T_1}{T_1 + T_3}, \quad (4)$$

$$F_1 = \frac{2T_1}{2T_1 + T_2 + T_3},$$

where T_1 (True Positives) indicates the number of branch pixels correctly segmented, T_2 (False Positives) denotes the number of background pixels which are incorrectly specified as branch pixels, and T_3 (False Negatives) denotes the number of nonsegmented branch pixels.

The IoU parameter is used to verify whether the algorithm can effectively detect all branches, as well as their size and location. Therefore, this paper calculates this parameter according to equation (5) and maps each branch of the ground truth value (g) to the segmentation suggestion (b) with maximum IoU overlap.

$$\text{IoU} = \frac{s(G_b \cap G_g)}{s(G_b \cup G_g)}, \quad (5)$$

FIGURE 6: Flow chart dense point cloud model construction.

TABLE 1: Characteristics of the datasets evaluated.

Processing object	Number of images	Pixel	Brightness (min/max)	Type
TB-Roses	400	640 * 480	80/140	Real

TABLE 2: Results obtained by the different algorithms evaluated for the processing object.

Method	Precision	Recall	F_1	IoU
DeepLabv3	45.42 ± 0.7	56.42 ± 0.8	54.84 ± 0.7	78.95 ± 5.6
U-net	63.63 ± 3.4	72.82 ± 2.5	67.74 ± 3.8	76.43 ± 5.8
FCSN	76.24 ± 0.5	81.48 ± 1.6	78.83 ± 2.1	85.26 ± 6.3
SegNet	68.37 ± 0.8	78.35 ± 3.5	75.39 ± 0.8	83.18 ± 8.2

where $s(G_b \cap G_g)$ represents the intersection between branch recommendations and basic facts, whereas $s(G_b \cup G_g)$ depicts its union.

As shown in Table 2, the comparison results of relevant indicators are obtained by different methods. It can be seen that the FCSN method obtains better results than other methods. In addition, by analyzing the Recall parameters at the pixel level, it is found that FCSN has the best overall segmentation performance. Thus, from the branch level measurement Recall, FCSN can effectively recover more branches.

In order to strictly analyze the results obtained, we compared the statistical significance by considering the nonparametric Wilcoxon signed rank test of paired samples [2]. The purpose is to evaluate whether the segmentation performance has been improved. Compared with other methods, the segmentation results obtained by FCSN are statistically significant. For this purpose, each data set and the folded results are compared in pairs. From the comparison results, if compared with SegNet, the p value obtained by this method is 0.005, while compared with U-Net and DeepLabv3 methods, the p value obtained is 0.002. Therefore, the test shows that this method is obviously better than the results obtained by other methods (considering the statistical significance threshold $p < 0.01$, the most restrictive threshold is usually used).

The segmentation results obtained by different methods on the sample image of the data set are shown in Figure 7. In order to make the graph more informative, we selected samples with obvious errors in all methods. Each column corresponds to the results obtained by each method, in

which the black and white areas represent the correct detection of branches and background, respectively, and the red and blue pixels represent the T_1 and T_2 of branches, respectively. These numbers can not only find the location of the error, but also know the accuracy of the visualization method. It is learned that due to the influence of non-branching elements (general background, leaves, etc.), using U-Net and SegNet methods will produce more noise, and using DeepLabv3 method is conducive to the segmentation of thicker objects. Although FCSN method can also make mistakes, these errors mainly occur in the contour of branches.

The postprocessing problem for combining segmented images and parallax images is evaluated below. Several depth measures in previous work were used to evaluate differential results [31], which are

$$\text{Rmse} = \sqrt{\frac{1}{T} \sum_i^T \|d_i - d_i^g\|^2}, \quad (6)$$

$$\text{Sq-Rel} = \frac{1}{T} \sum_i^T \frac{\|d_i - d_i^g\|^2}{d_i^g},$$

where T is the total pixels in the image, d_i is the estimated depth for the pixel i , and d_i^g is the ground truth depth.

The results of the experiment are shown in Table 3. The first four columns are the parallax evaluation results, and the last three columns are the segmentation results. The first two columns indicate the parallax of the complete image, while

FIGURE 7: Examples of results obtained by segmentation of sample images using different methods. (a) Object. (b) DeepLabv3. (c) U-net. (d) FCSN. (e) SegNet.

TABLE 3: The calculation results of image, branch, and slice related indexes obtained before and after the application of the algorithm.

Algorithm	Image		Branch		Slice		
	Rmse	Sq-rel	Rmse	Sq-rel	Precision	Recall	F_1
PMVS	0.5793 ± 0.4	0.0846 ± 0.3	0.2731 ± 0.4	0.0289 ± 0.2	89.72 ± 3.9	92.89 ± 2.7	90.37 ± 2.7
Improved PMVS	0.5016 ± 0.1	0.0549 ± 0.1	0.1064 ± 0.3	0.0079 ± 0.1	93.15 ± 3.6	92.16 ± 2.4	93.27 ± 2.4

TABLE 4: The average distance between 3D reconstruction and ground real point cloud.

Distance	Mean (m)	Std (m)
x	0.00248	0.00385
y	0.00373	0.00379
z	0.00539	0.00748

FIGURE 8: 3D reconstruction results obtained using improved PMVS.

the third and fourth columns represent that the branch only considers the pixels within the ground real value. From the results of branch level in this experiment, we know that the difference improvement process only affects branches, and the rest of the process only uses these depth values. As shown in the figure, these processes help to improve the original results of segmentation and parallax. The difference improved most significantly, and the branch level Rmse decreased from 0.2731 to 0.1064.

In order to evaluate the performance of 3D reconstruction, it is compared with the 3D skeleton ground authenticity of synthetic image. This ground truth has 3520 point clouds, and each point cloud contains a 3D skeleton of synthetic roses. During the evaluation, it is necessary to obtain the average value of the minimum distance between each point of the ground truth value and the reconstructed skeleton. This parameter can measure the average distance of the reconstructed skeleton relative to the ground truth value.

For objective evaluation, the depth of the reconstruction point is normalized by the farthest point in the ground truth of each plant.

The evaluation results are shown in Table 4, which lists the distances on each axis x , y , and z , where z represents the depth and y represents the axis pointing to the ground. It can be seen that the average distance is less than 1 cm, and the error on the x - y plane is less than the error on the z axis (calculated according to the depth). However, this error is still less than 1 cm. Because the opening size of the end effector is 1.5 cm and is curved at the tool tip, this accuracy indicates that the trimming process has met the requirements. An example of three-dimensional reconstruction obtained using different input images is shown in Figure 8, which can intuitively see that the results obtained by this method are basically consistent with the facts.

Figure 8 reflects the comparison between real photos and artificial reconstruction results rendered from equivalent

positions. Due to the vertex attribute transfer, the color gradient in the ground texture remains unchanged in 3D reconstruction. The experimental results show that the image data is structurally processed, and then the conversion from two-dimensional image to three-dimensional scene can be better realized by using this algorithm.

5. Conclusion

According to the requirements of 3D reconstruction of outdoor garden scene, the feature point extraction algorithm is studied, and the implementation principle and characteristics of the extraction operator are described. A method that can add stable feature points in weak texture region is proposed, which is based on the principle of SIFT algorithm. This method can effectively increase the number of feature points in the sparse texture region and provide more information for later processing. This paper focuses on the learning and research of PMVS dense reconstruction algorithm. Aiming at the problem of false matching in the feature point matching stage, a matching candidate point selection strategy is proposed. This method sorts the candidate points in descending order based on the linear distance in three-dimensional space. Experiments show that this method can effectively reduce false matching and improve the reconstruction accuracy. Although the method in this paper restores the complete structure of the scene to a great extent, there will still be holes in the reconstruction model in the case of large illumination differences. Therefore, the algorithm still has some problems to be improved and optimized. How to obtain the seamless texture of the scene in the environment of inconsistent illumination is a problem that needs to be solved in the future [32–35].

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares no competing interests.

Acknowledgments

This study was supported by the scientific research project of Xijing University “Research on the application of Chinese traditional decorative elements in modern interior design” with Grant no. xj160110.

References

- [1] M. Wang, D. Xiao, and Y. Xiang, “Low-cost and confidentiality-preserving multi-image compressed acquisition and separate reconstruction for internet of multimedia things,” *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1662–1673, 2021.
- [2] J. Demsar, “Statistical comparisons of classifiers over multiple data sets,” *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.
- [3] A. Lenjani, C. M. Yeum, S. Dyke, and I. Bilonis, “Automated building image extraction from 360° panoramas for post-disaster evaluation,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 3, pp. 241–257, 2020.
- [4] Y. F. Liu, X. Nie, J. S. Fan, and X. G. Liu, “Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 5, pp. 511–529, 2020.
- [5] A. Murtiyoso and P. Grussenmeyer, “Documentation of heritage buildings using close-range uav images: dense matching issues, comparison and case studies,” *Photogrammetric Record*, vol. 32, no. 159, pp. 206–229, 2017.
- [6] C. A. Perez-Ramirez, J. P. Amezcua-Sanchez, M. Valtierra-Rodriguez, H. Adeli, A. Dominguez-Gonzalez, and R. J. Romero-Troncoso, “Recurrent neural network model with Bayesian training and mutual information for response prediction of large buildings,” *Engineering Structures*, vol. 178, pp. 603–615, 2019.
- [7] B. Xu and C. Liu, “A 3D reconstruction method for buildings based on monocular vision,” *Computer Aided Civil Information*, pp. 1–16, 2021.
- [8] P. Tsarouchi, G. Michalos, S. Makris, and G. Chrysosolouris, “Vision system for robotic handling of randomly placed objects,” *Procedia CIRP*, vol. 9, pp. 61–66, 2013.
- [9] N. Brodu and D. Lague, “3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 68, pp. 121–134, 2012.
- [10] C. Tomasi and T. Kanade, “Shape and motion from image streams under orthography: a factorization method,” *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.
- [11] P. Debevec, J. Malik and C. Taylor, Modeling and rendering architecture from photographs,” *Acm Siggraph*, vol. 5, no. 2, pp. 11–20, 1999.
- [12] M. Pollefeys, R. Koch, M. Vergauwen, and L. V. Gool, “Automated reconstruction of 3D scenes from sequences of images,” *Isprs Journal of Photogrammetry & Remote Sensing*, vol. 55, no. 4, pp. 251–267, 2001.
- [13] M. Goesele, B. Curless, and S. M. Seitz, “Multi-view stereo revisited,” in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, pp. 2402–2409, New York, NY, USA, June 2006.
- [14] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multiview stereopsis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362–1376, 2010.
- [15] J. Hackenberg, M. Wassenberg, H. Spiecker, and D. Sun, “Non destructive method for biomass prediction combining tds derived tree volume and wood density,” *Forests*, vol. 6, no. 4, pp. 1274–1300, 2015.
- [16] F. Liu, Q. Zhao, X. Liu, and D. Zeng, “Joint face alignment and 3D face reconstruction with application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, pp. 664–678, 2020.
- [17] S. Paulus, H. Schumann, H. Kuhlmann, and J. Léon, “High-precision laser scanning system for capturing 3D plant architecture and analysing growth of cereal plants,” *Biosystems Engineering*, vol. 121, pp. 1–11, 2014.
- [18] M. Moeller, M. Benning, C. Schonlieb, and D. Cremers, “Variational depth from focus reconstruction,” *IEEE*

- Transactions on Image Processing*, vol. 24, no. 12, pp. 5369–5378, 2015.
- [19] K. Atsushi, H. Sueyasu, Y. Funayama, and T. Maekawa, “System for reconstruction of three-dimensional micro objects from multiple photographic images,” *Computer-Aided Design*, vol. 43, no. 8, pp. 1045–1055, 2011.
 - [20] N. Snaveley, S. M. Seitz, and R. Szeliski, “Photo tourism,” *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 835–846, 2006.
 - [21] S. Pertuz, D. Puig, and M. A. Garcia, “Analysis of focus measure operators for shape-from-focus,” *Pattern Recognition*, vol. 46, no. 5, pp. 1415–1432, 2013.
 - [22] R. Florian, H. Martin, and D. G. Giovanni, “Some radar imagery results using super resolution techniques,” *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 10, pp. 2611–2625, 2014.
 - [23] K. Suwa, T. Wakayama, and M. Iwamoto, “Three-dimensional target geometry and target motion estimation method using multistatic ISAR movies and its performance,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2361–2373, 2011.
 - [24] M. A. Stuff, P. Sanchez, and M. Biancalana, “Extraction of three-dimensional motion and geometric invariants from range dependent signals,” *Radar Signal Processing and Its Applications*, vol. 14, no. 1, pp. 161–181, 2003.
 - [25] M. Muja and D. G. Lowe, “Fast approximate nearest neighbors with automatic algorithm configuration,” in *Proceedings of the International Conference on Computer Vision Theory and Application (VISSAPP)*, pp. 331–340, Lisboa, Portugal, February 2009.
 - [26] N. Strisciuglio, G. Azzopardi, and N. Petkov, “Brain-inspired robust delineation operator,” in *Proceedings of the Computer Vision-ECCV 2018 Workshops*, L. Leal-Taixé and S. Roth, Eds., pp. 555–565, Munich, Germany, September 2018.
 - [27] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, *Rethinking Atrous Convolution for Semantic Image Segmentation*, CoRR, Berlin, Germany, 2017.
 - [28] O. Ronneberger, P. Fischer, and T. Brox, “U-net: convolutional networks for biomedical image segmentation,” in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, October 2015.
 - [29] H. Cuevas-Velasquez, A.-J. Gallego, and R. B. Fisher, “Segmentation and 3D reconstruction of rose plants from stereoscopic images,” *Computers and Electronics in Agriculture*, vol. 171, Article ID 105296, 2020.
 - [30] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: a deep convolutional encoder-decoder architecture for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
 - [31] D. Eigen, C. Puhrsch, and R. Fergus, “Depth map prediction from a single image using a multi-scale deep network,” in *Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2 NIPS’14*, pp. 2366–2374, MIT Press, Cambridge, MA, USA, December 2014.
 - [32] L. B. Michael, “Two-dimensional ESPRIT with tracking for radar imaging and feature extraction,” *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 2, pp. 524–532, 2004.
 - [33] P. Dias, M. Matos, and V. Santos, “3D reconstruction of real world scenes using a low-cost 3D range scanner,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 21, no. 7, pp. 486–497, 2006.
 - [34] P. Koppen, Z.-H. Feng, J. Kittler et al., “Gaussian mixture 3D morphable face model,” *Pattern Recognition*, vol. 74, pp. 617–628, 2018.
 - [35] K. T. Vierling, L. A. Vierling, W. A. Gould, S. Martinuzzi, and R. M. Clawges, “Lidar: shedding new light on habitat characterization and modeling,” *Frontiers in Ecology and the Environment*, vol. 6, no. 2, pp. 90–98, 2008.