IEEE Robotics and Automation Letters (RAL) paper presented at
2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

View Planning for Grape Harvesting Based on Active Vision Strategy Under Occlusion

Tao Yi[1], Dongbo Zhang[2], Lufeng Luo[3], and Jiangtao Luo[2]

***Abstract*— Replacing humans with robots for fruit harvesting is the trend of agricultural automation in the future. However, for grape harvesting robots, locating the picking point becomes a significant challenge in highly occluded environments due to the small fruit stem, which can be entirely obscured by fruit leaves when the observation angle is poor. In the work, a view planner based on an active vision strategy is proposed to address the occlusion problem. It aims to find the picking point by altering the observation perspective of the harvesting robot. The view planning process is achieved through multiple iterations. Each iteration consists of three key steps: randomly generating candidate views, predicting the ideal perspective using a score function, and guiding the robotic arm to change the viewpoint. To evaluate the degree of occlusion, a novel concept of Spatial Coverage Rate Metric (SC) is introduced. Based on this, the score function is improved by incorporating SC and motion cost. Finally, to validate the effectiveness of the planner, we conducted comparative experiments with other advanced view planners on a real grape harvesting robot. The experimental results demonstrate that our method achieves a higher picking success rate with lower computation time.**

I. INTRODUCTION

Robots have been widely employed in the field of fruit harvesting in agriculture due to the promising prospects of robots in improving productivity. A multi-purpose grape-harvester robot is introduced in [1], which can achieve harvest, green harvest, defoliation, monitoring, etc. To improve harvest efficiency, a rapid grape-harvesting robot with dual-arm is designed in [2]. However, the above-mentioned works do not mention how grape harvesting is achieved when grape stems are occluded. The issue of occlusion remains a pressing challenge in autonomous harvesting. This is particularly true for clustered fruits like grapes, where it is necessary to locate the stems of fruit clusters and perform harvesting operations through cutting. Fruit stems are typically small, and in cases of poor view, they can easily be completely obscured by fruit leaves, often resulting in failures in the harvesting process.

When manually harvesting grapes, if occlusion occurs, a common response is to change the perspective. After the adjustment, if the stem can be observed, subsequent harvesting operations proceed. Obtaining visual information favorable for harvesting operations by altering the viewing angle falls within the realm of active vision research. Inspired by this, we propose a view planning approach based on active vision strategies to address occlusion issues in robotic harvesting grapes.

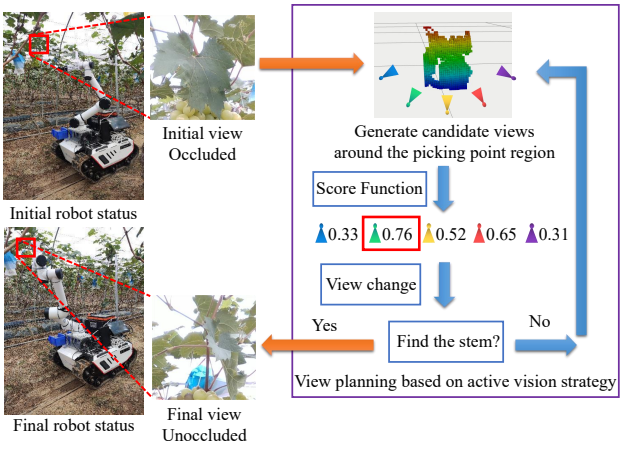


Fig. 1. **Overview of view planning for grape harvesting under occlusion.** The purpose of view planning is to find the ideal view that can detect the fruit stem. Each iteration consists of three main steps: generating candidate views around the picking point region, predicting the ideal perspective using a score function, and guiding the robotic arm movement based on the predicted view.

Before view planning, to enhance the perception of the 3D environment by the harvesting robot, the potential picking point region was determined based on the spatial relationships between fruit clusters and fruit stems, as well as the edges of cutting operations. Subsequently, an octomap [3] was utilized to reconstruct a voxel map of this region.

View planning methods can be categorized into synthetic methods and generate-and-test methods. The ideal view is directly computed under specific system and task constraints in synthetic methods. While these methods have low computational costs, their stability is not high. Generate-and-test methods find the next view through random sampling of candidate views and view evaluation. Our approach is based on the generate-and-test method. As shown in Fig.1, it involves multiple iterations, each comprising three key steps: randomly generating candidate views around the picking point region, predicting the ideal view using a score fuction, and guiding the robotic arm to change the view.

Similar view planning methods have been used in 3D scene and object reconstruction [4], 3D object recognition

*This work was supported by Key Project of Guangdong Fundamental and Application Fundamental Research Joint Fund [2020B1515120050] and the Joint Fund for Regional Innovation and Development of NSFC under [Grant U19A2083].

[1]Tao Yi is with the School of Mathematical and Computational Sciences, Xiangtan University, Xiangtan 411100, China fly199223@163.com

[2]Dongbo Zhang, and Jiangtao Luo are with the School of Automation and Electronic Information, Xiangtan University, Xiangtan 411100, China zhadonbo@163.com ,202121623013@smail.xtu.edu.cn

[3]Lufeng Luo is with the School of Mechatronic Engineering and Automation, Foshan University, Foshan 528000, China luolufeng617@163.com

[5], and pose estimation [6]. In recent years, viewpoint planning methods have also been used in agriculture to address tasks like fruit detection and pose estimation [7], fruit monitoring [8], and 3D crop environment exploration [9]. Unlike approaches that evaluate candidate views by computing information gain, the proposed method predicts the ideal view based on occlusion and motion cost. Consequently, improvements have been made to the score function.

The contributions of this work are summarized as follows:

- To address the occlusion challenges encountered in robot harvesting grapes, a view planning solution based on active vision strategies has been introduced for the first time.
- In order to reflect the degree of obstruction of the fruit pedicel, a novel concept of Spatial Coverage Rate Metric (SC) was introduced. Based on this concept, a view planning algorithm was proposed using the SC and motion loss score function.
- In the constructed laboratory and real-world outdoor experiments, the method presented in this paper was effectively validated. Compared to other advanced view planning methods, the proposed approach achieved a higher harvesting success rate in a relatively short amount of time, thereby fully confirming the feasibility of the method.

II. Related Work

The occlusion problem is a long-standing issue in harvesting robotics and has received much attention. For local occlusion, fruit detection can be directly based on the locally visible fruit region or prior knowledge of the fruit's shape. McCool et al. [10] introduced a Laplacian of Gaussian (LoG) multiscale spot detector to identify locally occluded segments of sweet peppers. Kierdorf J et al. [11] proposed using generative adversarial networks to address the challenge of leaf occlusion in berry picking for more accurate berry quantity estimation. They utilized patterns learned from images with unobstructed berries to generate highly probable scenarios behind leaves. To minimize harvesting damage, harvesting individual fruits sometimes involves detecting the peduncles and cutting them. Inkyu Sa et al. [12] use 3D Point Feature Histograms (PFH) and color information for robust and accurate peduncle detection of sweet pepper. Chris Lehnert et al. [13] introduced a novel sweet pepper peduncle segmentation system using an efficient deep convolutional neural network in conjunction with 3D post-filtering. Y. Pan et al. [14] built a panoptic volumetric map and estimated the complete shapes and 7 DoF poses of each fruit, even under substantial occlusions, using a deep neural network. While the methods above perform well in dealing with local occlusion scenarios, bunch-type grapes still suffer from locating and shearing the fruit stem, which is small and easily obscured by leaves. As for the wholly covered grape stem, it is hard to identify the picking point from prior knowledge and information on the current view. Therefore, our approach utilizes an active vision strategy to change the observation view for more helpful information.

Active vision has been introduced into the field of view planning. As early as the 1990s, Bajcsy et al [15] proposed the concept of active vision, which involves a robot controlling its camera's movement based on analysis results and task requirements to capture images from appropriate view for the task. In manual grape harvesting, actively changing the view is an effective method to cope with the occlusion problem. The same idea is applied in to reconstruct 3D scene and object, leading to the Next Best View (NBV) problem in active vision. Various viewpoint planning methods have been proposed.

Exploration-based viewpoint planning methods find widespread applications in autonomous exploration of unknown 3D environments [16], path planning [17], and environmental mapping [18]. Bircher et al. [19] introduced the Receding Horizon Next-Best-View Planner (RH-NBVP), where the views were generated by expanding RRTs and executing the first node of the best branch in a receding horizon fashion. M. Selin et al. [4] proposed the Autonomous Exploration Planner (AEP), which uses RH-NBVP as a local strategy and boundary-based planning as a global strategy to escape from local minima.

In agriculture, Zaenker et al. [20] suggested constructing a voxel map of fruit regions. On this basis, they sampled candidate viewpoints around the fruit region and used a utility function that considers the expected information gained to evaluate candidate viewpoints. In 2023, T. Zaenker et al. [8] introduced a novel graph-based view motion planning approach (VMP) for fruit monitoring. X. Zeng et al. [9] proposed a novel deep reinforcement learning (DRL) approach to determine the next best view for automatically exploring 3D crop environments. The network takes as input the encoded 3D observation map and the temporal sequence of camera view pose changes and outputs the most promising camera movement direction. T. Zaenker et al. [21] also presented a viewpoint planning method for fruit coverage, which combines local and global viewpoint planning to enable local occlusion avoidance while still covering large areas. Menon et al. [7] present a novel viewpoint planning approach that uses the missing surfaces of the fruit shapes to find the next best view.

We also adopted the approach of estimating the ideal observation viewpoint for detecting grape stems by constructing a 3D voxel map of the picking points region. Our method shares similarities with the AEP and RVP methods, but in scoring function design, we integrate SC and motion loss factors within the picking points region. Additionally, unlike the RVP planner that sets the Region of Interest (ROI) in the fruit region, we set the ROI in the picking points region. And we accelerate the mapping speed through a local mapping approach.

To the best of our knowledge, there seems to be a general lack of approaches that use active vision techniques to harvest fruits under occlusion, especially for harvesting clusters of fruit where the stem is completely occluded. The research presented in this paper is a valuable contribution to this field and offers a practical solution and approach for

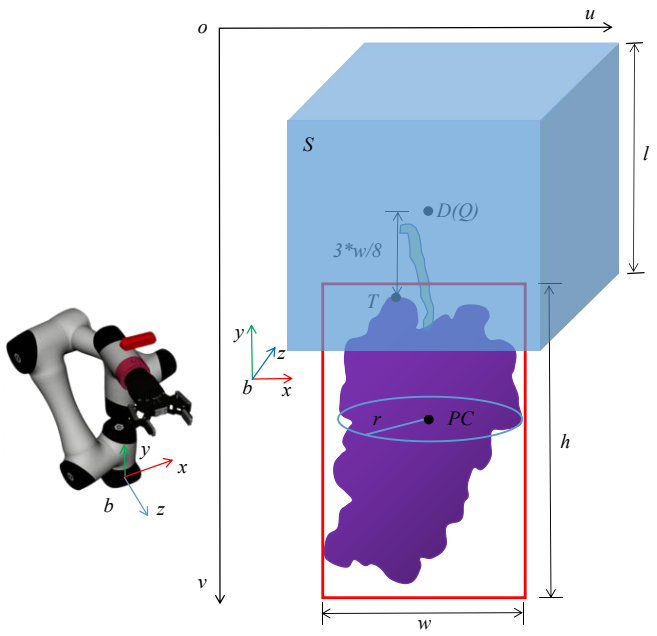


Fig. 2. **Overall schematic diagram of the picking point region.**The red box is the grape detection box of the Mask R-CNN output, and the blue cube S indicates the spatial region of the picking point.

similar problems.

III. Method Proposed

A. System Overview

In the context of this work, the goal of view planning is to find the ideal perspective that can detect the fruit stem within an occluded environment using a harvesting grape robot equipped with a depth camera. To address this, an NBV planner based on the Spatial Coverage Rate Metric (SC-NBVP) is proposed, which guides the robotic arm to adjust the camera view. The Spatial Coverage Rate Metric (SC) is a novel concept specifically designed to reflect occlusion in the picking point region.

Our proposed approach comprises three crucial stages: picking point region determination, rapid updating of the voxel map in picking point region, and view planning decision. The SC Rapidly Exploring Random Tree (SC-RRT) algorithm is the key to guiding the robotic arm movement.

B. Picking Point Region Determination

For grape harvesting robots, only the local region where leaves obstruct the grape stems will affect the harvesting operation. Therefore, we must first preprocess the images to identify potential picking point regions.

Following the approach in reference [22], grape clusters and stems are detected and segmented using the Mask R-CNN network. It is assumed that the center of the picking point region is located on the line connecting the centroid of the grape cluster to the ground, where the centroid of the grape cluster is typically the geometric center of the bunch in the 2D image. After obtaining segmentation results from the Mask R-CNN network, we calculate the grape cluster's centroid point $PC(u_c, v_c)$ based on the definition of image centroid moments.

The camera's internal and external parameters are called K_1 and K_2, respectively. The vertice of the grape contour is $T(u_t, v_t)$, and the width of the detection frame output by the Mask R-CNN network is w. Considering the operational clearance of the pruning mechanism, the center of the picking point region is defined as $D(u_c, v_t - 3w/8)$ in the pixel coordinate system Σuov, as illustrated in Figure 2. The 3D voxel map is constructed in base coordinates b, which requires a coordinate transformation of the D in Σuov. The depth Z_d of D is crucial for performing the coordinate transformation. Unfortunately, direct acquisition using the depth camera is not feasible due to occlusion caused by fruit leaves. To estimate Z_d, we can use the equatorial radius r of the bunch and the measured depth Z_c at the center of mass of the bunch.

$$Z_d = Z_c + r \tag{1}$$

The coordinate $Q(X_b, Y_b, Z_b)$ of the center point of the picking region in b is given by

$$\begin{bmatrix} X_b \\ Y_b \\ Z_b \\ 1 \end{bmatrix} = Z_d K_2^{-1} K_1^{-1} \begin{bmatrix} u_c \\ v_t \\ 1 \end{bmatrix} \tag{2}$$

In the b, a cubic region S with a side length of 0.1m is delineated around point Q as a potential picking point region.

C. Rapid Updating of the Voxel Map in Picking Point Region

We use Octomap [3] to model the 3D picking point region. The construction process of Octomap can be simplified as follows:

- Acquire an depth image from the current perspective.
- Convert the depth image into point cloud.
- Determine the probability of each voxel corresponding to each point cloud being occupied based on the Ray-casting algorithm.
- If the probability of a voxel being occupied is higher than a certain threshold, the voxel is considered occupied, and the octree's status is updated accordingly.

The voxel map update time is a critical metric during real-time robotic harvesting, directly affected by the number of point clouds. Therefore, the process can be accelerated by computing the point cloud only within the picking point region. The eight vertices of the picking point region are mapped to the depth image and obtain the maximum and minimum values of the u-axis and v-axis u_{max}, u_{min}, v_{max}, v_{min}. A rectangular region is selected on the depth image, which is centered on $((u_{max} - u_{min})/2 + u_{min}, (v_{max} - v_{min})/2 + v_{min})$, with a length of $((u_{max} - u_{min}) + \Delta)$ and a width of $((v_{max} - v_{min}) + \Delta)$, where $\Delta = 10$. When updating the voxel map, only data within the region are processed.

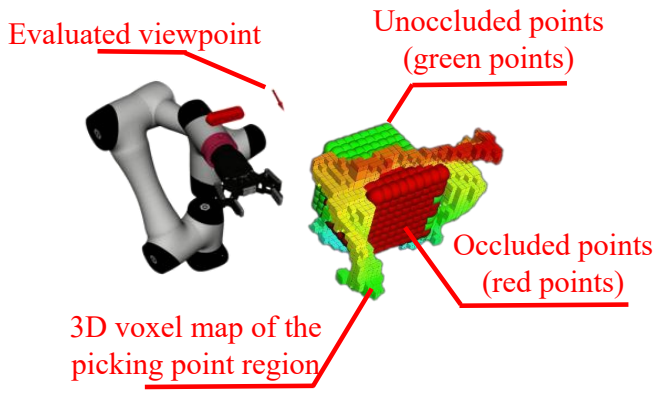


Fig. 3. **Visual interpretation of the Spatial Coverage Rate Metric (SC).** Discrete points in the picking point region are virtual points with fixed locations and totals. When the light is emitted from the evaluated view to all discrete points, the resulting points are classified as either occluded (red points) or unoccluded (green points) based on the current voxel map. The SC is the percentage of occluded points to total discrete points.

D. Spatial Coverage Rate Metric (SC)

To evaluate the occlusion of different candidate viewpoints, the Spatial Coverage Rate Metric (SC) was proposed. The basic idea is to use a virtual discrete space approximation to represent the picking point region, as shown in Fig.3. The spatial coverage rate is calculated as follows

$$SC(M, V.p) = OccludePoint(M, V.p) \\ /TotalDiscretePoint * 100\% \quad (3)$$

Here, $OccludePoint(M, V.p)$ represents the number of discrete points occluded by the voxel map M when emitting rays from viewpoint $V.p$ to all discrete points, and $TotalDiscretePoint$ represents all the discrete points within the discrete space.

The center Q of the picking point region is selected as the center point of the discrete space, and the distance between two neighboring points in the discrete space is called σ. One of the vertex coordinates of the picking point region is represented as $P(X_P, Y_P, Z_P)$. The total number of discrete points, $TotalDiscretePoint$, is given as follows:

$$TotalDiscretePoint = (2 * [|X_b - X_P|/\sigma] + 1) \\ * (2 * [|Y_b - Y_P|/\sigma] + 1) * (2 * [|Z_b - Z_P|/\sigma] + 1) \quad (4)$$

From the definition of the SC, it is known that any viewpoint around the picking point region can obtain a value reflecting the occlusion of that location.

E. View Planning SC-RRT Algorithm

The output of the SC-NBVP planner is the observation view $v = [p^{\mathrm{T}}, \alpha, \beta, \gamma]^{\mathrm{T}} \in R^6$ for the depth camera. It consists of the depth camera position $p = [x, y, z]^{\mathrm{T}} \in R^3$, the angles of roll α, pitch β, and yaw γ rotations around the x, y, and z axes, respectively. Under the definition of SC, we propose the SC Rapidly Exploring Random Tree (SC-RRT) algorithm. Algorithm 1 describes the workflow of the SC-RRT algorithm. The view planning algorithm involves multiple iterations. In each iteration, the current view's ability to detect grape stems is first determined using the Mask R-CNN network. If detection fails, the picking region's 3D voxel map is constructed using Octomap, which involves picking point region determination. Subsequently, n candidate viewpoints are randomly generated around the picking point region. To reduce the dimensionality of the search space, random sampling is applied only to the camera's position while the Euler angles are computed. The SC is calculated for each candidate viewpoint, and a score function assigns scores. The candidate viewpoint with the highest score is selected as the new predicted ideal viewpoint $V_{pred}.p$. Using $V_{pred}.p$, the tree expansion determines the closest node $V_{near}.p$ of the RRT. Then, we compute the new viewpoint $V_{new}.p$ on the straight line connection between $V_{pred}.p$ and $V_{near}.p$ at the step size l from $V_{pred}.p$. If the distance d between $V_{pred}.p$ and $V_{near}.p$ is less than l, then $V_{pred}.p$ is considered $V_{new}.p$. Finally, to ensure that the camera consistently faces the picking region, the orientation of V_{new} is computed based on the direction vector derived from the center point Q of the picking point region and the $V_{new}.p$.

The list of random candidate viewpoints for the SC-RRT algorithm is $Viewpoints = [V_1.p, V_2.p, ..., V_{n-1}.p, V_{last}.p]$. Under the spherical coordinate system with the Q as the sphere center, the global random sampling point $P_{rand} = (r, \theta, \phi)$ is defined, where $r \in [R_{min}, R_{max}]$, R_{min} is the minimum depth that the depth camera can measure, and R_{max} represents the maximum radius reachable by the robotic arm. $\theta \in [60^o, 150^o]$, $\phi \in [0^o, 180^o]$. When V_{last} is not empty, the last predicted viewpoint can help predict the current view. In the list of random candidate viewpoints, $n/2$ global random sampling points are included. The remaining candidate viewpoints are randomly sampled within the spherical region where $V_{last}.p$ is the sphere's center, and R is the radius.

A score function has been defined to evaluate random viewpoints. The function follows the principle of minimizing SC while considering the cost of motion.

$$score(M, V.p) = [SC(M, V_{init}.p) - SC(M, V.p)] \\ \exp\left[-\lambda \cdot L(V.p, V_{current}.p)\right] \quad (5)$$

$SC(M, V.p)$ and $SC(M, V_{init}.p)$ represent the SC of the evaluated and the initial viewpoint under the current voxel map. $L(V.p, V_{current}.p)$ indicates the distance between a random candidate viewpoint and the current viewpoint, measured using Euclidean distance. The constant λ determines the relative importance of the robot's movement cost and the expected reduction in spatial occlusion, with smaller values of λ prioritizing occlusion reduction and larger values favoring the shortest path. This constant is typically determined experimentally and is always positive. In each iteration, our goal is to find the viewpoint with the highest score as the predicted viewpoint $V_{pred}.p$:

$$V_{pred}.p = argmax_{p \in Viewpoints}(score(M, p)) \quad (6)$$

Algorithm 1: SC-RRT

input : initial view V_{init}, max nodes of RRT N_{max}, the center of the picking point region Q, number of candidate views n

output: Views //all planning view

$Views.init(V_{init})$;
$V_{current} = V_{init}$;
$V_{last} = V_{current}$;
$V_{pred} = \phi$;// Predicted view
$M_0 = \phi$;// 3D voxel map
$t \leftarrow 0$;
$finished \leftarrow false$;
while $t \leq N_{max}$ *or* $\neg finished$ **do**
  $RGB = Capture(V_{current}) | V_{current} \in Views$;
  $RGB_t \leftarrow RGB$;
  $p = Mask\_RCNN(RGB_t)$;
  if $p > 0.9$ **then**
    $Success()$;
    $finished \leftarrow true$;
  end
  $D = Capture(V_{current}) | V_{current} \in Views$;
  $D_t \leftarrow D$;
  $M_{t+1} = Octomap(M_t, RGB_t, D_t)$;
  $Viewpoints = getCandidates(n, V_{last}.p)$;
  $V_{pred}.p = argmax_{p \in Viewpoints}(score(M_{t+1}, p))$;
  $V_{near}.p = Near(V_{pred}.p, Views.p)$;
  $V_{new}.p = Steer(V_{pred}.p, V_{near}.p, StepSize)$;
  $V_{new} = caculatePose(V_{new}.p, Q)$;
  $V_{last} = V_{new}$;
  $Views.add(V_{new})$;
  $RobotArm.move(V_{new})$;
end

IV. EVALUATION

To assess the SC-NBVP planner's performance, we conducted laboratory and real-world outdoor experiments separately. In the context of exploration-based viewpoint planning methods, AEP [4] is the most representative. We employed an extended RRT approach to compute the next viewpoint with a continuous prediction method, similar to the AEP planner. In the agriculture, Zaenker et al. [20] demonstrated a ROI viewpoint planner (RVP), which constructs voxel maps of the fruit region and uses a utility function based on expected information gained in the fruit region to evaluate candidate viewpoints. Comparative experiments are conducted with AEP and RVP planners to assess and demonstrate the performance of the proposed view planning strategy.

A. Experimental Setup

The robot in the laboratory experiment consists of a 6-DOF collaborative robotic arm equipped with a harvesting gripper at the end-effector, as shown in Fig. 4. A RealSense D435 depth camera is installed at the end of the robotic arm to obtain a voxel map of the picking point region. The

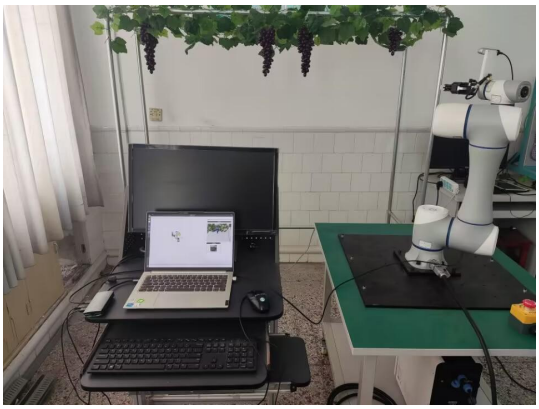


Fig. 4. Physical experiment platform for grape harvesting robot.

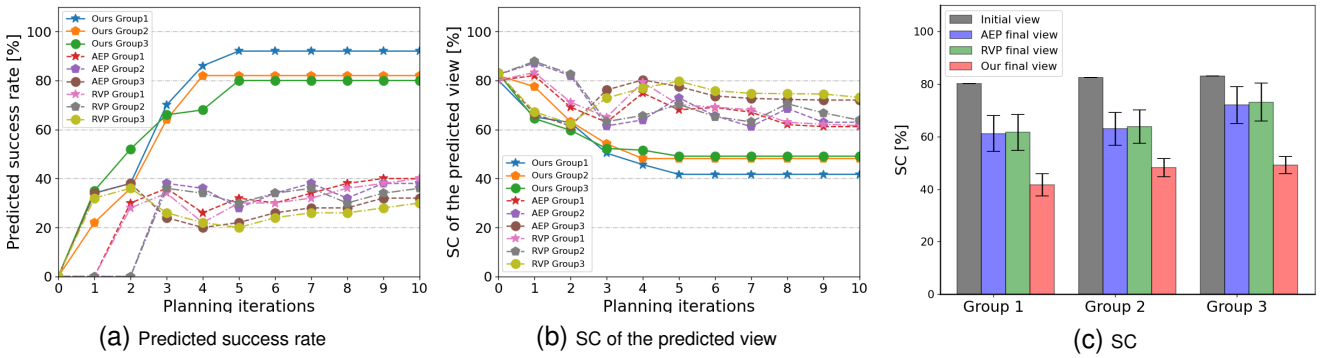
Fig. 5. The experimental groups are with different leaf occlusion angles, denoted as Group1 (Left), Group2 (Center), and Group3 (Right).

SC-NBVP planning algorithm runs on a PC with an Intel i7-10875H CPU with 2.3 GHz and 32 GB of RAM. To evaluate the performance of the SC-NBVP planner under different occlusion conditions, we selected three groups of grape images based on varying degrees of leaf occlusion angles, denoted as Group 1 (Left), Group 2 (Center), and Group 3 (Right), as shown in Fig.5.

The performance of the SC-NBVP planner was validated through two experiments.

- Each set of grapes started with an image perspective as the initial view. All planners independently repeated the view planning algorithm 50 times and recorded metrics such as the success rate of predicting the ideal view and the average SC for each iteration within the 50 planning runs. The prediction success rate refers to the proportion of times within the same iteration among the 50 planning runs when the grape stem was successfully detected.
- Fifty random spatial points were selected as initial viewpoints for testing. These initial viewpoints were chosen such that the fruit stem was occluded. Subsequently, each planner was employed for planning. The success rate of locating the fruit stem within ten planning iterations and the overall time cost of the planning process was recorded.

Before the comparison, we optimized the parameters of the AEP and RVP planners starting with reference values given in [4] and [20]. The parameters were fine-tuned based



(a) Predicted success rate (b) SC of the predicted view (c) SC

Fig. 6. Results obtained by the real grape harvesting robot for different occlusion scenarios during the evaluation experiments in the predicted view.

TABLE I

PLANNER PARAMETERS

	AEP	SC-NBVP	RVP
RRT step size L	0.1m	0.1m	n.a
RRT max.nodes N_{max}	10	10	n.a
Min.gain g_{zero} L	2.0	n.a	2.0
Number of candidate views n	50	50	50
Degrees coefficient λ	0.01m	0.01m	n.a
Sampling spherical space radius R	n.a	0.1m	n.a
Octomap resolution r	0.005m	0.005m	0.005m
Discrete point spacing σ	0.0125m	0.0125m	0.0125m
R_{min}	0.3m	0.3m	0.3m
R_{max}	1m	1m	1m

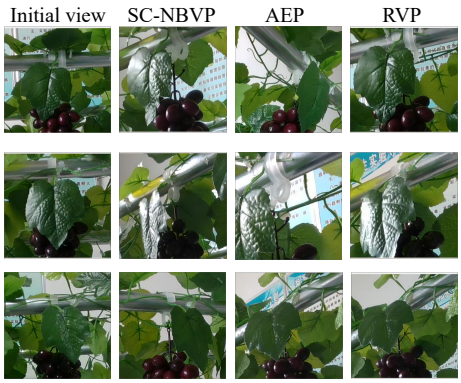


Fig. 7. Comparison of the AEP and RVP planners to the proposed planner's initial and final views, respectively. The AEP and RVP planners only sometimes find the fruit stem, while the final images from our proposed method are more convenient for finding the picking operation point.

on grape harvesting scenarios, and the adjusted parameters are presented in Table I. We incorporated a discrete point space into the parameters to compute the SC metric. Considering computation time and planner performance, we set the number of candidate viewpoints, denoted as n, to 50. The ROI for the AEP and SC-NBVP planners was assigned to the picking point region, while the RVP planner was specified for the fruit region.

B. Results and Analysis

Fig.6a shows that within five iterations, the proposed method achieves a prediction success rate of 80%. In comparison, the AEP and RVP planner has less than a 40% success rate, indicating that the SC-NBVP algorithm can find the ideal view with a higher probability within a limited number of iterations. Examining the changes in the success rate of the predicted view and the SC values in Fig.6a and Fig.6b, it can be observed that the SC metric exhibits a decreasing trend before and after planning. The lower the SC value, the higher the success rate in predicting view, suggesting that the proposed SC metric can reflect the occlusion effect in the picking point region. We also calculated the SC of the initial and final view before and after planner execution, as shown in Fig.6c. In the three experimental groups, the SC for the SC-NBVP planner decreased by 35.25%, 34.43%, and 31.55% compared to the initial view. However, the AEP planner's SC only decreased by 15.91%, 18.66%, and 11.11%, and the RVP planner's SC decreased by 15.49%, 17.84%, and 10.01%, indicating that all planners can reduce occlusion in the picking point region. Still, the SC-NBVP planner is more effective than AEP and RVP. During the experiments, we also recorded images of the picking point region from the robot's view before and after the planner's operation (see Fig. 7). It is clear that, compared to the AEP and RVP planners, the final view determined by the SC-NBVP planner was more conducive to accurately locating the picking point.

TABLE II

PLANNER PERFORMANCE COMPARISON IN LABORATORY EXPERIMENTS

Metric	Group	SC-NBVP	AEP	RVP
Success Rate [%]	Group 1	88%	40%	40%
Time Consumption [s]	Group 1	57.72	111.98	114.32
Success Rate [%]	Group 2	82%	38%	36%
Time Consumption [s]	Group 2	55.96	111.96	116.28
Success Rate [%]	Group 3	80%	38%	34%
Time Consumption [s]	Group 3	55.65	110.31	121.32

In the 2nd experiment, the experimental results are presented in Table II. As can be seen, the SC-NBVP planner outperforms others in the success rate metric. It surpasses the AEP planner by 48%, 44%, and 42%, and the RVP

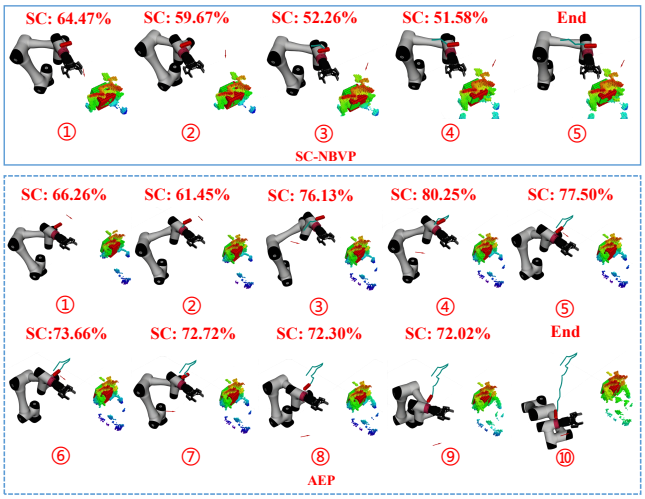


Fig. 8. **Comparison of the AEP planner to the SC-NBVP planner running process.** The sequence number indicates the running iteration and SC represents the Spatial Coverage Rate Metric of the predicted view.

planner by 48%, 46%, and 46%, respectively. An explanation for that could be the impact of the scoring function for candidate viewpoints. In the time consumption metric, the SC-NBVP planner reduces the time by 54.26s, 56s, and 54.66s compared to the AEP planner and by 56.6s, 60.32s, and 65.67s compared to the RVP planner. This indicates that the method proposed has a higher harvesting success rate in less computation time.

C. Impact of Score Function Improvement

To evaluate the improved score function on the performance of the view planner, the entire execution process of planners was recorded separately. Fig.8 provides a real-time comparison of the SC-NBVP and AEP planners. It is clear from the figure that the SC-NBVP planner's occlusion metric SC exhibits a gradual decreasing trend, while the SC of the AEP planner fluctuates. The primary reason for this phenomenon is the different principles underlying these two planners. SC-NBVP selects an ideal view based on a score function that combines SC and motion cost. In contrast, AEP aims to reduce unknown voxels as much as possible, which may not necessarily lead to an immediate reduction in occlusion. Regarding planning iterations, the AEP planner has approximately twice as many planning iterations as the SC-NBVP planner, indicating that the SC-NBVP planner has a shorter runtime.

D. Influence of the Improved Mapping Method

To assess the contribution of the improved mapping method, the time required for a single mapping update was separately recorded for Group 1 to Group 3 using both local mapping (LM) and global mapping (GM) methods. A single map update refers to the time required to refresh a regional map once the depth camera obtains a single depth image. Specific results are presented in Table III. The table shows that the GM method takes 1-2 seconds for map updates, while the QM method takes less than 0.2 seconds, achieving an efficiency improvement of approximately 12 times.

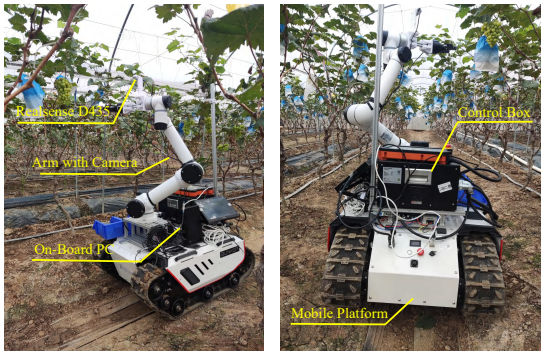


Fig. 9. Outdoor experiment platform for grape harvesting robot.

Fig. 10. Three different occlusion scenarios, denoted as Scene1 (Left), Scene2 (Center), and Scene3 (Right).

TABLE III
SINGLE MAPPING TIME FOR DIFFERENT METHODS

	LM Time Consumption [s]	GM Time Consumption [s]
Group 1	0.136	1.746
Group 2	0.162	2.072
Group 3	0.153	1.982

E. Real-World Outdoor Experiments

In additional experiments, we deployed our method onto a real grape-harvesting robot. The robot is equipped with a 6-DOF Aubo robotic arm mounted on a BUNKER-tracked chassis, as shown in Fig.9. For volumetric mapping, a Realsense D435 depth camera is installed at the end of the robotic arm. The viewpoint planning algorithm runs on an onboard PC with an Intel i7-1165G7 CPU operating at a frequency of 4.7GHz and 16 GB of RAM. The robot can control the movement of the mobile platform. Currently, the viewpoint planner is configured only to control the motion of the robotic arm. Experiments were conducted outdoors with three different occlusion scenarios, as shown in Fig.10. For each scenario, 50 random spatial points were selected as initial viewpoints, requiring the initial viewpoint to have the grape stems occluded by grape leaves. Subsequently, all planners were run separately, and metrics such as the success rate of locating the fruit stem within ten planning iterations and the overall time cost of the planning process were recorded.

Table IV displays the results of the outdoor experiments. All planners exhibit a decrease in the success rate compared to laboratory experiments, particularly evident in the case of the RVP planner in Scene 2. This phenomenon can be explained by the larger foliage in the outdoor experiments compared to the laboratory environment, leading to more severe occlusion of stems. Additionally, the ROI for the RVP planner is defined in the fruit region rather than the picking points region, and the expected information gain in the fruit region may not necessarily reduce the occlusion of stems in the picking region. The planning time for all planners increases as the decrease in success rate results in a longer average planning time. Nevertheless, the experimental results still demonstrate that our method achieves a higher picking success rate in less time than the AEP and RVP planners.

TABLE IV
PLANNER PERFORMANCE COMPARISON IN OUTDOOR EXPERIMENTS

Metric	Scenarios	SC-NBVP	AEP	RVP
Success Rate [%]	Scene 1	76%	34%	32%
Time Consumption [s]	Scene 1	65.83	128.75	132.16
Success Rate [%]	Scene 2	62%	30%	20%
Time Consumption [s]	Scene 2	69.52	135.83	145.67
Success Rate [%]	Scene 3	68%	32%	30%
Time Consumption [s]	Scene 3	63.79	127.54	136.78

V. CONCLUSION

To address the issue of grape stem occlusion during harvesting, we propose a view planning method based on an active vision strategy. The key to this method is the SC-NBVP view planner. This planner employs a random sampling approach to generate a list of candidate viewpoints around the picking point region and then continuously predicts the ideal view to guide the robot's view changes under a scoring function that considers both occlusion and motion costs. The concept of Spatial Coverage Rate Metric is introduced creatively to assess the occlusion of leaves on grape stems in the picking point region. We demonstrated in laboratory and outdoor experiments that our method achieved a higher harvesting success rate in less time, especially when fruit leaves occlude grape stems. In the future, we will conduct further research to enhance the efficiency of the viewpoint planner, aiming to improve its applicability.

REFERENCES

[1] E. Vrochidou, K. Tziridis, A. Nikolaou, T. Kalampokas, G. A. Papakostas, T. P. Pachidis, S. Mamalis, S. Koundouras, and V. G. Kaburlasos, "An autonomous grape-harvester robot: integrated system architecture," *Electronics*, vol. 10, no. 9, p. 1056, 2021.

[2] Y. Jiang, J. Liu, J. Wang, W. Li, Y. Peng, and H. Shan, "Development of a dual-arm rapid grape-harvesting robot for horizontal trellis cultivation," *Frontiers in Plant Science*, vol. 13, p. 881904, 2022.

[3] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "Octomap: An efficient probabilistic 3d mapping framework based on octrees," *Autonomous robots*, vol. 34, pp. 189–206, 2013.

[4] M. Selin, M. Tiger, D. Duberg, F. Heintz, and P. Jensfelt, "Efficient autonomous exploration planning of large-scale 3-d environments," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1699–1706, 2019.

[5] P. Hoseini, S. K. Paul, M. Nicolescu, and M. Nicolescu, "A one-shot next best view system for active object recognition," *Applied Intelligence*, vol. 52, no. 5, pp. 5290–5309, 2022.

[6] J. Hu and P. R. Pagilla, "View planning for object pose estimation using point clouds: An active robot perception approach," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9248–9255, 2022.

[7] R. Menon, T. Zaenker, N. Dengler, and M. Bennewitz, "Nbv-sc: Next best view planning based on shape completion for fruit mapping and reconstruction," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 4197–4203.

[8] T. Zaenker, J. Rückin, R. Menon, M. Popović, and M. Bennewitz, "Graph-based view motion planning for fruit detection," *arXiv preprint arXiv:2303.03048*, 2023.

[9] X. Zeng, T. Zaenker, and M. Bennewitz, "Deep reinforcement learning for next-best-view planning in agricultural applications," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2323–2329.

[10] C. McCool, I. Sa, F. Dayoub, C. Lehnert, T. Perez, and B. Upcroft, "Visual detection of occluded crop: For automated harvesting," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 2506–2512.

[11] J. Kierdorf, I. Weber, A. Kicherer, L. Zabawa, L. Drees, and R. Roscher, "Behind the leaves: estimation of occluded grapevine berries with conditional generative adversarial networks," *Frontiers in artificial intelligence*, vol. 5, p. 830026, 2022.

[12] I. Sa, C. Lehnert, A. English, C. McCool, F. Dayoub, B. Upcroft, and T. Perez, "Peduncle detection of sweet pepper for autonomous crop harvesting—combined color and 3-d information," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 765–772, 2017.

[13] C. Lehnert, C. McCool, I. Sa, and T. Perez, "Performance improvements of a sweet pepper harvesting robot in protected cropping environments," *Journal of Field Robotics*, vol. 37, no. 7, pp. 1197–1223, 2020.

[14] Y. Pan, F. Magistri, T. Läbe, E. Marks, C. Smitt, C. McCool, J. Behley, and C. Stachniss, "Panoptic mapping with fruit completion and pose estimation for horticultural robots," *arXiv preprint arXiv:2303.08923*, 2023.

[15] E. Krotkov and R. Bajcsy, "Active vision for reliable ranging: Cooperating focus, stereo, and vergence," *International Journal of computer vision*, vol. 11, no. 2, pp. 187–203, 1993.

[16] C. Stachniss, G. Grisetti, and W. Burgard, "Information gain-based exploration using rao-blackwellized particle filters." in *Robotics: Science and systems*, vol. 2, 2005, pp. 65–72.

[17] J. Rückin, L. Jin, F. Magistri, C. Stachniss, and M. Popović, "Informative path planning for active learning in aerial semantic mapping," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 11 932–11 939.

[18] E. Palazzolo and C. Stachniss, "Effective exploration for mavs based on the expected information gain," *Drones*, vol. 2, no. 1, p. 9, 2018.

[19] A. Bircher, M. Kamel, K. Alexis, H. Oleynikova, and R. Siegwart, "Receding horizon" next-best-view" planner for 3d exploration," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 1462–1468.

[20] T. Zaenker, C. Smitt, C. McCool, and M. Bennewitz, "Viewpoint planning for fruit size and position estimation," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 3271–3277.

[21] T. Zaenker, C. Lehnert, C. McCool, and M. Bennewitz, "Combining local and global viewpoint planning for fruit coverage," in *2021 European Conference on Mobile Robots (ECMR)*. IEEE, 2021, pp. 1–7.

[22] L. Luo, W. Yin, Z. Ning, J. Wang, H. Wei, W. Chen, and Q. Lu, "In-field pose estimation of grape clusters with combined point cloud segmentation and geometric analysis," *Computers and Electronics in Agriculture*, vol. 200, p. 107197, 2022.