参赛学生姓名：苏泓熹，黄景怡

中学：上海星河湾双语学校

省份：上海市

国家/地区：中国

指导老师姓名：朱佳铭，赵春宇

指导老师单位：上海星河湾双语学校，上海交通大学

论文题目：Geometry-Derived Fruit-Relevance Weighting for Next-Best-View Planning: Active Reconstruction of Occluded Strawberries for Robotic Harvesting

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Geometry-Derived Fruit-Relevance Weighting for Next-Best-View Planning: Active Reconstruction of Occluded Strawberries for Robotic Harvesting

Abstract

Leaves, stems, and neighboring fruits occlude strawberry surfaces visible by the RGB-D camera, which increases the complexity of geometric estimation on the target fruit for robotic grasping. Conventional next-best-view (NBV) planning chooses views to complete the fruit surface without explicitly distinguishing the target fruit surface from its surrounding structures. This paper presents geometry-derived fruit-relevance weighting for next-best-view planning (GFR-NBV), which prioritizes observations of fruit surfaces and improves the efficiency of target fruit reconstruction.

GFR-NBV estimates fruit-relevance weights based on local surface geometry without using color or texture of fruits. Strawberry surfaces have curvatures in both principal directions, leaf surfaces are almost flat, and stem surfaces curve in only one direction. Curvedness and shape index are derived from the principal curvatures and assign continuous fruit-relevance weight by comparing with reference statistics of strawberry fruits. Virtual rays from each candidate pose pass through the occupancy map to identify potentially visible unknown voxels. Then, the score of each candidate view is calculated by adding the expected entropy reductions of voxels that are weighted by their propagated fruit-relevance weights. Therefore, the candidate pose with the highest score is selected as the next best view.

After several GFR-NBV iterations, the observations provide a more complete 3D reconstruction of the target fruit. To obtain the parameters for robotic grasping, an initial axis is estimated from the fruit points, and the cross sections perpendicular to this axis are fitted with circles to identify the broadest section, whose center and radius define the fruit shoulder. The gradual reduction of cross-sectional radii defines the fruit tip, and finally the directed axis from the shoulder to tip is determined for grasping pose.

Simulations were conducted using 30 random scenes at each of three occlusion levels. The simulation results show that GFR-NBV mainly improves early-view reconstruction efficiency and reaches the 70% fruit-surface coverage target with the same or fewer observations than unweighted NBV. Physical experiments were carried out on a robotic platform under 7 strawberry-model scenes. GFR-NBV achieved the 70% point-reference earlier in 5 scenes and tied in the remaining 2 scenes, with the advantage under moderate and heavy occlusion.

Keywords: Next-best-view planning, Geometry-derived fruit-relevance weighting, Active reconstruction, Robotic strawberry harvesting.

2026 S.T. Yau High School Science Award
仅用于2026年丘成桐中学科学奖公示

Contents

1. INTRODUCTION 1

1.1 Research Background 1

1.2 Related Work and Research Gap 1

2. METHODOLOGY OF GFR-NBV 3

2.1 Assumptions 3

2.2 Geometry-Derived Fruit-Relevance Weighting 3

2.3 GFR-NBV Algorithm 5

2.4 Reconstructed Fruit Geometry and Pose Estimation 7

3. SIMULATION AND VALIDATION 10

3.1 Simulation Environment 11

3.2 Validation of Geometry-Derived Fruit-Relevance Weights 13

3.3 GFR-NBV Performance 15

3.4 Grasp-Oriented Pose Estimation 17

4. PHYSICAL EXPERIMENTS 18

4.1 Experimental Setup 18

4.2 Experimental Procedures 19

4.3 Results and Discussion 20

5. CONCLUSIONS AND FUTURE WORK 22

5.1 Conclusions 22

5.2 Limitations and Future Work 22

REFERENCES 23

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

1. Introduction

1.1 Research Background

Robotic strawberry harvesting is important to reduce labor and improve high-quality strawberry production. However, strawberries grow in field randomly in size, maturity, orientation, and distribution, making robotic harvesting particularly difficult [1][2]. Occlusion due to leaves, stems and neighboring fruits leads a fixed camera only capture part of fruit's surface, and consequently the incomplete geometry reduces the reliability of fruit localization, pose estimation, and grasp planning [1]. As the surface visibility depends strongly on camera pose, the camera must actively acquire views by choosing sensing poses. Although multi-view sensing and viewpoint selection have applied to the robotic fruit harvesting, their robustness and efficiency remain insufficient [3]. This motivates the study of next-best-view planning for strawberry harvesting.

Due to partial observations, the next-best-view (NBV) planning evaluates a set of feasible camera poses and chooses the one expected to provide the most useful additional information to complete the 3D model [4][5]. For strawberry harvesting, NBV means selecting a camera pose that reveals previously unobserved target-fruit surfaces rather than only increasing the coverage of surrounding foliage and neighboring fruits.

In continuous robotic strawberry harvesting, sensing and camera motion are repeated for each fruit. Reconstruction quality and sensing time directly affect the success rate and harvesting efficiency. Hence, an NBV planning should satisfy three requirements: (1) prioritize the target fruits rather than the entire plant scene, (2) acquire sufficiently complete fruit surface geometry for pose estimation and grasp planning, and (3) achieve the required reconstruction quality with few views.

1.2 Related Work and Research Gap

In a recent review, NBV methods were categorized into different types according to how information is used to evaluate candidate views [6]. For reconstruction of occluded strawberries, the most relevant methods are called the uncertainty-based, learning-based, prediction-based and task-aware NBV.

The uncertainty-based NBV represents a scene using a probabilistic occupancy map containing free, occupied, and unknown voxels, and then ranks candidate views according to expected entropy reduction [7]. This NBV is interpretable and requires no training. However, the evaluation is generally applied to the entire scene, and therefore an unknown foliage, i.e., leaves and stems, voxel may receive the same priority as an unknown region of the target fruit. This may navigate camera toward plant structures with little to fruit reconstruction.

The learning-based NBV reduces computational cost of the viewpoint evaluation by predicting information gain from accumulated point clouds. Most learning-based NBV rely on neural networks to evaluate candidate views [8][9]. However, as the gain is evaluated over the entire scene, these methods do not explicitly prioritize the target fruit.

The prediction-based NBV solves occlusion by estimating hidden geometry before selecting the next view [10]. Recent agricultural studies have applied this prediction to the

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

plant reconstruction and point-cloud completion to partially occluded strawberries [11][12]. However, the occluded fruit surface is actually predicted from strawberry shapes learned from training data rather than observed directly by the camera. If the target strawberry differs from those learned shapes, e.g., asymmetric shape or locally deformed surface, the reconstructed surface may be inaccurate, leading to errors in pose estimation and grasp planning.

The task-aware NBV directs viewpoint to prioritize the task-relevant regions, such as strawberry fruits. Attention-driven and semantics-aware NBV employ the predefined regions of interest or plant-part labels to focus viewpoint on relevant structures [13][14]. These methods improve target selection but rely on defined semantic regions or trained detectors. For avocado harvesting, an avocado is firstly detected and then a ring of candidate camera positions around the avocado are generated to find an un-occluded picking view [15].

Existing NBV methods meet only part of requirements of robotic strawberry harvesting. They either evaluate the whole scene, indicate hidden surfaces from shapes learned from training data, or rely on detailed predefined regions or trained detectors. A gap therefore remains for a target-focused method that efficiently reconstructs the surface of an occluded strawberry without fine-grained semantic labels or trained detectors.

To fill this gap, this study starts from a coarse target region and derives fruit-relevance weights from local point-cloud geometry within a target region and incorporates them into occupancy-based information gain for NBV planning. This guides camera views toward missing fruit surfaces without a trained semantic detector, while the fruit geometry is reconstructed from RGB-D observations to support pose estimation and grasp planning for the robotic harvesting.

The central question of this study is whether local surface geometry can provide a continuous target-relevance weight that redirects unweighted NBV toward missing strawberry surfaces without semantic labels. We further test whether the geometric fruit-relevance weighting reduces the observations required to obtain sufficient fruit geometry for grasp-oriented pose estimation.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

2. Methodology of GFR-NBV

2.1 Assumptions

The proposed geometry-derived fruit-relevance weighting next-best-view planning is called GFR-NBV, which is developed under the following assumptions:

(1) Before NBV planning, at least part of the target fruit is visible in the initial RGB-D observation and a coarse 3D region containing the target strawberry has been identified.
(2) Pose of RGB-D camera relative to the reconstruction map is known at every observation.
(3) The target region remains nearly static during the process.
(4) The visible point cloud has sufficient density and quality to enable a reliable estimation of local surface normal and curvature.
(5) No trained detector, semantic segmentation network, or learned shape-completion model is used in the proposed framework.

2.2 Geometry-Derived Fruit-Relevance Weighting

2.2.1 Physical Basis

The physical basis of the geometry-derived relevance arises from the different physical shapes of strawberries, leaves, and stems. Most visible regions of a strawberry have smooth convex surface that bends in two directions with two principal curvatures k_1 and k_2. Conversely, a leaf is close to a plane, so both curvatures are small. And a stem is approximately cylindrical, which bends strongly across its width but slightly along its axis.

Curvature-based features have been used to classify fruit, leaf, and stem in agricultural point clouds [16][17]. Those methods use trained classifiers, while GFR-NBV converts geometry directly into a continuous weight for the viewpoint planning.

These geometric differences are useful features for separating fruit-like surfaces from others. However, folded leaves, irregular fruits and other circular background objects may produce geometric features similar to the strawberry fruits. The geometric information only assigns a continuous relevance weight rather than make the classification of fruits and foliage. The weight represents local fruit-likeness.

2.2.2 Local Principal Curvature Estimation

For each observed point $\mathbf{x}_i$, its K nearest neighboring points can define a local surface patch. Before local geometry estimation, the accumulated point cloud was down-sampled using a 0.005 m voxel size, and $K = 30$ nearest neighbors were used for local surface fitting. The unit surface normal $\boldsymbol{n}_i$ is estimated by the covariance analysis [18]. The positive direction of the normal is chosen to be consistent with the known camera viewing direction.

The neighborhood points are put into a local coordinate system centered at $\mathbf{x}_i$, a local tangent plane, and aligned with this normal, as shown in Figure 1. A quadratic function is fitted to these neighboring points by least squares. The 2×2 Hessian matrix $\mathbf{H}_i$ of the fitted quadratic function contains the second derivatives of the surface, describing the local bending of the surface patch. Under the small-slope approximation, the two eigenvalues of the Hessian matrix approximate the principal curvatures k_1 and k_2, ($k_1 >= k_2$).

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

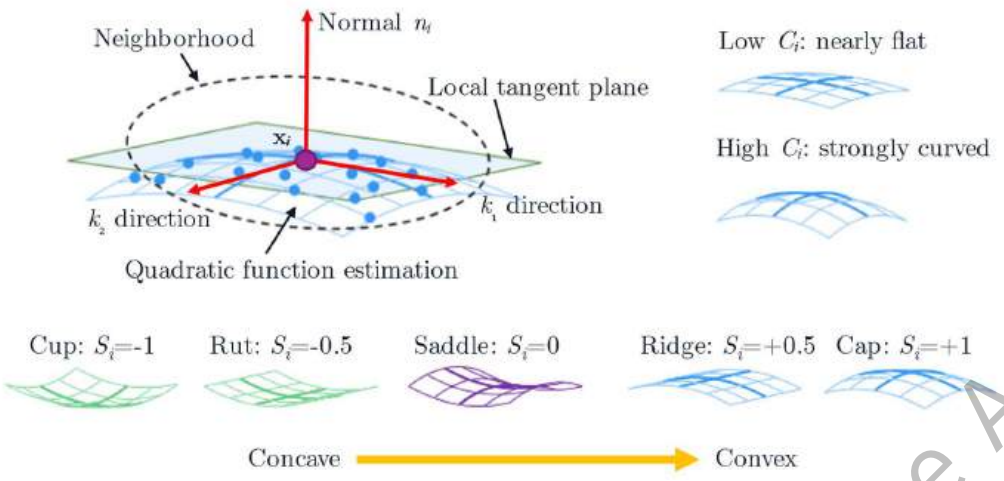


Figure 1 Schematic diagram of local principal curvatures k_1 and k_2, curvedness C, and shape index S for fruit-relevance estimation.

2.2.3 Curvedness and Shape Index

The two principal curvatures are converted into curvedness C and shape index S to describe the local geometry [19].

Curvedness is defined as the root of mean square of the two principal curvatures:

$$C = \sqrt{\frac{k_1^2 + k_2^2}{2}}. \tag{1}$$

The curvedness measures the overall bending strength of the local surface patch. It is always non-negative with the units of inverse length. As shown in Figure 1, a nearly flat region has a low C, while a strongly curved region has a high C. Specifically, a spherical surface of radius R has $C = 1/R$, and therefore a larger R with a smaller C means nearly flat, while a smaller R with a larger C means bending strongly.

Furthermore, the shape index is defined as

$$S = \frac{2}{\pi}\operatorname{atan2}\left(k_1 + k_2, k_1 - k_2\right), \qquad -1 \le S \le 1. \tag{2}$$

The two-argument function **atan2** is used to preserve the curvature-sign information and avoid division by zero when the two principal curvatures are equal (i.e., k_1 - $k_2 = 0$). Shape index is a dimensionless factor that describes the type of local shape. As shown in Figure 1, from concave to convex, S increases from -1 to 1, corresponding to different bending shapes.

Thus, the two factors, curvedness C and shape index S, provide geometry information for continuous fruit-relevance weight.

2.2.4 Fruit-Relevance Weight

For an observed point i, its difference from the fruit reference can be measured using the normalized squared geometric distance in terms of two factors C and S:

$$D_i^2 = \left(\frac{S_i - S_f}{\sigma_S}\right)^2 + \left(\frac{C_i - C_f}{\sigma_C}\right)^2, \tag{3}$$

where the mean factors of strawberries are S_f and C_f, and their standard deviations are σ_S

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

and σ_C. The first term measures the difference in local shape, and the second term measures the difference in bending strength. Only the dimensionless form of the two normalized terms can be added together. Mathematically, D_i is the Euclidean distance from (S_i, C_i) to (S_f, C_f) in the normalized shape–curvedness space. Thus, D^2=0 means the measured geometry exactly matches the fruit reference, while a larger D^2 indicates a greater geometric difference.

The relevance weight should approach 1 for a close geometric match and decrease toward 0 as the geometric difference increases. So, we construct the geometric distance into a Gaussian-shaped function as

$$w_i = \exp\left(-\frac{D_i^2}{2}\right) = \exp\left(-\left(\frac{S_i - S_f}{\sqrt{2}\sigma_S}\right)^2\right)\exp\left(-\left(\frac{C_i - C_f}{\sqrt{2}\sigma_C}\right)^2\right), \qquad 0 < w_i \le 1\,. \tag{4}$$

This fruit-relevance weight is determined by both the shape similarity and the curvedness similarity. The Gaussian-shaped weight function gives the maximum weight ($w_i = 1$) when the measured geometry matches the fruit reference, and smoothly reduces the weight as the geometric difference increases. Consequently, small deviations of real fruits or point-cloud noise do not cause an abrupt change in relevance, while surfaces that differ substantially from the strawberry reference receive low weights. Since the shape-index and curvedness terms are both considered in the weight function, a high weight requires similarity in both local shape and bending magnitude.

2.3 GFR-NBV Algorithm

GFR-NBV is based on the standard occupancy-based NBV model and modifies its viewpoint selection strategy according to the fruit relevance in candidate views.

2.3.1 Probabilistic Occupancy and Information Gain

GFR-NBV adopts the standard occupancy mapping and ray-casting algorithms in conventional NBV planning [20]. The coarse target region is input into NBV planning as voxels, and each voxel j is assigned an occupancy probability p_j. 0 and 1 indicate free and occupied space respectively, and 0.5 indicates that the voxel has not been sufficiently observed, meaning uncertainty. The occupancy probabilities are updated after each observation. The occupancy map used for NBV planning has a voxel size of 0.01 m.

The uncertainty of each voxel is quantified using the binary entropy [7]:

$$h_j = -p_j \log_2 p_j - (1 - p_j)\log_2(1 - p_j), \tag{5}$$

where entropy h_j is 0 to 1 bit. Reach 1 when $p_j = 0.5$, indicating the voxel is most uncertain, and approach 0 when p_j either 0 or 1, indicating the voxel is classified as free or occupied. Therefore, a new observation view v is effective for voxel j if the reduction of the voxel's entropy $\Delta h_j(v)$ or increase the information gain $IG_j(v)$ is positive, writing as

$$IG_j(v) = \Delta h_j(v) = h_j^{\text{before}} - h_j^{\text{after}} > 0\,. \tag{6}$$

For each candidate view v, rays are projected virtually through the current occupancy map to evaluate voxels expected to be visible without physically moving the camera. Let $R(v)$ denote the set of those voxels, the scene-level information gain is calculated as

2026 S.-T. Yau High School Science Award 仅用于2026正式提交科学论文公示

$$IG(v) = \sum_{j \in R(v)} \Delta h_j > 0 \tag{7}$$

A greater $IG(v)$ means that the candidate view v is expected to remove more uncertainty from the map, and thus the view with the greatest IG is chosen as the next best view.

However, Eq.(7) considers only the reduction of uncertainty of the scene-level view, including the target strawberry, surrounding foliage, and background. For GFR-NBV, we must add the weight of fruit-relevance as well as the uncertainty reduction to evaluate the best next view.

2.3.2 Fruit-Relevance-Weighted View

A fruit-relevance weight w_i is applied to each observed surface point $\mathbf{x}_i$. However, candidate-view evaluation concerns unknown voxels that have not yet been measured. Therefore, the relevance of an unknown voxel is estimated from the geometric weight available at nearby observed points.

Let $\mathbf{c}_j$ denote the unknown voxel j, and let O_t the set of observed surface points at planning step t. For each unknown voxel within the coarse target region R, its nearest observed point i^* is identified, and the point weight w_{i^*} is assigned to the voxel:

$$i^*(j) = \arg\min_{\mathbf{x}_i \in \mathcal{O}_t} \left\| \mathbf{c}_j - \mathbf{x}_i \right\|_2, \qquad \hat{w}_j = w_{i^*(j)}. \tag{8}$$

This estimation of weights for unknown voxels assumes that local surface relevance changes gradually around the observed boundary. It does not predict the complete shape of the hidden strawberry. Instead, this estimation indicates which unknown regions within R should receive greater attention during viewpoint planning. Once an unknown region is observed, its estimated relevance is replaced by a weight calculated directly from the newly measured point-cloud geometry.

For a candidate view v, virtual ray provides the set $R(v)$ of unknown voxels expected to become visible. GFR-NBV evaluates this view by combining the expected entropy reduction $\Delta h_j(v)$ with the estimated fruit relevance $\hat{w}_j$:

$$G_{\mathrm{FR}}(v) = \sum_{j \in R(v)} \hat{w}_j \Delta h_j(v) \tag{9}$$

The two factors in Eq. (9) have complementary meanings. The expected entropy reduction measures how much new spatial information may be obtained, while the fruit-relevance weight measures how useful that information is for target-fruit reconstruction. Hence, an unknown voxel contributes strongly to the view score only when it is both observable from that view and related to the target fruit. Foliage and background regions receive smaller contributions due to small fruit-relevance weights. If all unknown voxels are assigned $\hat{w}_j = 1$, reduces to the conventional information gain, $G_{\mathrm{FR}}(v) = IG(v)$. Thus, GFR-NBV retains the conventional NBV, but redirects scene-level information gain toward completion of the target fruit.

2.3.3 View Selection and Iterative GFR-NBV Algorithm

At planning step t, GFR-NBV uses the current occupancy map M_t, the accumulated point cloud P_t, and a set of feasible candidate views $\boldsymbol{V}_t$. The candidate views are evaluated

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

virtually. The camera remains at its current position during this evaluation. For each candidate view v, ray casting identifies the unknown voxels $R_t(v)$ expected to become visible, and the fruit-relevance-weighted view gain $G_{\mathrm{FR},t}(v)$ is calculated. The next view v^* at the step t is selected as

$$v_t^* = \arg\max_{v \in \boldsymbol{V}_t} G_{\mathrm{FR},t}(v). \tag{10}$$

After all candidate views have been evaluated, the camera moves to the view v^* and acquires a new RGB-D observation. The new point cloud is added to the accumulated observations, and the occupancy map is updated. Local geometric features and fruit-relevance weights are then recalculated using the newly observed surfaces. Thus, estimated relevance for previously unknown regions is replaced by weights from actual RGB-D.

The procedure repeats until the maximum fruit-relevance-weighted gain among all feasible candidate views falls below the stopping threshold of 0.05, or until the maximum number of views $T_{\max}$ is reached. The resulting accumulated fruit surface can subsequently be used for pose estimation and grasp planning.

The algorithm maximizes fruit-relevant information obtained from each selected view. Camera travel is not included in this study, because the main objective is view selection.

2.3.4 Computational Complexity

According to the algorithm of GFR-NBV, the computational complexity is counted based on the three stages of one planning cycle. Let N_o be the number of points in the newly acquired RGB-D observation, N_p the number of points in the accumulated point cloud, N_c the number of candidate views, N_r the number of rays cast for each view, K the number of neighboring points used for local geometry estimation, and L the average number of voxels traversed by one ray. The computational complexity of one GFR-NBV planning cycle is

$$T_{\mathrm{step}} = O(N_o L) + O\left[N_p(\log N_p + K)\right] + O(N_c N_r L). \tag{11}$$

The first term represents occupancy-map updating, which represents each of the N_o observed points generates a ray that traverses approximately L voxels. The second term represents geometric weight calculation. For each of the N_p accumulated points, finding K neighbors requires an average of $O(\log N_p + K)$ operations, while covariance analysis and quadratic fitting require operations proportional to K. The third term represents candidate-view evaluation, where N_c views each cast N_r rays through approximately L voxels. The memory complexity is $O(N_v + N_p)$, where N_v is the number of allocated voxels.

Overall, the additional computation introduced by GFR-NBV is the local geometric analysis and weight calculation, which grows near-linearly with the number of cloud points for fixed K. Virtual ray casting remains the main computational cost, which increases with the numbers of candidate views, rays, and traversed voxels. Thus, GFR-NBV includes the computation of conventional NBV and additional geometry-weighting costs.

2.4 Reconstructed Fruit Geometry and Pose Estimation

The output of GFR-NBV is the fruit-relevant surface which is accumulated from multiple views. After active reconstruction, the accumulated fruit-relevant surface is evaluated for grasp-oriented fruit pose estimation.

2026 S-T Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公开

2.4.1 Physical Assumption and Pose Definition

The purpose of active reconstruction is to identify a suitable grasping region to determine the position and orientation. The pose estimation method requires some assumptions about strawberry geometry.

The strawberry body is assumed to be a smooth, tapered, and approximately axisymmetric object, as shown in Figure 2. The cross-sections of the fruit are perpendicular to the longitudinal axis and can be considered as circles. The largest circle cross-section is defined as the fruit shoulder, which is normally close to the top of fruit body. From this shoulder toward the narrow tip, the radii of cross-sections are expected to decrease smoothly.

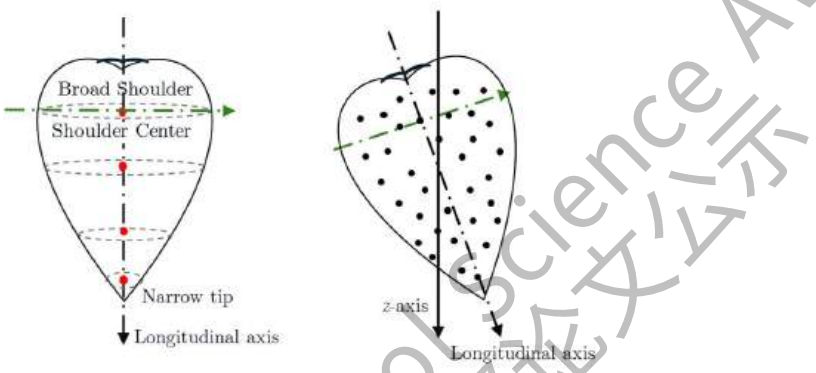


Figure 2 Schematic diagram of strawberry shape in pose estimation

Based on this assumption of strawberry fruits, the centers of fitted cross-sections lie along the longitudinal axis approximately. And the direction from the maximum-radius to minimum-radius indicates the directed axis from the fruit shoulder to the tip. The shoulder is selected as the intended grasping region because its large radius provides a smooth convex surface where the Fin-Ray fingers can properly fit. The larger contact area is expected to distribute the grasping force and reduce contact pressure [21].

Accordingly, the fruit pose is determined by 3 parameters: the shoulder center, the shoulder radius, and the fruit axis direction from shoulder to tip. The shoulder center determines the gripper position, the radius determines the required initial finger opening, and the fruit axis determines the gripper orientation. This model may become unreliable for scenes when fruits are nearly spherical or strongly deformed, or when the reconstructed point cloud covers little of the fruit surface.

2.4.2 Fruit-Relevant Component Extraction

During NBV planning, the weight w_i is used to evaluate how strongly each observed point is related to the target fruit. After the reconstruction is completed, a set of strong fruit-relevant points is gathered for pose estimation.

Although the target region is defined around one strawberry, the fruit-relevant point set P_f may still contain fruit-like points from foliage and neighboring fruits:

$$P_f = \{\mathbf{x}_i \in \mathcal{P} \,|\, w_i \geq \tau\}, \tag{12}$$

where P is the accumulated point cloud from all observations, and $\tau = 0.02$ is the fruit-relevance threshold fixed from calibration scenes.

Because P_f may contain points from multiple spatially separated fruits, spatial clustering is applied to divide P_f into candidate fruit-relevant components C_m:

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

$$C_m\big|_{m=1}^{M} = \text{Cluster}(P_f). \tag{13}$$

Each component C_m is independently examined for the strawberry-like cross-sectional geometry for pose estimation.

2.4.3 Shoulder and Directed-Axis Estimation

For each candidate fruit-relevant component C_m, the grasp-oriented geometry parameters, i.e., the shoulder center, shoulder radius, and the directed axis from shoulder to tip, are estimated from cross-sectional geometry of the fruit-relevant points, as shown in Figure 3.

The weighted centroid of fruit-relevant points can be used as the origin of the local coordinates of this fruit. As the reconstructed surface may be incomplete and nonuniformly sampled, the centroid of C_m may not be the physical center of the fruit. A weighted centroid is evaluated as

$$\mathbf{o}_m = \sum_{\mathbf{x}_i \in C_m} w_i \mathbf{x}_i \Big/ \sum_{\mathbf{x}_i \in C_m} w_i . \tag{14}$$

An initial undirected axis is estimated from the weighted spatial distribution of C_m. Using the weighted centroid as the local origin, the weighted covariance matrix is computed as

$$\mathbf{S}_m = \sum_{\mathbf{x}_i \in C_m} w_i (\mathbf{x}_i - \mathbf{o}_m)(\mathbf{x}_i - \mathbf{o}_m)^T \Big/ \sum_{\mathbf{x}_i \in C_m} w_i . \tag{15}$$

The eigenvector corresponding to the largest eigenvalue of $\mathbf{S}_m$, denoted by $\mathbf{a}_m$, gives the direction of maximum spatial extension and can be considered as the initial fruit axis. However, the direction for grasping has not been determined.

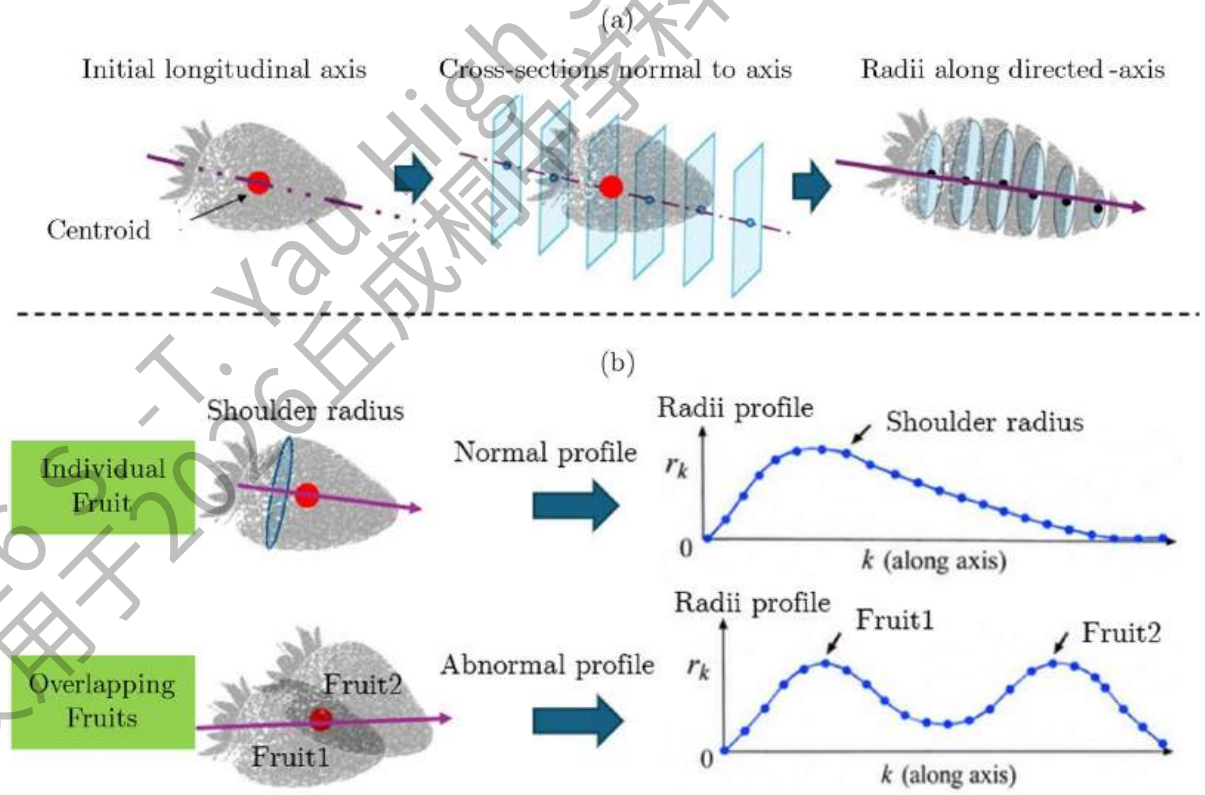


Figure 3 Schematic of cross-sectional analysis for strawberry geometry. (a) A centroid and initial axis are estimated and used to generate cross-sectional slabs for radii and centers estimation. (b) The sectional radius profile exhibits a single broad shoulder followed by progressive tapering, while overlapping fruits produce multiple broad-radius regions.

2026 S-T-Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

The fruit component is then divided into a sequence of thin slabs perpendicular to this initial axis, as shown in Figure 3(a). In each slab, the surface points are projected to the local cross-sectional plane and fitted with a circle. The fitted circle provides a cross-sectional center c_k and a radius r_k. A sequence of cross sections is therefore obtained as $(c_k, r_k)\big|_{k=1}^{K}$. The cross-sectional parameters also provide a geometric consistency check for rejecting components formed by multiple overlapping fruits. A typical single-fruit radius profile is expected to have a smooth sequence of sectional radii, and a single broad region followed by progressive tapering toward one end, as shown in Figure 3(a).

Components involving multiple comparable broad-radius regions, as shown in Figure 3(b), may have abrupt center-line discontinuities, which are regarded as overlapped components and can be excluded from grasping pose estimation.

For a geometrically consistent component, the centers of the cross sections $\mathbf{c}_k$ are fitted by a straight line. This fitted line $\hat{\mathbf{a}}_m$ is defined as the axis:

$$\hat{\mathbf{a}}_m = \mathrm{LineFit}\left(\{\mathbf{c}_k\}_{k=1}^{K}\right). \tag{16}$$

Along the axis, the fitted radii form an axial radius profile, as shown in Figure 3(a). The shoulder is identified from the broadest region of this profile. The corresponding fitted circle defines the shoulder center $\mathbf{c}_s$ and shoulder radius $\mathbf{r}_s$:

$$k_s = \arg\max_k \tilde{r}_k, \quad \mathbf{c}_s = \mathbf{c}_{k_s}, \quad r_s = \tilde{r}_{k_s}, \tag{17}$$

where $\tilde{r}_k$ is the smoothed value based on the initial estimated radius.

The axis direction is determined from the radius variation on the two sides of the shoulder. The side along which the cross-sectional radius gradually decreases toward the smallest cross-section is considered as the fruit tip. The refined axis is then oriented from the shoulder toward the tip which could provide the directed shoulder-to-tip axis.

3. Simulation and Validation

This simulation is designed to validate the proposed methodology. There are three key points to be validated: (1) whether geometry-derived weights provide useful fruit relevance, (2) whether these weights improve active fruit reconstruction, (3) whether the reconstructed geometry supports shoulder and directed-axis estimation.

The overall simulation workflow is shown in Figure 4. The procedure starts with an initial RGB-D observation, the accumulated point cloud and voxel occupancy map. Voxel entropy and geometry-derived fruit-relevance weights are progressively updated in parallel, and each candidate view is evaluated using propagated fruit-relevance weights. The camera then moves to the highest scoring candidate view to obtain new observation and consequently update the reconstruction. The iteration cycle repeats until the maximum number of views is reached, or the information gain drops below the stopping threshold. The accumulated fruit surface is then adopted to estimate the shoulder center, shoulder radius, and directed shoulder-to-tip axis. As for comparison, unweighted NBV follows the same procedure with all relevance weights set to one, while the random selection views replace the view-ranking step with random selection of candidate views.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

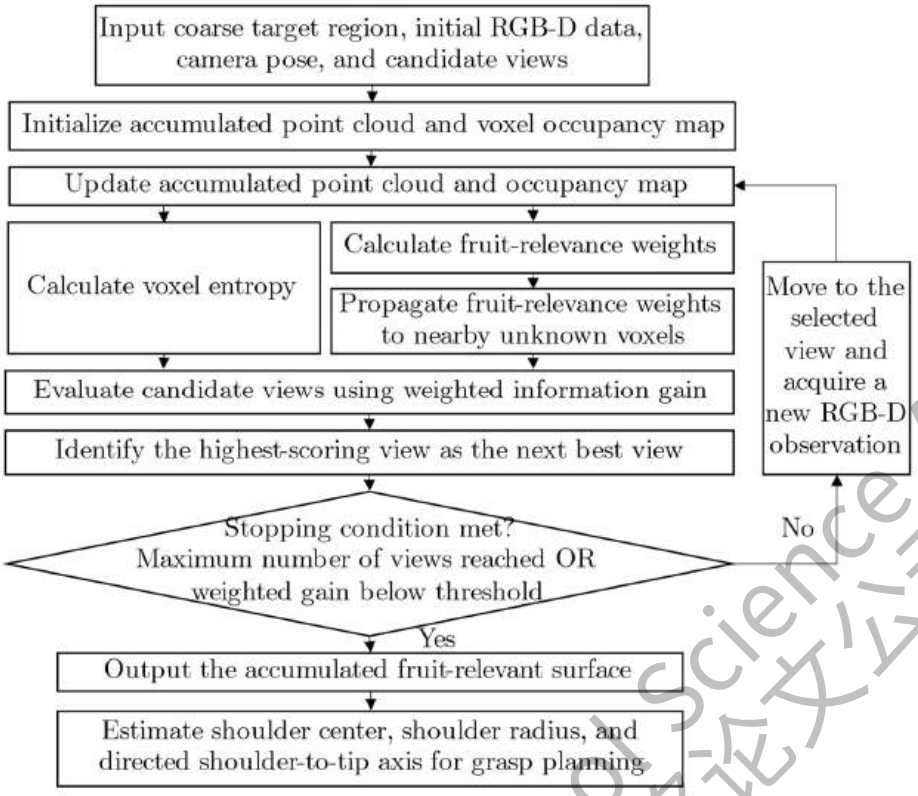


Figure 4 Overall simulation flow

3.1 Simulation Environment

3.1.1 Gazebo Environment and Camera Viewpoints

We set up the simulation environment in Gazebo to provide controlled strawberry geometry, occlusion, camera viewpoints, and ground truth for quantitative evaluation. Each virtual plant consists of a strawberry with leaves, stems, and neighboring fruits. Figure 5 shows scenes with different fruit arrangements and foliage distributions.

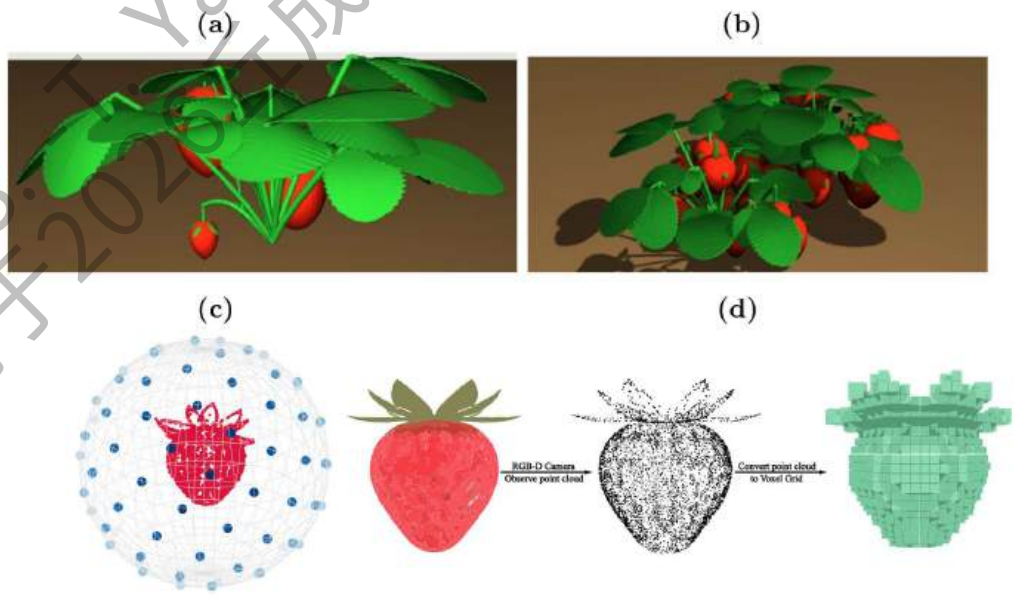


Figure 5 Gazebo environment: (a) a single plant scene, (b) a complex plant scene, (c) 96 candidate views from possible camera poses, and (d) From observation to RGB-D to voxels.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

A simulated RGB-D camera is placed around the plant and toward the target region. The camera resolution is 640 × 400 pixels, and the camera-to-target distance is 0.30 m. Same camera parameters are used for all NBV methods comparison. Simulated depth measurements are converted into point clouds and incorporated into a voxel-based map, which provides the spatial representation for occupancy estimation and view evaluation. This Strawberry is used as an example in this experiment. Only simulated RGB-D observations are supplied to NBV planning, but the ground-truth geometry is reserved for evaluation only.

To evaluate potential viewpoints, 96 candidate camera poses are sampled by using a Fibonacci-sphere pattern and distributed uniformly on a 0.30 m radius sphere around the target region, with all optical axes directed at the target center. This arrangement provides a dense but finite viewpoint set where both unweighted NBV and GFR-NBV can select their next best candidate observation.

3.1.2 Parametric Fruit and Occlusion

The strawberry is constructed with a two-dimensional fruit profile which is represented by a cubic Bezier curve. P_0 and P_3 lie on the axis of revolution and control the upper and the narrow tip. P_1 and P_2 control shoulder width, body curvature, and tapering rate. Rotation of the profile generates a 3D fruit surface, as shown in Figure 6.

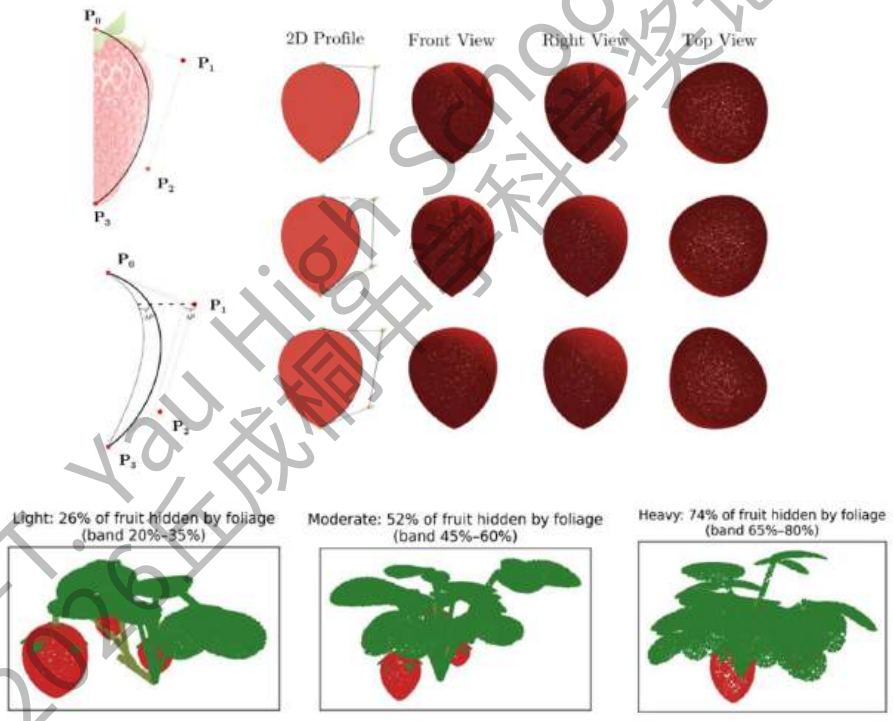


Figure 6 Parametric strawberry model with a Bezier profile fitted to a strawberry with control points P_0-P_3. Different occlusion levels from down-left to down-right.

When the surface, longitudinal axis, and cross-sectional radii are known, they provide the ground truth to evaluate the fruit surface coverage, shoulder position, shoulder radius, and directed axis. The Bezier-curve control points are fitted from physical strawberry profiles to generate different fruit sizes and shapes. As an ideal surface of revolution is perfectly axisymmetric, we add additional radial variation to simulate some slight asymmetry without changing the basic strawberry topology.

The leaf and stem are generated by using a parametric Lindenmayer system (L-system).

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

The L-system has the advantage of modeling rule-based plant topologies [22][23].

Target fruits are modelled in parameters, while leaves and stems are created with bounded random variation in length, curvature, and branching angles. This approach aims to generate geometrically and topologically valid occlusion patterns with known semantic identities and controllable spatial arrangements.

The plant environment is assembled by iteratively adding L-system leaves, stems and attaching Bezier-modeled strawberry fruits at the stems' end points. This can generate a continuous range of fruit occlusion ratios, providing a statistically controlled environments with different occlusion levels, such as light, moderate, and heavy.

The fruit occlusion is quantified directly from the known ground-truth fruit surface. For a camera viewpoint, let $N_{\mathrm{vis}}(v)$ denote the number of ground-truth fruit surface samples visible from that viewpoint and N_{GT} the total number of samples on the complete fruit surface. The visible-surface ratio is

$$R_{\mathrm{vis}}(v) = \frac{N_{\mathrm{vis}}(v)}{N_{\mathrm{GT}}}, \tag{18}$$

and the corresponding occlusion ratio is defined as

$$R_{\mathrm{occ}}(v) = 1 - R_{\mathrm{vis}}(v). \tag{19}$$

Because visibility depends on camera pose, a single scene produces different occlusion ratios from different candidate viewpoints. Simulations are grouped according to the occlusion ratio of the initial observation. To ensure test quality and statistically meaningful results, scenes with extreme occlusion rates, e.g., <20% or >80%, are excluded. Some gaps of occlusion are kept between adjacent intervals to prevent boundary ambiguity and guarantee clear statistical separation among the three occlusion levels. As shown in Figure 6, we set three separated occlusion intervals based on R_{occ}: (1) light occlusion, 20% - 35%, (2) moderate occlusion, 45% - 60%, and (3) heavy occlusion, 65% - 80%.

3.2 Validation of Geometry-Derived Fruit-Relevance Weights

3.2.1 Spatial Distribution of Fruit-Relevance Weights

Figure 7 presents a simulated strawberry plant scene and the observed point cloud colored based on the fruit-relevance weight w. The scene contains fruits, leaves, and stems, and the spatial distribution of the weights on plant surfaces. Each point is assigned a continuous weight using the local geometric information, without using semantic labels.

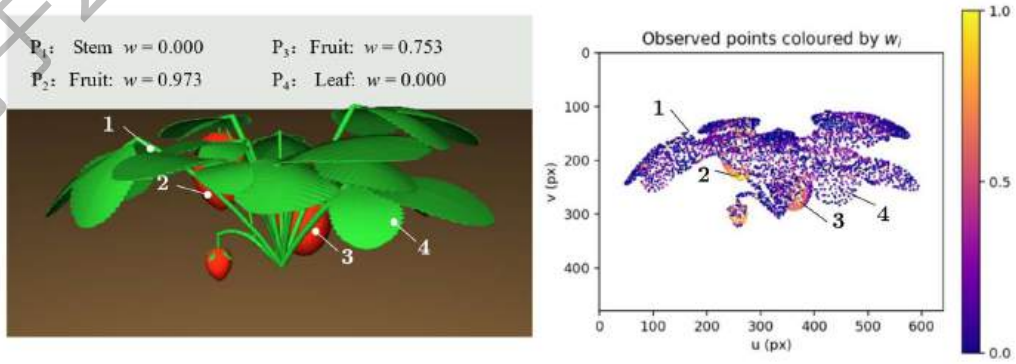


Figure 7 Observation of Plant-132 scene (left), geometric fruit-relevance weights (right).

2026 S. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Table 1 Parameters of weight calculation for typical points

Point	Organ	κ_1 (1/m)	κ_2 (1/m)	S (-)	C (1/m)	Shape factor	Curvedness factor	w
P_1	Stem	1250.3	-7.4	0.496	884.1	0.329	0.000	0.000
P_2	Fruit	92.5	38.3	0.750	70.8	0.980	0.992	0.973
P_3	Fruit	59.6	29.1	0.790	46.9	0.898	0.839	0.753
P_4	Leaf	2.5	-6.9	-0.275	5.2	0.000	0.182	0.000

In this example, a small fruit near the bottom is clearly visible and two large fruits are occluded by foliage. By comparison in weights, the foliage shows low- and intermediate-weight regions mixed with a few of high-weight points, and the fruits show majority high weights mixed with a few of low-weight points. Therefore, some overlap remains between fruit and non-fruit surfaces, showing that local geometry does not provide perfect separation.

Table 1 presents four selected points to illustrate the calculation of weights. The two principal curvatures are used to calculate shape index S and curvedness C. These values are compared with the reference statistics for strawberry fruit surfaces to obtain a shape factor and a curvedness factor. Their product gives the final fruit-relevance weight. Here, the reference parameters were determined from a separate calibration set containing 50 strawberry-model samples. The mean and standard deviation of S and C are $S_f = 0.8$, $\sigma_S = 0.15$ and $C_f = 50$ (m^{-1}), $\sigma_C = 12$ (m^{-1}), respectively.

At the stem point P_1, one principal curvature is much larger in magnitude than the other, and the final weight is therefore close to zero. At the fruit point P_2, both principal curvatures are positive, and the resulting shape index and curvedness closely match the fruit reference, producing a high weight of approximately 0.973. The second fruit point, P_3, has lower shape and curvedness factors, giving a weight of 0.753. This indicates different locations on a fruit have different weights. At the leaf point P_4, both principal curvatures are small in magnitude, giving a low curvedness, so that its shape factor is close to zero and the final weight is also close to zero.

3.2.2 Quantitative Evaluation of Weight Discrimination

Ground-truth labels from the simulation models are used to identify observed points on fruit surface as positive samples and points on leaves and stems as negative samples. Figure 8(a) shows the weight distributions for Plant 132 scene. The two probability distributions of weights of fruit points and non-fruit points are compared. Fruit points are mainly concentrated at higher weights, while non-fruit points are mainly concentrated at lower weights. The overlap shows that some non-fruit points receive high weights and some fruit points receive low weights. Thus, the weights provide a useful but imperfect distinction between the two groups of weights.

Discrimination is quantified using the area under the receiver operating characteristic curve (AUROC). Points with weights at or above a threshold are classified as fruit points. Sweeping this threshold produces a ROC curve of the true positive rate (the fraction of fruit points correctly identified) over the false positive rate (the fraction of non-fruit points incorrectly identified). The 6 scenes have AUROC values ranging from 0.860 to 0.971. Plant 132 has the lowest value of 0.860 and is also a heavily occluded scene, as shown in Figure 8(b).

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

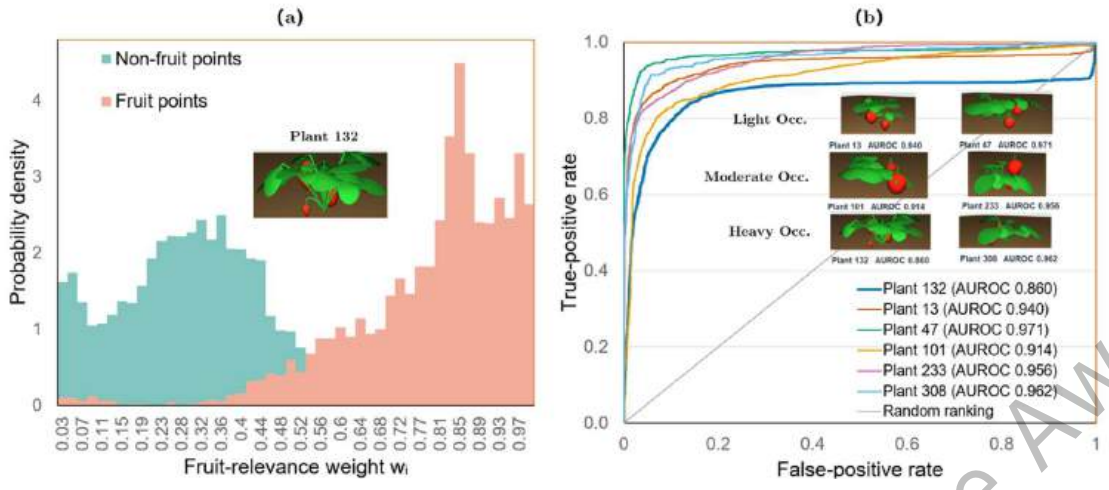


Figure 8 (a) Separately normalized weight distributions for fruit and not-fruit points of Plant-132 scene. (b) ROC curves for 6 different plant scenes including Plant-132.

The other 5 scenes have values above 0.90. Performance does not decrease consistently with occlusion level, and the heavily occluded Plant 308 has an AUROC of 0.962, showing that occlusion level alone does not determine discrimination performance.

3.3 GFR-NBV Performance

3.3.1 Evaluation Criteria

Reconstruction performance is evaluated in terms of fruit surface coverage and observation efficiency. All methods use the same target geometry and initial observation. Ground-truth information is not used in the view selection algorithm.

$N_{\text{covered},k}$ denotes the number of unique reference fruit-surface samples observed after k views. $N_{\text{reference}}$ denotes the whole fruit-surface samples. And the fruit-surface coverage ratio (FCR) SC_k is defined as

$$\text{SC}_k = \frac{N_{\text{covered},k}}{N_{\text{reference}}} \tag{20}$$

Observation efficiency is measured by the number of observations required to first reach a predefined coverage level, $\text{SC}_{\text{target}}$. The common initial view is counted as the first observation. Fewer observations indicate that the selected viewpoints acquire useful fruit-surface information more efficiently.

3.3.2 View Selection and Reconstruction Progress

The contribution of geometric weighting to active reconstruction is evaluated by comparing GFR-NBV with unweighted NBV. GFR-NBV uses the fruit-relevance weights to guide next view. Unweighted NBV follows the same mapping and view-selection procedure, with all relevance weights set to one, isolating the contribution of geometric weighting.

The two methods select different next viewpoints from the first step and follow different observation sequences. GFR-NBV evaluates expected information gain using fruit-relevance weights, while unweighted NBV assigns equal relevance to all evaluated voxels. The comparison shows how geometric weighting changes view selection.

The point-cloud views provide a visual comparison of fruit observations along the two

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

sequences. To assess whether these differences improve reconstruction, accumulated fruit-surface coverage SC_k is provided. The increase of consecutive observations, (SC_k - SC_{k-1}), shows how much additional coverage k view provides. Thus, the camera positions illustrate the view-selection process, while the corresponding coverage values quantify its effectiveness. In this case, it is obvious that the increase of fruit-surface coverage rate of GFR-NBV (75% - 26% = 49%) is much greater than that of unweighted NBV (51% - 26% = 25%) from the second observation.

Figure 9 compares the camera-view sequences selected by GFR-NBV and unweighted-NBV in a representative scene. Both methods begin from the same initial observation, with an FCR of 26%, but select different subsequent viewpoints. GFR-NBV increases the accumulated FCR to 75% at the second view, whereas unweighted NBV reaches only 51%. GFR-NBV therefore exceeds the 70% reconstruction target after one additional observation, while unweighted NBV remains below the threshold until the fifth view in this example. By the fifth observation, the difference decreases, with coverage reaching 92% and 89%, respectively. This comparison indicates that geometric weighting mainly improves early acquisition of useful fruit-surface information rather than producing a large difference in final coverage.



Figure 9 Representative camera-view sequences and accumulated fruit-surface coverage for (a) GFR-NBV and (b) unweighted NBV.

Figure 10 compares fruit-surface coverage ratio against the number of views under light, moderate, and heavy occlusion. Each occlusion level includes 30 sample scenes, and the solid curves show the mean coverage for each method. The first acquired view is the initial observation shared by all methods within each scene.

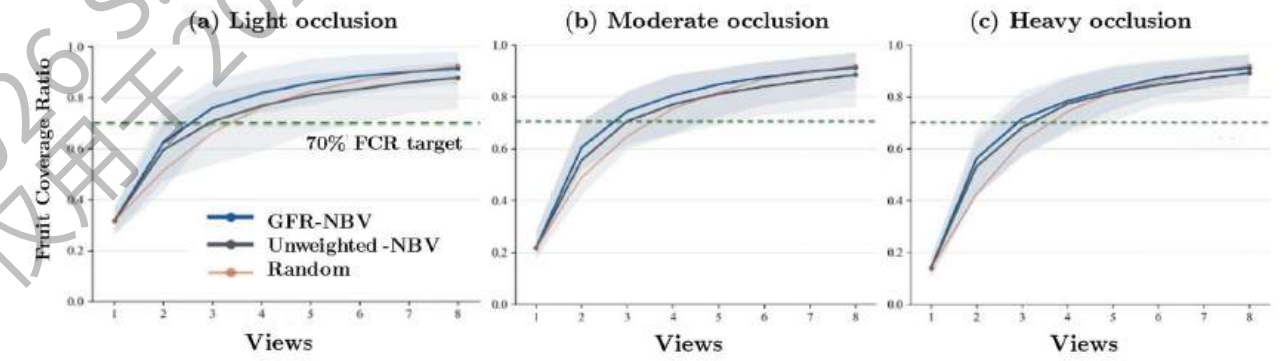


Figure 10 Fruit-surface coverage ratio (FCR) over the number of views under (a) light, (b) moderate, and (c) heavy occlusion. Each occlusion level contains 30 simulated scenes.

2026 S. T. Yau High School Science Award
仅用于2026年成利中学科学奖论文公示

Results show that the primary advantage of GFR-NBV lies in early-stage fruit-surface acquisition. As the first view is a common initial observation, the differences from the second view are due to the selected view planning. Under all occlusion levels, GFR-NBV provides the steepest increase in the first two observations and therefore the mean FCR reaches **70%** reconstruction target by third view. Thus, only two additional observations are required after initialization. Conversely, the random selection requires one additional view under light and moderate occlusion and up to two additional views under heavy occlusion.

Although the initial visible fruit surface decreases as occlusion becomes heavier, GFR-NBV continues to reach the target by the third view. As more observations are acquired, the differences between the methods gradually decrease. By the eighth view, all methods approach a fruit-surface coverage of approximately 88–92%. Therefore, GFR-NBV reaches the target fruit within two observations after initialization, showing early-view efficiency in all three occlusion levels.

3.4 Grasp-Oriented Pose Estimation

Figure 11 shows the results of the cross-sectional analysis of fruits. For the single fruit, as shown in Figure 11(a), the fitted section centers remain close to a common line, supporting the estimate of the longitudinal axis. The corresponding radius profile exhibits a single maximum and a sustained decrease. The maximum identifies the fruit shoulder and the maximum radius, while the decreasing profile of radii indicates the tip side. The resulting directed axis therefore points from the estimated shoulder toward the fruit tip.

Figure 11(b) shows two overlapping fruits whose radius profile no longer follows the smooth and single peak expected for one fruit. Instead, it contains a sharp radius discontinuity followed by a second radius structure. This indicates that the component should not be treated as a single fruit. It can therefore be excluded from grasp-pose estimation.

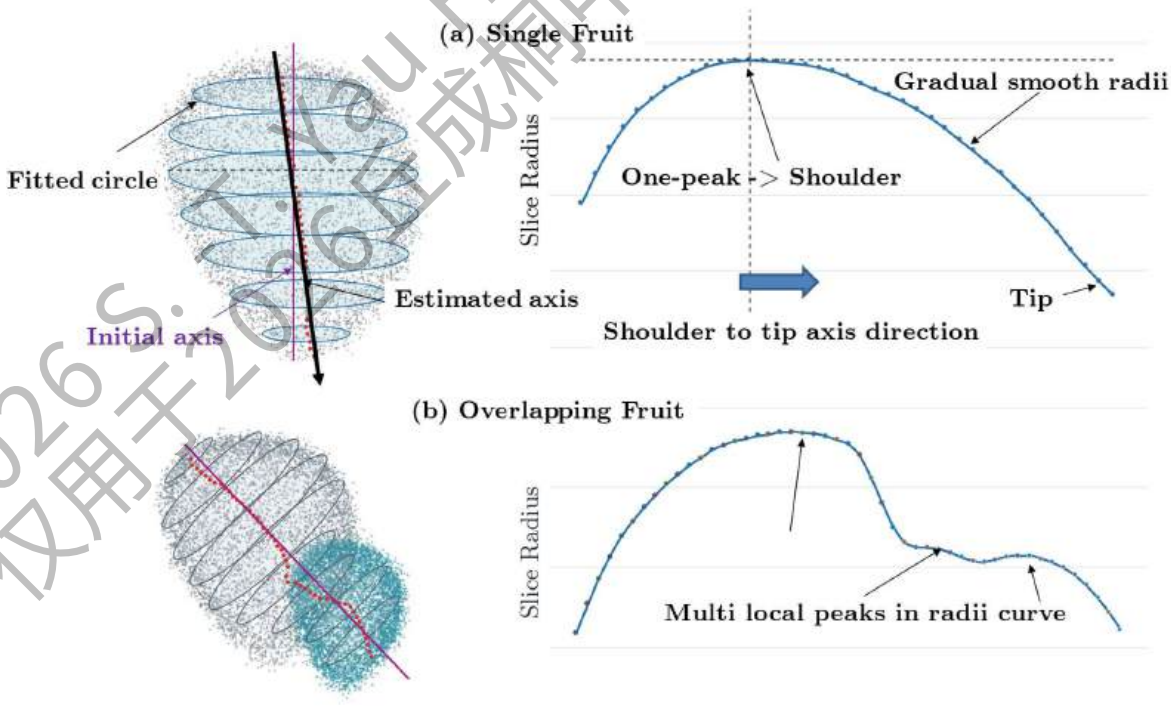


Figure 11 Representative results of cross-sectional shoulder and directed-axis estimation.

2026 S. Liyau High School Science Award
仅用于2026日成利中学科学奖公示

4. Physical Experiments

Although simulation enables validation of GFR-NBV in controlled environments, influences in terms of real RGB-D camera measurements, calibration errors of vision, physical occlusions, control errors and constraints of the robot arm can only be evaluated in physical experiments. Physical experiments are therefore carried out to assess whether the proposed method can acquire sufficient fruit-surface information and estimate shoulder geometry and orientation for robotic grasping.

4.1 Experimental Setup

4.1.1 Robotic Experimental Platform

Figure 12(a) shows the laboratory experimental setup. In the harvesting application, the robotic arm is mounted on a mobile vehicle and approaches strawberries in the lower crop row, as shown in Figure 12(b). In order to carry out the experiment conveniently, we fixed the robotic arm on the table and the plant model was raised, keeping the relative positions of the arm and plant in application. The plant model was placed within the reachable zone of the robotic arm, and different occlusions were adjusted by changing the surrounding leaves.

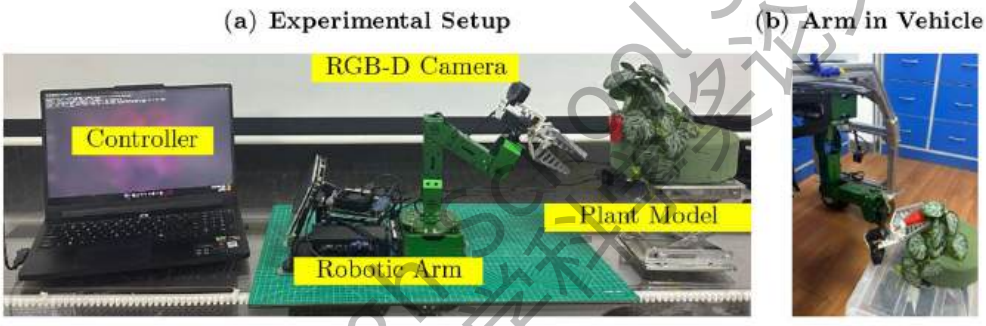


Figure 12 Experimental platform to demonstrate the robotic arm grasping: (a) experimental setup, and (b) demonstration of a robotic arm installed in a vehicle to grasp strawberries.

This platform includes a Hi Wonder Jet-Arm robotic arm, an STM32 embedded unit, an ORBBEC Gemini 335 RGB-D camera, a laptop controller, and a strawberry plant model with artificial strawberry fruits, leaves, and stems. The RGB-D camera was installed at the end of the arm so that the target can be easily observed from multi-views via motion of the end of arm. The algorithm was run in the laptop controller to realize point-cloud reconstruction, GFR-NBV evaluation, shoulder and directed-axis estimation, and robot motion control. The motion and grasping orders were sent to the STM32 unit which directly controlled the robotic arm and gripper.

4.1.2 Fin-Ray Gripper

The grasping gripper was developed for picking strawberry, as shown in Figure 13. Since ripe strawberries are fragile, the gripper uses three compliant Fin-Ray fingers that adapt to local surface variations during closure. The Fin-Ray fingers were designed to conform the curve of strawberry fruits, and were manufactured by 3D printing with TPU95 material. TPU95 provides the elasticity required for compliant strawberry grasping.

2026 S-T·Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

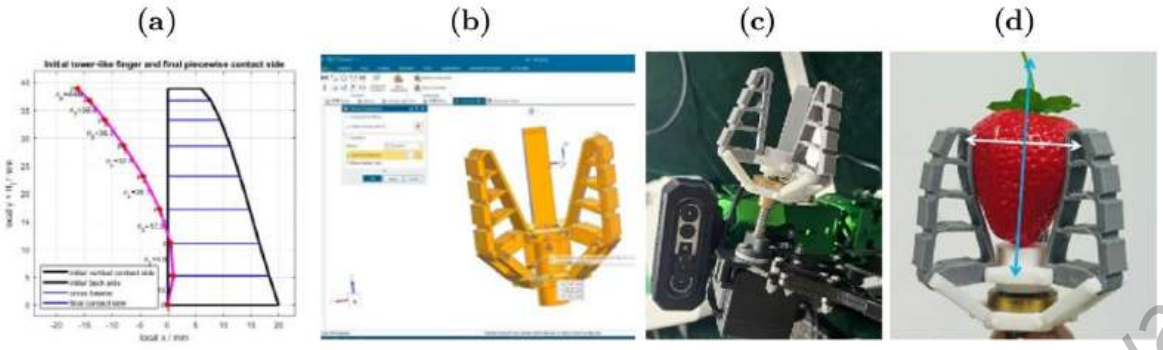


Figure 13 Fin-Ray fingers and strawberry grasping: (a) numerical simulation of a Fin-Ray finger to fit typical strawberry fruit surface, (b) 3D design of a Fin-Ray gripper, (c) installed Fin-Ray gripper, and (d) Fruit grasping pose in the Fin-Ray gripper.

4.2 Experimental Procedures

4.2.1 Experimental Scenes

Three different strawberry fruits were used to construct the physical test scenes. For each fruit, the fruit pose and surrounding leaves were manually adjusted to create different levels of occlusion, as shown in Figure 14.



Figure 14 Initial observations from the camera. Three strawberries combined with different foliage placements to generate light-, moderate-, and heavy-occlusion.

Unlike the simulation, the physical scenes cannot be generated accurately based on the predefined occlusion level. Instead, the fruit and foliage were first arranged in the experimental workspace, and the occlusion level was determined after the initial RGB-D observation. Each scene was qualitatively classified into light, moderate, and heavy occlusion according to the measured initial view of the target fruit. This kind of classification resulted in an unequal number of scenes of occlusion levels. Since the purpose of the physical experiment was to evaluate GFR-NBV under real arrangements, we did not adjust the arrangements for balance. The experiments include combinations of fruit samples, fruit orientation, foliage distribution, and target visibility within the same platform.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

4.2.2 Experimental Procedures

The physical experimental procedure is briefly presented in Figure 15. For each scene, GFR-NBV and unweighted NBV were tested independently from the same initial observation. Other parameters, such as candidate-view set, robotic arm settings, and a maximum of 8 observations, were the same.

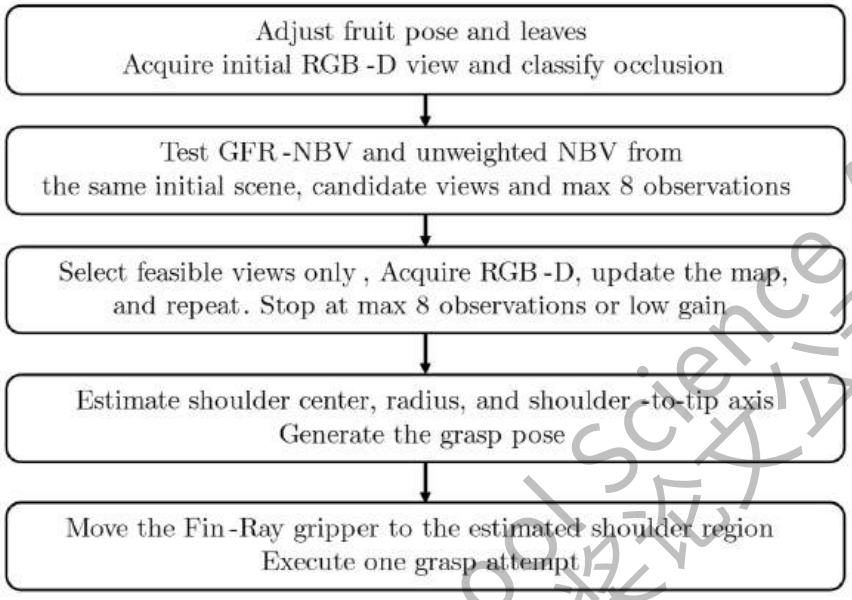


Figure 15 Experimental procedures for GFR-NBV and unweighted NBV under the same initial conditions. Candidate views were restricted to the reachable workspace of the arm.

Unlike the ideal spherical viewpoints used in simulation, physical candidate views were restricted by the arm reachable range and camera range. GFR-NBV used fruit-relevance weights, while the unweighted NBV with all fruit-relevance set to one.

After reconstruction, the fruit-relevant surface was used to estimate the shoulder center, radius, and directed shoulder-to-tip axis, which were converted into a grasp pose. The 7 scenes were tested once with both GFR-NBV and unweighted NBV, and a total 14 physical NBV scenes were carried out.

4.3 Results and Discussion

Table 2 provides the experimental results for seven scenes. The maximum observation number was eight. As we cannot count the real fruit point number, we define N_{ref} as the larger final target-fruit point count obtained by GFR-NBV and unweighted NBV. For post experiment comparison, 70% N_{ref} was used as a common point acquisition reference, which is adopted to compare efficiency and does not represent true fruit-surface coverage.

GFR-NBV reached the 70% reference level earlier than unweighted NBV in five of the total seven scenes and required the same number of observations in the remaining two scenes. The advantage was more evident under the heavy and moderate occlusion. For Fruit 1, GFR-NBV reached 70% N_{ref} in three observations for heavy and moderate occlusions, whereas unweighted NBV did not reach 70% with all eight observations. For Fruit 2 under the moderate occlusion, GFR-NBV reached 70% in two observations, while the unweighted method again failed to reach it within the observation limit. Fruit 3 under heavy occlusion

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

required three and five observations for GFR-NBV and unweighted NBV, respectively.

Under light occlusion, the difference between the two NBV was smaller. Fruit 1 and Fruit 3 reached the target in the same number of observations with both NBVs, while GFR-NBV required one fewer observation for Fruit 2. This is reasonable because a larger portion of the fruit is visible from the initial view, reducing the benefit of relevance-guided view selection.

GFR-NBV also produced more final fruit points in 5 of 7 scenes. However, final point number and the early reconstruction efficiency were not always consistent. For example, under the light occlusion, unweighted NBV acquired more final points for Fruit 1 and Fruit 2, while GFR-NBV reached 70% reference level equally early or earlier. This means that the main benefit of GFR-NBV is not necessarily maximizing the final number of reconstructed points, but acquiring sufficient target-fruit information with fewer observations.

Once the reconstruction reached 70% of N_{ref}, the fruit pose could be successfully estimated, allowing the Fin Ray gripper to approach the fruit with the correct orientation.

Table 2 Experimental results of GFR-NBV and unweighted NBV in 7 scenes with different levels of occlusion. Final Points: accumulated points of the fruit with maximum 8 observations. Steps to 70% N_{ref}: step number to reach 70% of larger final target-fruit point count obtained by GFR-NBV and unweighted NBV.

		Heavy		Moderate		Light	
		GFR NBV	Unweighted NBV	GFR NBV	Unweighted NBV	GFR NBV	Unweighted NBV
Fruit 1	Final Points	1529	914	940	633	3783	4467
	Steps to 70% N_{ref}	3	>8	3	>8	2	2
Fruit 2	Final Points			2511	845	1657	1973
	Steps to 70% N_{ref}			2	>8	2	3
Fruit 3	Final Points	1228	1220			1837	1587
	Steps to 70% N_{ref}	3	5			1	1

The physical experiments provide the similar results as the simulation. The fruit-relevance weighting improves early-view reconstruction efficiency, especially under moderate and heavy occlusions. For all seven scenes, GFR-NBV reached the 70% point-reference level earlier in five scenes and tied in the remaining two scenes than the unweighted NBV, but not always has greater final reconstruction points. Hence, the main benefit of GFR-NBV is to obtain sufficient target-fruit information earlier.

2026 Shi Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

5. Conclusions and Future Work

5.1 Conclusions

This study presents geometry-derived fruit-relevance weighting into occupancy-based next-best-view planning for active strawberry reconstruction. Local curvedness and shape index are calculated and propagated to nearby unknown voxels in virtual ray casting. The candidate views are ranked based on the score from the expected uncertainty reduction and their fruit-relevance weights. Thus, this method requires only geometric reference information, without semantic detectors or learned shape priors.

The reconstructed fruit surface is employed to estimate the grasp-oriented pose. Cross-sectional analysis of fruit points provides the fruit shoulder center and radius. Moreover, the sectional radii variance determines the direct shoulder-to-tip axis. With these parameters, the grasp location, orientation, and initial gripper opening are determined.

Simulation results show that GFR-NBV mainly improves early-view reconstruction efficiency. Under light, moderate, and heavy occlusions, GFR-NBV reached the 70% reconstruction target with the same or fewer observations than the unweighted NBV. Physical experiments were conducted on a robotic platform with a robotic arm, Fin-Ray gripper and different strawberry scenes.

In the seven physical scenes, GFR-NBV reached the 70% point-reference earlier in five scenes and tied in the remaining two scenes. These results validate that geometry-derived fruit-relevance weighting can help acquire sufficient target-fruit information earlier.

5.2 Limitations and Future Work

The effectiveness of GFR-NBV planning depends on the quality of fruit surface detection. Sensing noises and folded leaves may reduce the accuracy of curvature-based fruit surface identification. The current study has not systematically consideration of sensor noise or natural-foliage complexity. In addition, the propagation of fruit relevance to adjacent unknown voxels also introduces uncertainty, particularly when voxels are near object boundaries.

Finally, this study does not evaluate the motion path of the robotic arm and its execution time. Future work should consider motion time cost and extend evaluation to real strawberry plants in field harvesting.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

References

1. Zhou, H., Wang, X., Au, W., Kang, H., & Chen, C. (2022). Intelligent robots for fruit harvesting: recent developments and future challenges. Precision Agriculture, 23, 1856-1907.
2. Cui, T., Dong, Y., Zhang, R., Wu, Y., Zhang, Y., Lu, X., & Tang, Z. (2026). From Field to Market: Evolution of Strawberry-Harvesting Techniques and Research Progress in Intelligent Robotic Systems. *Applied Sciences*, *16*(15), 7522.
3. Magalhães, S.A., Moreira, A.P., Santos, F.N.d. *et al.* Active Perception Fruit Harvesting Robots — A Systematic Review. *J Intell Robot Syst* **105**, 14 (2022).
4. C. Connolly, "The determination of next best views," *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, St. Louis, MO, USA, 1985, pp. 432-435.
5. Pito, R. (1999). "A Solution to the Next Best View Problem for Automated Surface Acquisition." IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10), 1016–1030.
6. Alsadik, B., Mahdi, H. A., & Abdulateef, N. A. (2025). A structured review and taxonomy of next-best-view strategies for 3D reconstruction. ISPRS Open Journal of Photogrammetry and Remote Sensing, 17, 100098.
7. Isler, S., Sabzevari, R., Delmerico, J., & Scaramuzza, D. (2016). An Information Gain Formulation for Active Volumetric 3D Reconstruction. Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA).
8. Zeng, R., Zhao, W., & Liu, Y.-J. (2020). PC-NBV: A Point Cloud Based Deep Network for Efficient Next Best View Planning. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 7050–7057.
9. Ci, J., van Henten, E. J., Wang, X., Burusa, A. K., & Kootstra, G. (2025). SSL-NBV: A self-supervised-learning-based next-best-view algorithm for efficient 3D plant reconstruction by a robot. Computers and Electronics in Agriculture, 233, 110121.
10. Menon, R., Zaenker, T., Dengler, N., & Bennewitz, M. (2023). NBV-SC: Next Best View Planning Based on Shape Completion for Fruit Mapping and Reconstruction. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2023).
11. Li, X., He, W., Park, J., Reberg-Horton, S. C., et al. (2025). SemP-NBV: Semantic-Aware Predictive Next-Best-View for Autonomous Plant 3D Reconstruction. Proceedings of the 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
12. Abouzeid, A., Mansour, M., Hu, C., & Song, D. (2025). A Point Cloud Completion Approach for the Grasping of Partially Occluded Objects and Its Applications in Robotic Strawberry Harvesting. arXiv:2506.14066.
13. Burusa, A. K., Scholten, J., Wang, X., Rapado Rincon, D., van Henten, E. J. (2024). Semantics-aware next-best-view planning for efficient search and detection of task-relevant plant parts. Biosystems Engineering, 248, 1–14.
14. Burusa, A. K., van Henten, E. J., & Kootstra, G. (2024). Attention-driven next-best-view planning for efficient reconstruction of plants and targeted plant parts. Biosystems Engineering, 246, 248–262.
15. Song, X., & Karydis, K. (2025). GS-NBV: a Geometry-based, Semantics-aware Viewpoint Planning Algorithm for Avocado Harvesting under Occlusions. arXiv:2506.23369.
16. Avendano, J., Ramos, P. J., & Prieto, F. A. (2017). A system for classifying vegetative structures on coffee branches based on videos recorded in the field by a mobile device. Expert Systems with Applications, 88, 178–192.
17. Ge, Y., Xiong, Y., & From, P. J. (2022). Three-dimensional location methods for the vision system of strawberry-harvesting robots: development and comparison. Precision Agriculture, 24(2), 764–791.
18. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., & Stuetzle, W. (1992). Surface Reconstruction from Unorganized Points. Proceedings of SIGGRAPH '92, pp. 71–78
19. Koenderink, J. J., & van Doorn, A. J. (1992). Surface shape and curvature scales. Image and Vision Computing, 10(8), 557–564.
20. Vasquez-Gomez, J. I., Sucar, L. E., Murrieta-Cid, R., & Lopez-Damian, E. (2014). Volumetric Next-best-view Planning for 3D Object Reconstruction with Positioning Error. International Journal of Advanced Robotic Systems, 11(10), 159.
21. Crooks, W., Vukasin, G., O'Sullivan, M., Messner, W., & Rogers, C. (2016). Fin Ray® Effect Inspired Soft Robotic Gripper: From the RoboSoft Grand Challenge toward Optimization. Frontiers in Robotics and AI, 3, 70.
22. Prusinkiewicz, P., & Lindenmayer, A. (1990). The Algorithmic Beauty of Plants. New York: Springer-Verlag.
23. Guan, Z., Abd-Elrahman, A., Whitaker, V., Agehara, S., Wilkinson, B., Gastellu-Etchegorry, J.-P., & Dewitt, B. (2022). Radiative Transfer Image Simulation Using L-System Modeled Strawberry Canopies. Remote Sensing, 14(3), 548.

2026 T-Rau High School Science Award 项目-2026日晟杯高中生科学论文大赛