Abstract—Robotic systems in agriculture do not only enable increasing automation of farming activities but also represent new challenges for robotics due to the unstructured environment and the non-rigid structures of crops. Especially, active perception for fruit mapping and harvesting is a difficult task since occlusions frequently occur and image segmentation provides only limited accuracy on the actual shape of the fruits. In this paper, we present a viewpoint planning approach that explicitly uses the shape prediction from collected data to guide the sensor to view as yet unobserved parts of the fruits. We developed a novel pipeline for continuous interaction between prediction and observation to maximize the information gain about sweet pepper fruits. We adapted two different shape prediction approaches, namely parametric superellipsoid fitting and model based non-rigid latent space registration, and integrated them into our Region of Interest (RoI) viewpoint planner. Additionally, we used a new concept of viewpoint dissimilarity to aid the planner to select good viewpoints and for shortening the planning times. Our simulation experiments with a UR5e arm equipped with a Realsense L515 sensor provide a quantitative demonstration of the efficacy of our iterative shape completion based viewpoint planning. In comparative experiments with a state-of-the-art viewpoint planner, we demonstrate improvement not only in the estimation of the fruit sizes, but also in their reconstruction. Finally, we show the viability of our approach for mapping sweet peppers with a real robotic system in a commercial glasshouse.

I. INTRODUCTION

The OECD-FAO Agricultural Outlook 2022-2031 [1] projects a requirement of 24% agricultural productivity growth to meet the twin challenges of zero hunger and reduction in greenhouse gas emissions. However, with a projected decrease of 20% in real food prices by 2031, farming has become an increasingly unattractive economic proposition, especially with the increasing industrialization and urbanization of low income countries. This necessitates increasing automation of agricultural activities, with a shift from large-scale mechanization and indiscriminate chemical interventions to targeted precision agriculture using robotics and artificial intelligence.

Robotics is being used for varied activities in agriculture such as crop monitoring, land preparation, plant treatment, harvesting, plant phenotyping, and yield estimation [2]. In the context of precision agriculture, fruit detection and mapping is key to yield estimation and harvesting. Unlike objects found in industrial and household scenarios, fruits and plants change their color, size and shape over time, requiring spatio-temporal mapping of their positions and sizes for phenotyping and crop management decisions. However, frequent occlusion of fruits and the variation in their position on plant structures make the reliable perception of fruits a challenging task. State-of-the-art active perception for horticulture robots focuses on image-based detection and segmentation as well as visual servoing techniques presenting a research gap in shape-based active perception [3]. Whereas we earlier investigated the benefits of shape completion for improving the accuracy of fruit size estimation using data fused from different poses generated by our RoI viewpoint planner [4], [5], we now close the loop by feeding the shape completion results to inform the viewpoint planner where to plan viewpoints for maximizing the information gain.

In more detail, we present a framework for shape completion based viewpoint planning that uses the predicted shape structure to focus the sensor’s attention on the intersection between unknown regions and missing surfaces. As can be seen in Fig. 1, shape completion enables not only to predict
the size and position of partially detected fruits, but also to guide the sensor to the next best view using the predicted shapes.

We utilize two different approaches to shape completion [4], [6] and perform a comparative analysis of their efficacy in guiding the viewpoint planning.

To summarize, our contributions are the following:

- Adaption and integration of two shape completion methods to predict fruit shapes.
- A novel viewpoint planning approach that uses the predicted surfaces of the fruit shapes for finding the next best view.
- Formulation of viewpoint dissimilarity to find new viewpoints in far away regions.
- Quantitative simulation experiments that demonstrate superior performance of our planner compared to viewpoint planning without shape completion in terms of estimated volume and reconstruction accuracy
- Qualitative experiments with a real robotic platform measuring sweet pepper plants in a commercial glasshouse.

II. RELATED WORK

With robots being increasingly deployed outside structured industrial scenarios, active perception is a key factor in improving their efficacy [7]. Viewpoint planning is a subset of active perception where the sensor pose or a sequence of poses is planned to maximize the information gain, i.e., minimize the entropy about the state of the environment or target objects, subject to constraints such as obstacle avoidance and movement cost. For detailed oriented tasks, especially at the object level, such as active recognition, pose estimation [8], mapping or reconstruction [9], [10], manipulators or mobile manipulators are typically used with attention-driven next best view (NBV) planning.

In reconstruction tasks, it is typically assumed that the object is not occluded by the environment and, given enough views, can be completely perceived by the viewpoint planning system. However, fruit mapping is a task where the objects of interest are highly occluded due to fruits growing under the leaves and hence might never be fully reconstructed. Soria et al. [11] developed a multi-view reconstruction method for apples in commercial orchards using probabilistic segmentation, whereas Sarabu et al. [12] proposed a dual arm system for cooperative apple picking with both arms equipped with RGB-D sensors for volumetric surveying and grasping, using graph-based planning. Zaenker et al. [5] developed RoI targeted viewpoint planning where contours of the detected RoIs are selected as targets for the next best view. While this approach shows an improvement in volume estimation compared to general viewpoint planners, it does not utilize shape information to find the next view. Burusa et al. [13] demonstrated that placing 3D bounding boxes on different parts of the plant such as stem, leaf nodes, and the whole plant, as an attention mechanism to guide the volumetric NBV planner led to significant improvement in accuracy and speed of reconstruction. However, the 3D bounding boxes were defined by the user and not autonomously generated.

Unlike in industrial scenarios, where object congruence in terms of shape is a reasonable assumption, in agricultural and household scenarios, even objects belonging to the same class are only similar in shape leading to the need for deformable shape completion [14]. In the field of manipulation planning, shape completion is being increasingly used to improve the robustness of grasp planning [15]–[17], whereas in the agricultural context, shape completion has been used for fruit mapping and localization [18]–[20], without using its output for viewpoint planning. Volumetric occupancy mapping with probabilistic depth completion [21] and prediction of unobserved space using depth map augmentation [22] have been used for navigation planning of mobile robots and micro-aerial vehicles, respectively, without any exploration planning. In the context of exploration planning, shape priors have been used for guiding the exploration of objects using active tactile sensing [23], [24]. Recently, Schmid et al. [25] have demonstrated the application of incremental semantic scene completion for informative path planning of micro-aerial vehicles for efficient environment coverage. While this work is similar to our work, Schmid et al. focus on efficient coverage of complete scenes and not on the precise mapping of particular objects. Furthermore, in contrast to Schmid et al. who update the actual scene mapping based on the scene completion, we mark shape completed regions as unknown in terms of occupancy but with high region of interest probability scores, thus leading to a more conservative application of shape predictions while still using it for guiding the NBV planner to regions of interest.

To the best of our knowledge, next best view planning for object mapping and reconstruction based on iterative deformable shape completion has not been carried out till date.

III. OUR APPROACH

Our work extends the RoI based viewpoint planner [5] with the output of shape completion not only to estimate the location and size of the fruits, but also to provide feedback to the viewpoint planner to provide useful viewpoints on predicted regions of interest. Fig. 2 gives an overview of our shape completion based RoI viewpoint planning. The RoI detection module detects the fruit point clouds which are then fed to the shape completion module as well as the RoI occupancy mapping module. The shape completion module estimates the completed shapes for the fruit clouds, which are then fed to the occupancy mapping module to estimate the predicted regions of interest, with their predicted RoI values. Finally, viewpoints are sampled around the predicted RoIs to obtain better views of the fruits in which turn are used to generate an improved shape prediction. The individual steps are described in more detail in the following subsections.

A. Region of Interest Detection

The color image from the RGB-D camera at every sensor pose is forwarded to an HSV based segmentation method
in the case of red sweet peppers in the simulated scenario, and to a Mask R-CNN [26] based sweet pepper detector, in the real-world experiments for detecting red, green, and yellow peppers, to form the RoI, i.e. fruit masks. These masks are fused with the depth data from the RGB-D camera to form the RoI point cloud which is then forwarded to the occupancy mapping module for NBV planning, and to the surface mapping module for shape completion.

B. Shape Completion

The shape completion process runs in parallel to viewpoint planning, with the latest shape completion result being used for planning new viewpoints. It is also used for estimating the size and location of the fruits. It takes as input the fruit point clouds and outputs the completed shapes.

1) Surface Mapping and Clustering

The RoI point clouds of the fruits at every observation pose are fed to the voxblox mapping system [27], which accumulates the point clouds iteratively to form a truncated signed distance fields (TSDF) map, from which the surface point cloud of the fruits are extracted. The surface point cloud is then clustered using the method described in [4]. We also estimate the centroid of each cluster by computing the surface normals and calculating the least-squares solution to the intersection. We improved the clustering by performing a sanity check on the cluster size, i.e., clusters whose bounding box dimensions are larger than the typical sweet pepper dimensions are split further.

2) Shape Fitting

We integrated the following shape completion methods to provide a common shape completer interface which provides feedback to the viewpoint planner.

Superellipsoid Fitting (SE): As in our previous work [4], the clustered surface point clouds are fed to the superellipsoid matcher, which fits a superellipsoid to the respective cluster by optimizing a cost function that minimizes the deviation in cluster center while simultaneously imposing constraints on its dimensions.

Non Rigid Shape Registration (SR): Rodriguez et al. [6] developed a category level shape completion approach using a learned latent space projection, and local non-rigid registration based on coherent point drift [28], for predicting shapes of partially observed objects for grasp planning. We adapted the approach to predict shapes of fruits by learning a canonical model using meshes of the sweet peppers used in simulation, as shown in Fig. 3. As the shape registration approach can only deal with local rigid transforms, we first shift the input cluster to the centroid calculated in Sec. III-B.1. Then, the shape registration calculates the deformed completed shape for each cluster along with a small local rigid transform. The deformed shape is then shifted back to its original centroid location and corrected using the generated transform.

C. Occupancy Mapping with Predicted Regions of Interest

The RoI point clouds generated in Sec. III-A as well as the complete point cloud at every observation pose are fed to the RoI Octomap developed in our previous work [5], by augmenting the Octomap [29] nodes with region of interest values in addition to the existing occupancy information. We then calculate the missing surfaces from the completed shapes and their predicted RoI value, to augment the RoI Octomap with information about predicted RoIs.

1) Missing Surface Estimator

The shape completion approaches detailed above fit a complete shape based on the clustered point cloud. However, for viewpoint planning we are only interested in the missing surfaces. Thus, we perform a nearest neighbour search for the predicted cloud on the input cloud and store the Eu-
yield estimation and for robotic harvesting. To enable better estimation of their location and volume for information gain about the objects of interest, i.e., fruits, we use viewpoint planning to maximize the viewpoint information gain. Viewpoint planning is used here to maximize the utility value of the viewpoint for new fruit clusters being detected. The occupancy map does not belong to the deleted list before the insertion. This occurs when the occupancy value of a node is lower than a threshold (chosen according to Eq. (2), similar to Duberg et al. [30]:

\[
\text{state}_{\text{occ}} = \begin{cases} 
\text{Free}(F) & p_{\text{occ}} \leq 0.45 \\
\text{Occupied}(O) & p_{\text{occ}} \geq 0.55 \\
\text{Unknown}(U) & \text{otherwise}
\end{cases}
\]  

Similarly, we also modified the RoI states from a binary RoI-NonRoI state to account for predicted regions of interest PredRoI in Eq. (3), which are nodes whose occupancy state is unknown and whose RoI value is lower than that for regions of interest:

\[
\text{state}_{\text{roi}} = \begin{cases} 
\text{RoI} & \text{state}_{\text{occ}} = O \land (p_{\text{roi}} \geq 0.75) \\
\text{PredRoI} & \text{state}_{\text{occ}} = U \land (0.5 < p_{\text{roi}} < 0.75) \\
\text{NonRoI} & \text{otherwise}
\end{cases}
\]  

As Octomap does not explicitly store unknown regions as nodes, while inserting predicted RoI nodes, we check the Octomap for the predicted cloud point, and only if it does not exist, we insert a new node with \( p_{\text{occ}} = 0.5 \) and the respective \( p_{\text{roi}} \). With every observation, we check for the state of the predicted region nodes. The predicted RoI nodes that are observed in the current view and change their occupancy state to Free or Occupied are removed and added to the list of deleted predicted nodes. In the next round of insertion of predicted RoI nodes, it is verified that they do not belong to the deleted list before the insertion. This ensures the predicted RoIs decrease over time in the absence of new fruit clusters being detected. The occupancy map with predicted RoIs and their predicted RoI values, is used for viewpoint generation as described below.

### D. Predicted RoI Based Viewpoint Planning

The viewpoint planner has no model of the environment and hence we use a random sampling based next best view planner. Viewpoint planning is used here to maximize the information gain about the objects of interest, i.e., fruits, to enable better estimation of their location and volume for yield estimation and for robotic harvesting.

#### 1) Viewpoint Sampling

As in our previous work [5], we use a combination of two sampling methods: predicted RoI targeted sampling, which uses the predicted regions of interest to find viewpoints that can observe hitherto unseen predicted parts of the fruits, and exploration sampling to explore unknown regions to find new regions of interest, i.e., fruits if the utility values of all predicted RoI viewpoints fall below a threshold. We sample candidate viewpoint poses from the target nodes by sampling random sensor distances and viewing directions. We then cast rays from the viewpoint pose according to the sensor’s field of view parameters, until they hit an occupied node or the target node. Additionally, we formulated two new methods for sampling around predicted regions of interest. In the first method, we randomly sample nodes from the list of predicted RoI nodes (pRoI). In the second method, we calculate the RoI weighted mean of the missing surfaces (MSC). These missing surface centers are used as target nodes for viewpoint pose generation. For predicted RoI sampling, nodes with \( \text{state}_{\text{occ}} = \text{Unknown} \) in the ray’s path are summed to calculate the \( \text{entropyGain} \) whereas the RoI value of nodes with \( \text{state}_{\text{roi}} = \text{predRoI} \) in the 6-neighbourhood \( N_{6} \) of the unknown node are summed to calculate the \( \text{roiGain} \). The expected information gain \( IG \) of a ray is calculated as the weighted gain of \( \text{entropyGain} \) and \( \text{roiGain} \) in Eq. (4):

\[
IG = \alpha * \text{roiGain} + (1 - \alpha) * \text{entropyGain}
\]  

#### 2) Viewpoint Dissimilarity

We observed that the occlusion of fruits by leaves and the limited reachability of manipulators led to the problem of repeated sampling of similar viewpoints which was exacerbated by inserting predicted RoIs. To mitigate this problem,
we formulated the concept of viewpoint dissimilarity and filter sampled viewpoints using a dissimilarity threshold as shown in Alg. 1. The dissimilarity metrics between two viewpoints \( v_{p1} \) and \( v_{p2} \), with origins \( t_{v_{p1}} \) and \( t_{v_{p2}} \), and viewing directions \( dir_{v_{p1}} \) and \( dir_{v_{p2}} \) respectively are formulated as follows:

\[
VpD_{\perp}(v_{p1}, v_{p2}) = 1 - \left( \frac{1}{VpD_{\perp}} \right) \\
VpD_{\text{origin}}(v_{p1}, v_{p2}) = \min\left( \frac{\|t_{v_{p1}} - t_{v_{p2}}\|}{\text{dist}_{\text{cutoff}}}, 2.0 \right) \\
VpD(v_{p1}, v_{p2}) = \min(VpD_{\perp} \cdot VpD_{\text{origin}}, 1.0) \tag{5}
\]

\( VpD_{\perp} \) indicates the dissimilarity in viewing direction whereas \( VpD_{\text{origin}} \) indicates the dissimilarity in the origin of viewpoints, with \( \text{dist}_{\text{cutoff}} \) being a scaling factor. The dissimilarity index \( VpD \) is in the range \([0,1]\), with 0 indicating a high similarity and 1 indicating a high dissimilarity. Every successfully attained viewpoint, whether from predicted RoI sampling or exploration sampling, is added to a list of past viewpoints as depicted in line 14 of Alg. 1. During the sampling of new viewpoints, each viewpoint is compared to the past viewpoints. If the dissimilarity index falls below a threshold (chosen as 0.1), the viewpoint is discarded, otherwise the information gain \( IG \) is weighted with the dissimilarity index. The dissimilarity indices can be varied to achieve a balance between focusing on currently discovered regions of interest and discovering new regions of interest. The dissimilarity index based viewpoint rejection also leads to a narrower sampling space over time, thus allowing to terminate the viewpoint sampling if no new useful dissimilar viewpoints are available for a certain consecutive number of iterations (line 6 of Alg. 1).

IV. EXPERIMENTS

We performed experiments in simulation using the Gazebo framework [31] to provide a quantitative analysis of the performance of our shape completion based viewpoint planning in comparison to [5].

For the evaluation, we used the three scenarios shown in Fig. 4. For shape completion based viewpoint planning (NBV-SC), we apply two shape completion methods: superellipsoid (SE) and shape registration (SR), and two predicted RoI viewpoint samplers: predicted RoI node sampling (pRoI) and missing surface center sampler (MSC). We compared the performance of our shape completion based viewpoint planners to the RoI viewpoint planner (RVP) [5], with superellipsoid based shape completion only for size and position estimation as in [4]. We carried out 10 trials in each scenario and use a plan length of 120 s for each trial, where plan length corresponds to the trajectory duration.

In Scenario 1 (Fig. 5a), the number of detected fruits are similar for viewpoint planning with and without shape completion. In Scenario 2 (Fig. 5b), and Scenario 3 (Fig. 5c), RVP detects clusters faster compared to NBV-SC planners. However, as it does not have shape completion to find new targets for RoI targeted sampling, it keeps performing exploration sampling with time, which leads to fruits being re-detected from larger distances, causing errors in perception. In contrast to RVP, the NBV-SC planners detect new clusters over time, with the number of clusters detected slightly lower as compared to RVP. This is due to the fact it keeps finding new predicted RoI viewpoints, thus leading to less exploration.

Within the NBV-SC planners, it can be seen that shape registration has less variation both in the number of fruits detected and the volume accuracy. SE is more deformable as it tries to fit a general superellipsoid shape to fruit clusters whereas SR being model based, has less variation in the completed shapes. SR tends to underestimate the completed shape volume whereas SE generally overestimates the shapes. With regards to the sampling approach, while MSC is computationally faster as it uses the cluster centers instead of individual predicted RoI nodes, its performance in terms of volume accuracy, especially in Scenario 3, is worse than that of pRoI sampling for superellipsoid fitting.

We also performed an ablation study by disabling the viewpoint dissimilarity check for SE-MSC method and adding the viewpoint dissimilarity check to RVP, and conducted 10 trials for Scenario 3. It can be seen from Tab. I that addition of viewpoint dissimilarity check leads to either termination of the sampling process or encourages the planner to find useful dissimilar viewpoints, resulting in an overall lower planning time.

We additionally calculated the Chamfer distance of the generated fruit point cloud from the surface map to the ground truth to evaluate the final quality of our reconstructed fruit meshes. Tab. II shows that with the new approaches, the average distances are 2-3 mm more accurate compared to our old planner. While this is only a small improvement, it applies consistently across all scenarios. We

| Planner              | Total Time (seconds) |
|----------------------|----------------------|
| SE-MSC               | 377.2 ± 80.7         |
| SE-MSC-NOVPD         | 715.0 ± 101.0        |
| RVP-VPD              | 411.0 ± 81.7         |
| RVP                  | 993.8 ± 236.7        |

TABLE I: Effect of Viewpoint Dissimilarity in Scenario 3 of Fig. 4. SE-MSC-NOVPD is viewpoint dissimilarity check disabled for SE-MSC whereas RVP-VPD is RVP augmented with viewpoint dissimilarity check. Total time is the time taken for the planner to achieve a plan length of 120 seconds, or the time at termination of sampling.
Fig. 5: Volume accuracy of completed shapes for fruit clusters. SE denotes superellipsoid, SR denotes shape registration, MSC denotes missing surface center sampler, pRoI denotes predicted RoI node sampling, and RVP denotes the RoI Viewpoint planner [5]. NBV-SC planners and RVP are similar in number of detected clusters, however NBV-SC planners outperform RVP in the accuracy of volume estimation of completed shapes, especially in scenarios 2 and 3. NBV-SC planners plan more views on individual fruits leading to better individual coverage. Volume accuracy is the accuracy of the completed shape’s volume compared to the groundtruth shape’s volume calculated as in [4].

| Planner   | Scenario 1 | Scenario 2 | Scenario 3 |
|-----------|------------|------------|------------|
| SE-pRoI   | 8.2 ± 0.2  | 7.2 ± 0.2  | 7.2 ± 0.2  |
| SE-MSC    | 8.2 ± 0.3  | 7.5 ± 0.3  | 8.2 ± 1.0  |
| SR-pRoI   | 8.3 ± 0.2  | 8.0 ± 1.0  | 7.4 ± 0.2  |
| SR-MSC    | 8.0 ± 0.8  | 7.2 ± 0.3  | 7.6 ± 0.4  |
| RVP       | 10.1 ± 2.0 | 10.5 ± 2.0 | 10.9 ± 2.0 |

TABLE II: Chamfer distances (in mm) of the detected point clouds generated from the surface map [27], obtained after a plan length of 120 s or the time at termination of sampling, compared to the groundtruth. All NBV-SC methods achieved significantly better results compared to RVP [5].

We performed a one-sided MannWhitney U test of RVP with all other methods, to check the statistical significance of the Chamfer distance improvement. In all scenarios, NBV-SC methods were significantly better with $p < 0.05$. The new methods generate more viewpoints per fruit, which leads to this improvement. In real-world scenarios, with inaccurate sensors and potentially more cluttered environments with more occlusions, this behavior could prove to be even more beneficial.

For our real-world experiments, we applied the approach for mapping sweet peppers in a commercial glasshouse, with the UR5e arm equipped with a Realsense L515 sensor, mounted on the trolley presented in [32]. An example of the performed shape completion in this environment is shown in Fig. 1. We provide a demonstration of the viewpoint selection in the accompanying video.

V. SUMMARY

In this paper, we presented a novel approach to active perception in agricultural robotics using iterative deformable shape completion based viewpoint planning for fruit mapping. We adapted two shape completion approaches, namely superellipsoid fitting and non-rigid shape registration for predicting regions of interest that can be used as potential targets for next best view planning. Our simulation experiments with a UR5e arm equipped with a Realsense L515 show that shape completion based viewpoint planning leads to more complete coverage and thereby better reconstruction of individual fruits as well as to a more accurate size estimation. We also formulated a new concept of viewpoint dissimilarity to strike a balance between individual versus total coverage of fruits using manipulator based exploration and demonstrated its effects in reducing the planning time.
