3D Reconstruction Of Strawberries In Complex Outdoor Environment To Assist Evaluation And Faster Automated Picking Task

Student Work by Shashank Bhoite

Abstract : This research proposes a novel method to create a robust 3D reconstruction of strawberry plants in a complex environment using readily available depth cameras by overcoming the limitation of distance for depth sensing using synthetic data prediction using GANs. This study will cover most aspects of robotic strawberry picking from fruit picking to state-of-the-art reconstruction methods highlighting weaknesses and propose a new technique to resolve them. This project not only focuses on developing novel method but also making it faster than state of the art available method which will efficiently fit in picking pipeline. The proposed architecture was designed with modularity in mind so that in the future, each part of the pipeline can be improved and reinserted into the system without major changes.

Why is there a need for automated strawberry picking using robot?

Time Consuming process:
- Strawberry picking is highly selective process.

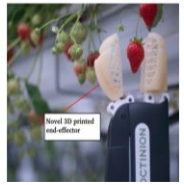
Loss of interest in local workers:
- Over the last decade, there is significant amount of loss interest in fruit by local workers.

Brexit:
Brexit resulted in reduction of amount of seasonal immigrant workers.

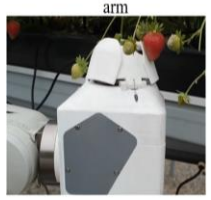
Problems with current state of the art robots

Uneven berry identification

Disease transfer



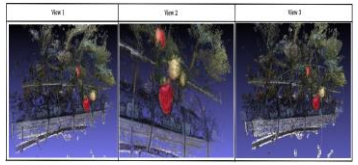
Damage to the cluster from picking arm



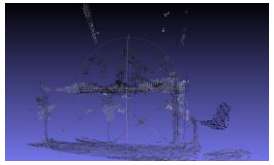
Sensor distance limitation

Current State of art methods evaluated for scope of strawberry picking task :

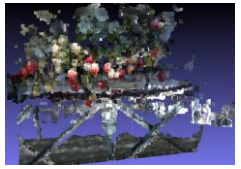
Structure from motion



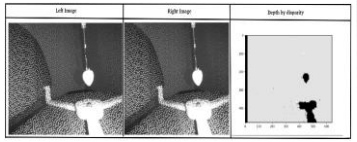
RGBD SLAM ZEDFU

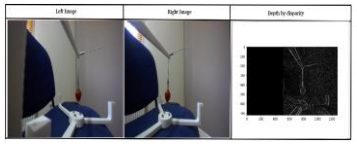


RTABmap

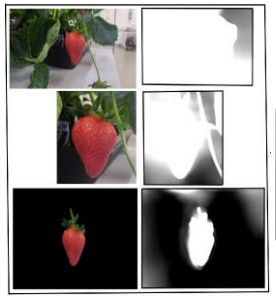


Depth by Disparity

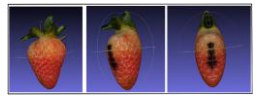




Midas

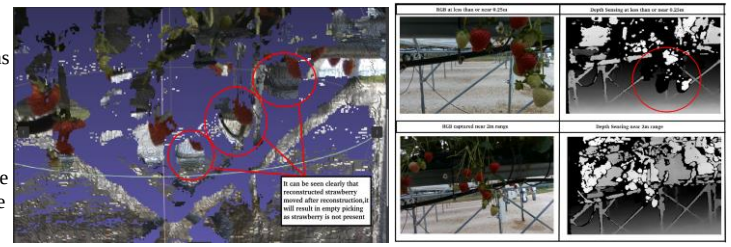


Soft-body reconstruction



Sensor's limitations

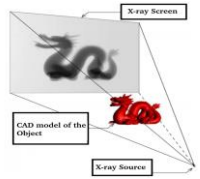
Current state of the art robots mostly use RGBD sensors and stereo cameras to percept objects and for efficient picking.
These sensors have limitation on distance range on they can perceive depth example: RealSense can only sense between 0.25m – 3m



Evaluation

Method Name	Reconstruction of canopy	Reconstruction of plant	Reconstruction of strawberry including picking point at stalk	Affected by environmental factors such and wind and reflection on strawberry	Time Taken	Any technical challenges or shortcomings	Chances of feasible application for strawberry picking and reason
Structure from Motion	Very high Quality	Very high Quality	Very high Quality	Yes, Movements due to wind creates very bad results for SFM	20-25 mins	No	Very low chances due to very high processing time and highly affected by environmental factors
Open3D Reconstruction (RealSense)	Good Quality but noisy	Good Quality but noisy	Good Quality but noisy	Yes, Movements due to wind creates very bad results for Open3D reconstruction	15-20 mins	No	Very low chances due to very high processing time and highly affected by environmental factors along with depth sensing limitation
RTABMAP (ZEDFU)	Very high Quality	No reconstruction	No reconstruction	Not applicable as it fails to captures fruit itself	4-5 mins	No	Very low chances as it fails to reconstruction the plant it self due to depth sensing distance limitation
RTABMAP (RealSense)	Good Quality	Poor Quality	Poor Quality	Yes, Movements due to wind creates very bad results for RTABMAP as well	5 mins	Yes, frequent crashes while reconstruction and lag while capturing data	Very low chances due to poor quality representation
Depth by disparity map (RealSense)	Poor Quality	Poor Quality	Poor Quality	No	5-10 ms (near real-time)	As the RealSense Sensor is not true stereo camera for RGB, it only outputs stereo image for infrared	Very low chances due to poor quality representation
Photogrammetry	Not tested for canopy	Poor Quality	Poor Quality	No	5-10 ms (near real-time)	No	Very low chances due to poor quality representation
Soft Body Reconstruction	Not tested for canopy	Poor Quality	Poor Quality	No	5-10 ms (near real-time)	No	Very low chances due to poor quality representation
MIDAS	Poor Quality	Poor Quality	Poor Quality	No	20-30 ms(near real-time)	No	Very low chances due to poor quality representation
Proposed method	Not tested for canopy	Good Quality but noise	Very high Quality	No	20-30 ms(near real-time)	No	High Chances of feasibility as it is near real-time and provides good quality representation for the strawberry and stalk

Synthetic X-ray



MIDAS

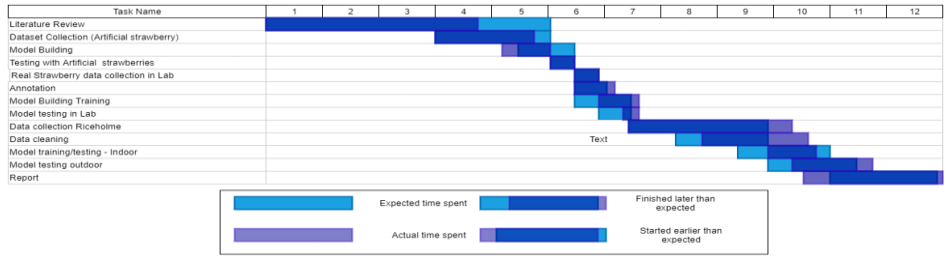
Conclusion for comparison of SOTA methods

As seen in the above results and evaluation, Mentioned algorithms possess capabilities to do a high-quality reconstruction, it is the current generation of sensors that is failing to provide high quality sensing for smaller fruit due to distance limitation and environmental factors such as wind.

Inspiration

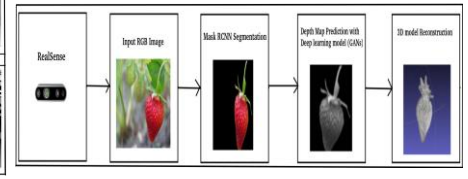
In such cases deep learning comes into picture for filling the gap to predict missing data with synthetic data. Powerful deep learning architecture such GANs(Generative adversarial network) have ability to predict synthetic with very high accuracy.
The proposed architecture uses GANs to predict synthetic depth map inspired from methods MIDAS and prediction of synthetic X-ray using GANs(thesis) by Haidarbhai Mustafa.

Gantt Chart

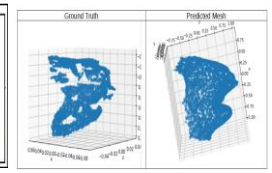


Proposed Architecture

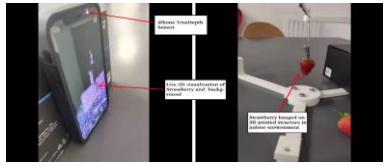
3D Reconstruction Pipeline



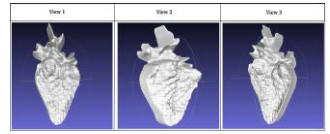
Bad Validation Data

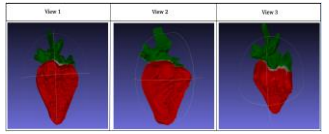


Validation Data Collection

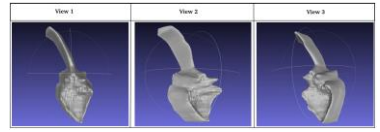


Results





Failed Output



Conclusion

The proposed architecture uses state-of-an-art GANs technique to create a synthetic depth map, which gives a very high-quality representation of the strawberry including the surface of the strawberry and picking point at the stalk. As this method is near real-time the proposed architecture also states the possibility of pair 2D detection and tracking of strawberry and then translating those movements to strawberry representation in 3D to track the strawberry in 3D. This will allow the strawberry picking robot to pick the strawberry unaffected by the real-time movement of the strawberry and without having to go back to reconstruct the scene again.