Fruity: A Multi-modal Dataset for Fruit Classification and 6D Pose Estimation in Precision Agriculture

Mahmoud Abdulsalam ¶, Zakaria Chekakta, Nabil Aouf, and Maxwell Hogan

¶Corresponding author

School of Science and Technology, Department of Electrical and Electronic Engineering
City, University of London, ECV1 0HB London, United Kingdom
Email:{mahmoud.abdulsalam, zakaria.chekakta, nabil.aouf, maxwell.hogan}@city.ac.uk

***Abstract*—The application of robotic platforms for precision agriculture is gaining traction in modern research. However, the demand for a complete fruit dataset is still not satisfied. In this paper, we present fruity, a multi-modal fruit dataset with a variety of use cases such as 6D-pose estimation, fruit detection, fruit picking applications, etc. To the best of our knowledge, this dataset is the first-ever multi-modal fruit dataset tailored specifically for fruit 6D pose estimation in precision agriculture. The dataset is collected over a range of multiple sensors consisting of an RGB-D camera, thermal camera and an indoor tracking camera for ground truth poses. Fruity features RGB images, stereo depth images, thermal images, camera 6D-poses, fruit 6D-poses and relative 6D-poses between the cameras and fruits. The classes of the dataset are commonly harvested fruits which include: apples, oranges, bananas, avocados and lemons. It is also enriched with a clustered class to account for occlusion scenario. The dataset is recorded over multiple trajectories implemented with multiple platforms. The dataset alongside the documentation and utility tools is publicly available at: https://github.com/MahmoudYidi/Fruity.git.**

I. Introduction

Precision agriculture is poised to be the solution of global food shortage. Robotics in agriculture is often considered to be a good form of precision agriculture. However, the shortage of accurate and complete datasets is restricting the exploitation of robots in agriculture. Detecting and estimating 6D-poses find application in object grasping, Virtual Reality (VR), Augmented Reality (AR), and autonomous driving. However, the availability of datasets has limited this application in Agriculture. Having a single sensor for this purpose introduces limitations that are associated with the sensing mechanism. For example, RGB cameras are not usable for complex computer vision applications in low illumination scenarios [1], [2]. RGB Cameras are utilized for edge detection, colour-based classifications [3] and 2D localization [4]. However, using these images for 3D localization is a challenging task [5], [6]. Understanding this compromise, researchers complement the weakness of one sensor with the strength of another thereby arriving at the concept of multi-modal sensoring. Even though datasets are collected peculiar to a given application, it is pertinent that they are collected with completeness, accuracy, and richness to facilitate the development and evaluation of newer innovations.

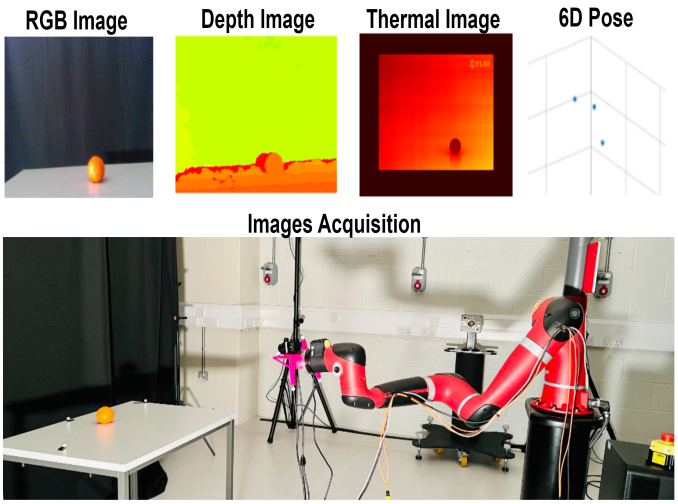

Fig. 1. Figure showing the modalities of our dataset (RGB, Depth and Thermal) including the 6D Pose and the robotic manipulator used for acquiring the dataset.

Over the years, many datasets have been collected for autonomous driving [5], pedestrian detection [7] and odometry [1]. However very few have been collected for fruit detection and picking. In this paper, we propose Fruity: A multi-modal dataset for fruit recognition and 6D-Pose Estimation in precision agriculture. This dataset can be utilized to facilitate and evaluate new innovative methods in fruit picking, detection, and 3D localization. Fruity consists of 6 classes namely: apple, banana, orange, avocado, lemon, and a fruit cluster class. The modalities of the dataset include a thermal modality, RGB modality and a depth modality as seen in Fig. 1. Each class of the dataset is accompanied by 6D poses of the fruits and the camera alongside the relative poses between them. The dataset is acquired through different trajectories on multiple platforms. A customised sensor rig was designed and constructed to house the sensors while

being mounted on the platforms. The data acquisition and synchronization is possible through the Robotics Operating System (ROS) framework [8] as shown in Fig. 2. The outputs of the system are the RGB, depth, and thermal images. The 6D poses of the cameras and the targets are also obtained from the system. These poses are also used to compute the relative poses between the camera and the target. The major contributions of this dataset can be summarised as follows:

- We present a multi-modal indoor fruit dataset that encompasses data from modern sensors. This is the first multi-modal fruit dataset that is tailored for 6D-pose estimation in precision agriculture which finds application in autonomous fruit-picking and harvesting.
- We have acquired the dataset over different trajectories implemented on multiple platforms (manipulator and UAV) to provide a variety of 6D poses to facilitate effective training.
- We provide a toolkit to easily manage and utilize the multi-modal dataset in form of plug-and-play codes as well as providing documentation on how to easily use this data.

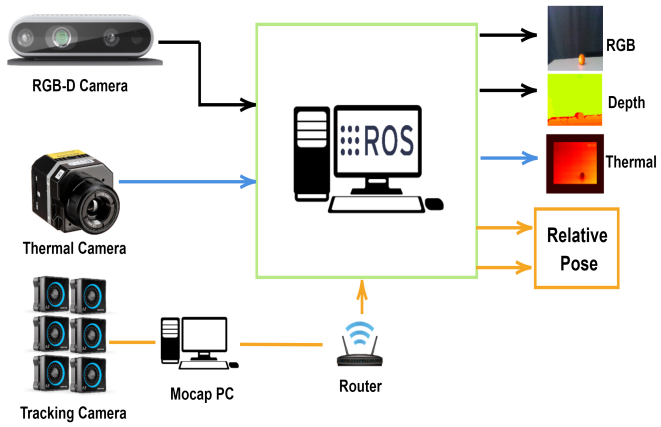

Fig. 2. Figure showing the overview of the dataset acquisition pipeline featuring RGB-Camera, Thermal Camera and Tracking system to record data simultaneously via ROS

II. Related Work

We will discuss works related to the following:

A. Fruit Datasets

The most used fruit dataset for classification problems is the Fruits 360 [9]. The dataset consists of 90,483 RGB images of different varieties of fruits. A Logitech C920 camera was used to capture the fruits on a white background. Even though the dataset is rich in terms of the fruit variety, the dataset fails to capture multiple modalities thereby restricting the dataset to only classification and detection problems. Hence, 6D-pose estimation is not possible using this dataset. Additionally, images are of size 100×100 thereby making it difficult to utilize the data for very detailed classification problems. Additional modality becomes a necessity since the RGB camera can not provide the optimal dataset for all the listed applications. More works in this regard were carried out on RGB-D setups [10]–[13] while others exploited the combination of RGB with Near Infrared (NIR) images [14], [15]. Even though these works have added an additional modality, the problem of 3D fruit localization of the detected fruit still persists since the 6D poses of the camera and fruits are not readily provided as ground truths. Moreover, they focus on a single fruit as a class hence reducing the diversity of these data.

B. 6D Pose Datasets

The LineMOD Dataset [16] is a well-known dataset for 6D pose estimation. It consists of RGB-D images and poses of 15 objects. The OCCLUSION dataset [17] has properties like the former but also accounts for testing 6D poses of occluded objects. Another dataset with similar property is the T-LESS Dataset [18]. This dataset is accompanied with 3D card models. YCB-Video Dataset [19], a popular dataset for 6D pose estimation consists of household objects displayed in short videos. Other 6D pose estimation datasets are also proposed [20]–[22]. Although these datasets are tailored for 6D-pose estimation, none of them is targeted towards 6D-pose estimation of fruits to facilitate fruit picking application in precision agriculture.

C. Other Multi-Modal Datasets

The KITTI dataset [23] is a popular multi-modal dataset having data from lidar, stereo and IMU sensors. Due to the richness of this dataset, it has facilitated the emergence of state-of-art methods for 3D-object detection. The H3D dataset [24] is another multi-modal dataset where objects are annotated from multiple views as opposed to KITTI. Other significant multi-modal dataset have been proposed over the years [25]–[29]. However, majority of the dataset are proposed for autonomous driving and can barely find application in fruit detection, fruit 6D-pose estimation and fruit-picking.

D. Platforms

In terms of platform, majority of the indoor multi-modal datasets are acquired with handheld methods [30], [31]. This does not account for more dynamic and rich poses.

III. Experimental Setup

To collect data on moving platforms, it becomes necessary to house the sensors on a befitting sensor rig to allow seamless data acquisition. We designed a CAD model of a sensor rig peculiar to our application. The sensor rig is capable of carrying all the required cameras (RGB-D and thermal) as seen in Fig. 3. It also has reflective markers onboard to allow for tracking and retrieval of its ground truth 6D pose. The sensors used for the data capture include an Intel RealSense D435i stereo camera, a FLIR vue pro thermal camera and an optitrack tracking system. All the Intel

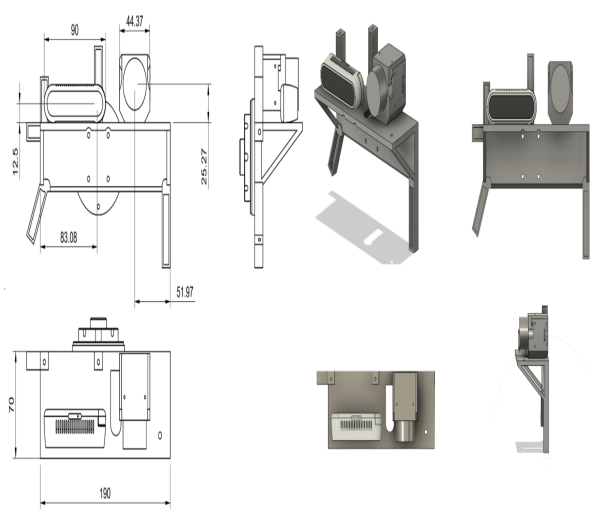

Fig. 3. Sensor Rig design drawing and CAD model

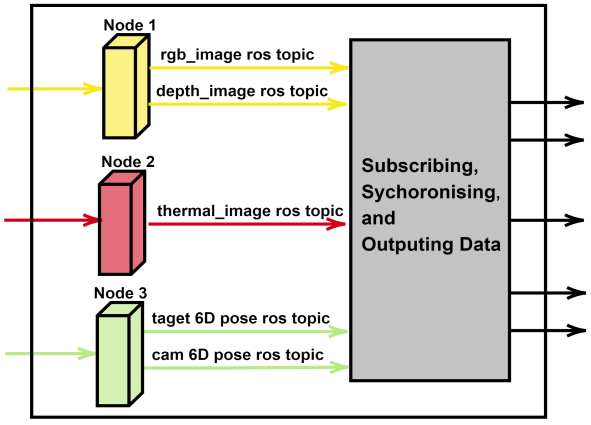

Fig. 4. ROS workflow with Node 1,2 and 3 representing the camera packages of our sensors. Respective data are published in form of ROS topics which are the utilised in a script to synchronise and save the data

RealSense camera settings were used as default as assigned by ROS. The Thermal camera was used with the grayscale filter and all other parameter used as default. The outputs of these sensors are summarized in Table I. The stereo camera provides the RGB image and the depth image. The thermal camera provides the thermal information of the fruit and the tracking cameras are used to provide the 6D positions and orientations of the camera and the fruit which is taken as the center of mass of the camera rig.

Robotics Operating System (ROS) is used for the sensor interfacing. For a consistent dataset, we require all the data to be synchronous and in real-time. We implemented this by collecting all the data from the sensors simultaneously as rostopics. Fig. 4 shows the interfacing of the sensors. The RGB-D camera is connected to the workstation and accessed through the realsense camera package. This packages collects the RGB and the depth data which are then published as rostopics. The data is then subscribed and synchronized before outputting both images. The same pipeline applies for the thermal camera but in this case using a thermal camera package. The tracking cameras are hosted on another PC running the optitrack software. The 6D poses are collected through a client package over wireless communication.

The platforms used for the dataset collection are the sawyer manipulator from Rethink Robotics, and a customized UAV shown in Fig. 5. The camera rig is also held at hand to collect more data otherwise difficult to obtain from both platforms. This enriches the data with various forms of moving thereby constituting more dynamic relative poses between the cameras and the fruit.

IV. Dataset Collection

The images and 6D poses of the fruits are collected in a sequential manner in the form of trajectories. The platforms equipped with the required sensors are subjected

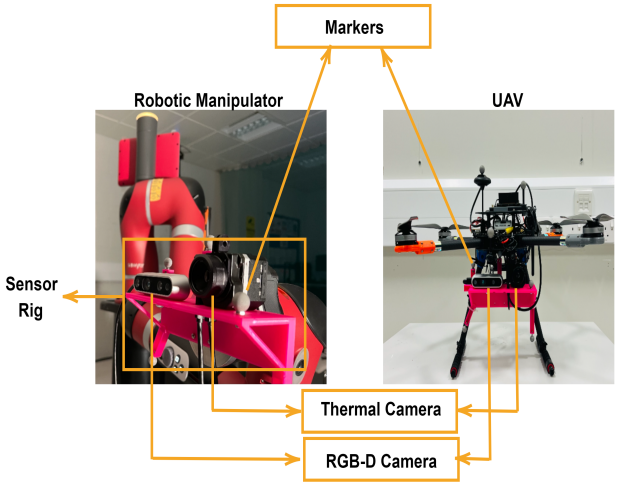

Fig. 5. Figure showing the platforms used to implement the trajectories for data acquisition. The Robotic Manipulator and UAV platforms are equipped with an RGB-D and thermal camera which are housed in the sensor rig

to a trajectory. This gives us a rich dataset with a variety of poses as there are many different relative poses between the cameras and the fruits along each trajectory. The trajectories are determined by the cameras' Field Of View (FOV), and the freedom of the platforms' joints. A definitive trajectory is likely to have frames with no objects in sight which will not only confuse the network to be trained but also add more frivolous volume to the dataset. Thus, the trajectories are intuitively implemented to cater for these limitations. Frames were captured at 30Hz along the trajectory while simultaneously recording the relative 6D pose. The 6D pose P is represented as follows:

$$P = [X_t, Y_t, Z_t, X_q, Y_q, Z_q, W_q]^T \quad (1)$$

TABLE I
SENSOR SPECIFICATIONS, TYPES AND OUTPUTS

Sensor	Type	Output
1 × RGBD Camera	Intel realSense D435i stereo camera	30Hz 8bit 640×480 RGB image 30Hz 16bit 640×480 depth image
1 × Thermal Camera	FLIR Vue Pro thermal camera	30Hz 16bit 640×480 thermal image
6 × Tracking Camera	Optitrack motion tracking system	3-dimensional position 3-dimensional orientation

where X_t, Y_t, Z_t are the 3D-translation in X, Y, Z while X_q, Y_q, Z_q, W_q are the quaternions.

A. Collection on Manipulator

We used a Sawyer robotic manipulator for the collection. It has 7 degree of freedom with a payload capacity of up to 4kg. The 3D-printed sensor rig carrying the camera setup was mounted as an end-effector to the manipulator. The trajectories were created as waypoints on the Intera software [32]. The trajectories were such that the FOV and the manipulator restraints were not impacted. A total of 5 trajectories were conducted with the manipulator. Fig. 6 shows the trajectories and the 3D-translation profile in X, Y, Z direction of the manipulator. Fig. 7 shows the quaternions of the trajectories.

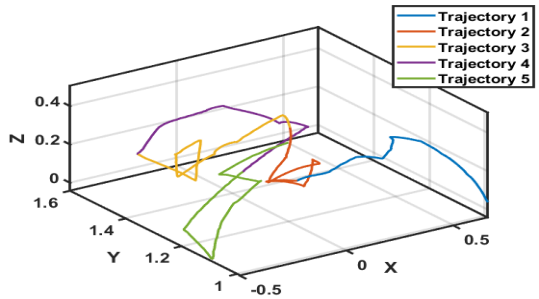

Fig. 6. Figure showing the 3D-translation $[X_t, Y_t, Z_t]$ in X, Y, Z direction of the manipulator's trajectories. Each trajectory is represented by a coloured line

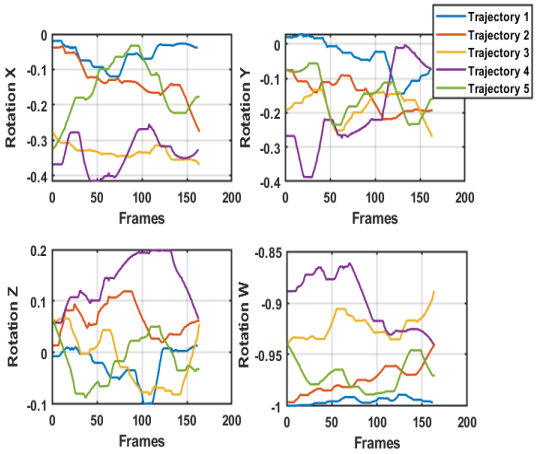

Fig. 7. Figure showing the quaternion profile $[X_q, Y_q, Z_q, W_q]$ of the manipulator's trajectories. Each trajectory is represented by a coloured line

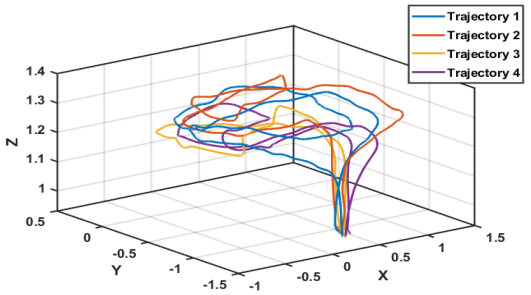

Fig. 8. Figure showing the 3D-translation $[X_t, Y_t, Z_t]$ in X, Y, Z direction of the UAV's trajectories. Each trajectory is represented by a coloured line

B. Collection on UAV

For more dynamic and rich poses, we used a UAV to collect data from more distinct poses. The sensor rig was transferred to a UAV equipped with an onboard processing unit. The UAV was manually controlled to perform trajectories such that the target fruit remains in the FOV of our sensors. A total of 4 trajectories were conducted. Fig 8 shows the UAV's 3D-translation in X, Y, Z direction withe quaternion shown in Fig 9.

C. Collection on Handheld Rig

The sensor rig was handheld to implement trajectories that are rather not possible on both the manipulator and the UAV to provide more coverage and multiple poses. The trajectories were implemented by randomly moving through a space to cover the maximum possible area without influencing the FOV. A total of 6 trajectories were conducted. As seen in Fig 10 and Fig. 11, more area was covered thereby providing more 6D-pose.

V. DATASET

A total of 33,195 images were collected which are distributed across the 3 modalities (RGB, depth, and thermal). The camera 6D-pose, fruit 6D-pose, and the relative 6D-pose between the camera and the fruit were also acquired. 15 distinct trajectories were implemented on 3 platforms to offer a variety of poses and images. Fig 12 shows the distribution of the dataset. The apple class has a total of 1869 images for each modality while the avocado class has 1900 for each modality. The banana, lemon, orange and cluster classes have

TABLE II
COMPARISON WITH OTHER FRUIT DATASETS

Dataset	Platform	Class > 1	RGB	Depth	Thermal	6D-Pose GT
Fruit 360 [9]	✗	✓	✓	✗	✗	✗
Sa et al. [14]	✗	✓	✓	✗	✗	✗
Kuang et al. [33]	✗	✓	✓	✗	✗	✗
Tu et al. [12]	✗	✗	✓	✓	✗	✗
Gene et al [15]	✗	✗	✓	✓	✗	✗
Wang et. al [11]	✗	✗	✓	✓	✗	✗
Tian et. al [13]	✗	✗	✓	✓	✗	✗
lehnert et. al [10]	Manipulator	✗	✓	✓	✗	✗
Ours	UAV/Manipulator	✓	✓	✓	✓	✓

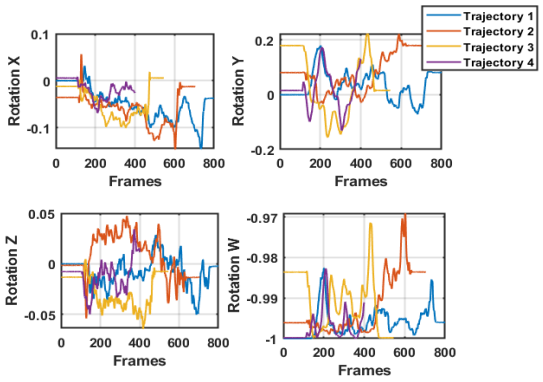

Fig. 9. Figure showing the quaternion profile $[X_q, Y_q, Z_q, W_q]$ of the UAV's trajectories. Each trajectory is represented by a coloured line

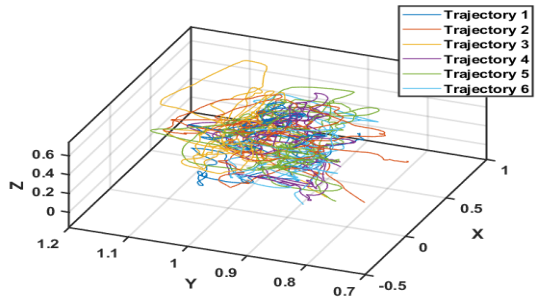

Fig. 10. Figure showing the 3D-translation $[X_t, Y_t, Z_t]$ in X, Y, Z direction of the handheld trajectories. Each trajectory is represented by a coloured line

1805, 1788, 1719, and 1984 images respectively for each modality. Each of the captured images is accompanied with 6D pose of the cameras, fruit, and the relative pose between them.

We compared our dataset with other fruit related work/dataset available in Table II. The dataset were compared based on modalities, classes, the platform used for dataset acquisition and 6D pose Ground Truth (GT). Our dataset proves to be a more complete dataset for fruit recognition and 6D-pose estimation.

The qualitative result of the dataset is shown in Fig 13

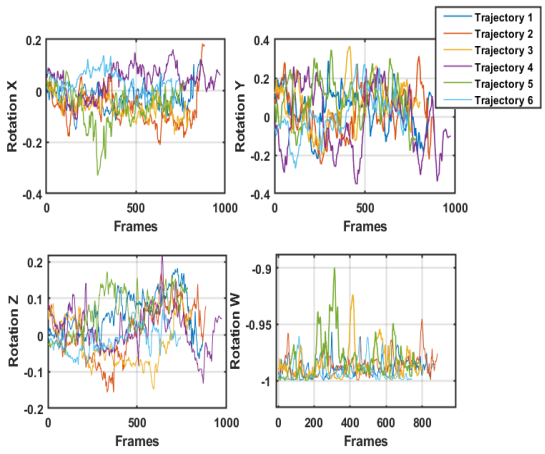

Fig. 11. Figure showing the quaternion profile $[X_q, Y_q, Z_q, W_q]$ of the handheld trajectories. Each trajectory is represented by a coloured line

for each modality captured at different 6D-poses. The depth and thermal image are displayed in different colormaps to provide visual distinction between the modalities. However, the original images in the dataset are in grayscale as in most conventional datasets.

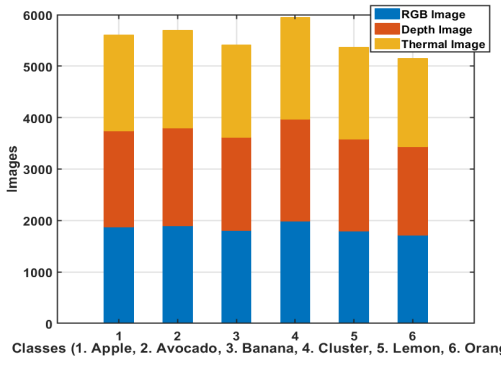

Fig. 12. Figure showing the distribution of the dataset, each colour represent a modality in the dataset

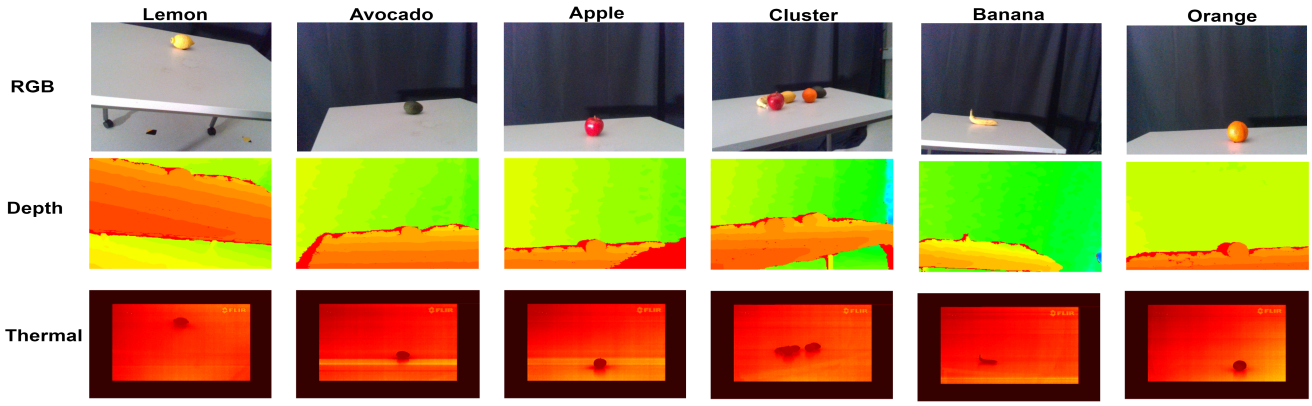

Fig. 13. Examples of Multi-Modal our dataset -top row from RGB-cameras, middle row from depth cameras and bottom row from thermal cameras The sensor rig was utilised to capture the individual fruit classes at different 6D-poses.

VI. Conclusion

This paper presents the Fruity dataset. The dataset is acquired with a stereo camera, thermal camera, and tracking camera to provide more modalities of the target (fruits). We collected the dataset on different platforms in multiple trajectories to provide a variety of 6D poses to enrich the dataset. The multiple modalities can enhance the performance of neural networks in the detection and 6D pose estimation of fruits which can be applied to fruit picking. Owing to the increase in demand for robotics in agriculture, we aim to provide the first fruit-based 6D pose estimation dataset to facilitate the use of artificial intelligence in precision agriculture. Our near future plan is to enrich the dataset with more fruit classes and also provide more scenarios such as outdoor, fruits on tree, leaves-occluded frames and more dynamic trajectories.

References

[1] Peize Li, Kaiwen Cai, Muhamad Risqi U Saputra, Zhuangzhuang Dai, and Chris Xiaoxuan Lu. Odombeyondvision: An indoor multi-modal multi-platform odometry dataset beyond the visible spectrum. *arXiv preprint arXiv:2206.01589*, 2022.

[2] Oualid Araar, Nabil Aouf, and Jose Luis Vallejo Dietz. Power pylon detection and monocular depth estimation from inspection uavs. *Industrial Robot: An International Journal*, 42(3):200–213, 2015.

[3] Mahmoud Abdulsalam and Nabil Aouf. Deep weed detector/classifier network for precision agriculture. In *2020 28th Mediterranean Conference on Control and Automation (MED)*, pages 1087–1092. IEEE, 2020.

[4] Mahmoud Abdulsalam, Kenan Ahiska, and Nabil Aouf. A novel uav-integrated deep network detection and relative position estimation approach for weeds. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, page 09544100221150284, 2023.

[5] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[6] Duarte Rondao and Nabil Aouf. Multi-view monocular pose estimation for spacecraft relative navigation. In *2018 AIAA Guidance, Navigation, and Control Conference*, page 2100, 2018.

[7] Peishan Cong, Xinge Zhu, Feng Qiao, Yiming Ren, Xidong Peng, Yuenan Hou, Lan Xu, Ruigang Yang, Dinesh Manocha, and Yuexin Ma. Stcrowd: A multimodal dataset for pedestrian perception in crowded scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19608–19617, 2022.

[8] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan, 2009.

[9] Horea Mureşan and Mihai Oltean. Fruit recognition from images using deep learning. *arXiv preprint arXiv:1712.00580*, 2017.

[10] Christopher Lehnert, Andrew English, Christopher McCool, Adam W Tow, and Tristan Perez. Autonomous sweet pepper harvesting for protected cropping systems. *IEEE Robotics and Automation Letters*, 2(2):872–879, 2017.

[11] Zhenglin Wang, Kerry B Walsh, and Brijesh Verma. On-tree mango fruit size estimation using rgb-d images. *Sensors*, 17(12):2738, 2017.

[12] Shuqin Tu, Yueju Xue, Chan Zheng, Yu Qi, Hua Wan, and Liang Mao. Detection of passion fruits and maturity classification using red-green-blue depth images. *Biosystems Engineering*, 175:156–167, 2018.

[13] Yuyu Tian, Huichuan Duan, Rong Luo, Yan Zhang, Weikuan Jia, Jian Lian, Yuanjie Zheng, Chengzhi Ruan, and Chengjiang Li. Fast recognition and location of target fruit based on depth information. *IEEE Access*, 7:170553–170563, 2019.

[14] Inkyu Sa, Zongyuan Ge, Feras Dayoub, Ben Upcroft, Tristan Perez, and Chris McCool. Deepfruits: A fruit detection system using deep neural networks. *sensors*, 16(8):1222, 2016.

[15] Jordi Gené-Mola, Verónica Vilaplana, Joan R Rosell-Polo, Josep-Ramon Morros, Javier Ruiz-Hidalgo, and Eduard Gregorio. Multi-modal deep learning for fuji apple detection using rgb-d cameras and their radiometric capabilities. *Computers and Electronics in Agriculture*, 162:689–698, 2019.

[16] Stefan Hinterstoisser, Vincent Lepetit, Slobodan Ilic, Stefan Holzer, Gary Bradski, Kurt Konolige, and Nassir Navab. Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In *Asian conference on computer vision*, pages 548–562. Springer, 2013.

[17] Eric Brachmann, Alexander Krull, Frank Michel, Stefan Gumhold, Jamie Shotton, and Carsten Rother. Learning 6d object pose estimation using 3d object coordinates. In *European conference on computer vision*, pages 536–551. Springer, 2014.

[18] Tomáš Hodan, Pavel Haluza, Štepán Obdržálek, Jiri Matas, Manolis Lourakis, and Xenophon Zabulis. T-less: An rgb-d dataset for 6d pose estimation of texture-less objects. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 880–888. IEEE, 2017.

[19] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. *arXiv preprint arXiv:1711.00199*, 2017.

[20] Alykhan Tejani, Danhang Tang, Rigas Kouskouridas, and Tae-Kyun Kim. Latent-class hough forests for 3d object detection and pose estimation. In *European Conference on Computer Vision*, pages 462–477. Springer, 2014.

[21] Andreas Doumanoglou, Rigas Kouskouridas, Sotiris Malassiotis, and Tae-Kyun Kim. 6d object detection and next-best-view prediction in the crowd. In *CVPR*, volume 1, page 2, 2016.

[22] Ziang Xie, Arjun Singh, Justin Uang, Karthik S Narayan, and Pieter Abbeel. Multimodal blending for high-accuracy instance recognition. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2214–2221. IEEE, 2013.

[23] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[24] Abhishek Patil, Srikanth Malla, Haiming Gang, and Yi-Ting Chen. The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 9552–9557. IEEE, 2019.

[25] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020.

[26] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, et al. A2d2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*, 2020.

[27] Yukyung Choi, Namil Kim, Soonmin Hwang, Kibaek Park, Jae Shin Yoon, Kyounghwan An, and In So Kweon. Kaist multi-spectral day/night data set for autonomous and assisted driving. *IEEE Transactions on Intelligent Transportation Systems*, 19(3):934–948, 2018.

[28] Sijie Zhu, Taojiannan Yang, Matias Mendieta, and Chen Chen. A3d: Adaptive 3d networks for video action recognition. *arXiv preprint arXiv:2011.12384*, 2020.

[29] John Houston, Guido Zuidhof, Luca Bergamini, Yawei Ye, Long Chen, Ashesh Jain, Sammy Omari, Vladimir Iglovikov, and Peter Ondruska. One thousand and one hours: Self-driving motion prediction dataset. In *Conference on Robot Learning*, pages 409–418. PMLR, 2021.

[30] Alex Junho Lee, Younggun Cho, Sungho Yoon, Youngsik Shin, and Ayoung Kim. Vivid: Vision for visibility dataset. In *ICRA Workshop on Dataset Generation and Benchmarking of SLAM Algorithms for Robotics and VR/AR*, 2019.

[31] Weichen Dai, Yu Zhang, Shenzhou Chen, Donglei Sun, and Da Kong. A multi-spectral dataset for evaluating motion estimation systems. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5560–5566. IEEE, 2021.

[32] Rethink Robotics. Intera, n.d. Accessed May 3, 2023. https://www.rethinkrobotics.com/intera/.

[33] Hulin Kuang, Cairong Liu, Leanne Lai Hang Chan, and Hong Yan. Multi-class fruit detection based on image region selection and improved object proposals. *Neurocomputing*, 283:241–255, 2018.