A comparative data set of annotated Broccoli heads from a moving vehicle with accompanying depth mapping data

Oliver Hardy, Karthik Seemakurthy and Elizabeth Sklar
Lincoln Institute for Agri-food Technology
University of Lincoln, UK
ohardy|esklar@lincoln.ac.uk, Karthikvjit@gmail.com

The *LAR Broccoli* dataset [1] is a collection of manually annotated video footage of broccoli heads from a UK-based organic farm late in the harvesting season. Our new data set provides a high number of annotated frames captured at 30 frames per second with relatively low cost commercially available cameras that utilise two different depth sensing methods, the RealSense D435 and Stereolabs ZED 2 camera. The broccoli images have a variety of sizes, levels of occlusion and additional interesting complications such as weeds and previously harvested broccoli stems. We also provide an annotated RGB data set of the same crop recorded with the tractor running at 3 km/h capturing blurring effects that detection algorithms will need to adapt to if autonomous broccoli harvesting machines want to operate at these speeds. The data set with accompanying annotations is a resource intended to provide an introduction point for those researching computer vision and machine learning in agriculture to apply detection algorithms to broccoli heads. The high number of annotated frames renders this data set able to be implemented with ease and allows focused efforts on algorithm development. The use of low-cost cameras enables models trained on our data set to then be applied to a field-test, again with relative ease. One limitation of our data set is the lack of ground truth data preventing absolute sizing, however the variations in the crop and extensive annotations make this a powerful data set for detection purposes. This data set is shareable and may be accessed by contacting the first author of this paper. A summary of our data set compared with the others highlighted in Section 2 is shown in Table 1.

1. Data Collection

The data were collected in October towards the very end of the harvest season from a row of the Ironman cultivar at an organic farm in Nottinghamshire, UK. The average height of the broccoli plants was approximately 55cm to the top of the head and this crop had at least one previous manual harvest performed on it before the data capture. On the day of data capture, the weather was overcast providing even lighting conditions. The three different imaging systems were mounted to a custom-built frame attached to the rear of a standard agricultural tractor and raised to a height of 1m above ground level. *As the three cameras were used on the same crop, there is an accompanying chart with the dataset that shows which annotated object (i.e. broccoli head) in each dataset corresponds to one another.* The cameras were used on individual passes of the tractor over the crop and have different fields of view therefore there are differences in the number of broccoli heads observed in each camera's data set. In addition, for the RealSense dataset, there was a loss in image capture for a period where the camera came loose from its mount; therefore the data set is split into two separate files, A and B (before and after the period of lost data). The number of broccoli heads observed in each annotated data set varied from 130 to 147 depending on the run due to changes in field of view between cameras or levels of occlusion.

2. Related work

We highlight three previous studies of identifying broccoli with computer vision that have made their datasets publicly available, as a basis of comparison for our contribution. Kusumam *et al*. [7] provide colour and depth data from three sites in Lincolnshire, UK and one site in Spain representing a range of weather conditions and plant maturity levels. The Kusumam dataset was collected from a tractor driving slowly over the crop and imaged with an RGB-D camera (RealSense D435 [8]) and a Kinect 2 [3] for creating point-cloud images. Bender *et al*. [1] produced a multi-modal 10-week time series dataset of broccoli and cauliflower images captured using an autonomous robot in New South Wales, Australia. The plants were grown in beds subject to different growing conditions such as levels of irri-

[1] "LAR" stands for Lincoln Agri-Robotics, named after a grant that funded part of the work presented here (see acknowledgments for details concerning the funding).

	Location	Cultivar	Camera(s)	Frame rate (FPS)	Number of annotated frames	Vehicle speed (km/h)	Type(s) of annotation	Ground truth head size
Kusumam *et al*. [7]	UK 1 UK 2 Spain	Ironman Ironman Titanium	RealSense D435 and Kinect 2	7.5 3.3 6.4	2,201 1,580 1,518	(walking)	centroid	yes
Bender *et al*. [1] taros	Australia	N/A	CMOS, thermal and hyper-spectral	3-4 to 6-8	1,248	N/A	bounding boxes	yes
Blok *et al*. [2]	US	Avenger	RGB and monochrome	4-10	2,560	0.5	polygons and circles	yes
	Netherlands	Ironman	RealSense D435	5-10		0		
LAR Broccoli	UK	Ironman	RealSense D435	30	**7045**	1.2	polygons and bound-ing boxes	no
			Zed 2 (stereo)		**16,178**	1.5		
			Logitech Brio (RGB)		**2,693**	3.4		

Table 1. Summary table comparing data sets

gation to create variation in growth. The robot was equipped with a stereo imaging system (two 2.3 Megapixel CMOS cameras [6]) as well as thermal and hyper-spectral imaging capabilities and a number of environmental factors were recorded from sensors not on the robot platform. Blok *et al*. [2] created a large dataset with systematic levels of *occlusion*, the issue of a broccoli head being partially or fully obscured from the camera. Blok *et al*. provide depth image data on two Broccoli fields: one in Santa Monica, California, US using a combined RGB camera (IDS UI-5280FA-C-HQ [9]) and monochromatic stereo-vision camera (IDS Ensenso N35 stereo [5]) and the other in the Netherlands using the RealSense D435 RGB-D camera. In addition to these three large datasets Zhou *et al*. [10] show an example of computer vision and deep learning based segmentation and yield estimation from digital images with no accompanying depth sensor, the images available on request from the corresponding author.

3. Annotations

The data sets were each annotated by the same person using CVAT.AI (Computer Vision Annotation Tool), an online annotation tool [4]. For each frame that a broccoli head is visible in the central row of the crop, a single polygon was carefully drawn around the perimeter of the largest visible area of a broccoli head. Where a broccoli head is completely exposed, the polygon will cover the total area of the head and where, for example, a broccoli leaf bisects the head creating two distinct visible regions of the same broccoli, only the larger of the two regions will be annotated. For the ZED 2 dataset, the left- and right-hand images of the stereo camera were independently annotated as separate items. An example of a frame from each imaging system and the corresponding annotations are shown in Figure 1

Figure 1. The same head of broccoli from each of the three cameras with polygonal annotations. Top left: Logitech, Top Right: RealSense, Bottom: Left and Right ZED 2 Stereo image

4. Acknowlegments

This work began with Innovate UK Project No: 10028115 and was completed support from LAR, Lincoln Agri-Robotics project fund for which the dataset is named.

References

[1] Asher Bender, Brett Whelan, and Salah Sukkarieh. A high-resolution, multimodal data set for agricultural robotics: A ladybird's-eye view of brassica. *Journal of Field Robotics*, 37(1):73–96, 2020.

[2] Pieter M Blok, Eldert J van Henten, Frits K van Evert, and Gert Kootstra. Image-based size estimation of broccoli heads under varying degrees of occlusion. *biosystems engineering*, 208:213–233, 2021.

[3] L. Caruso, R. Russo, and S. Savino. Microsoft kinect v2 vision system in a manufacturing application. *Robotics and Computer-Integrated Manufacturing*, 48:174–181, 2017.

[4] `https://www.cvat.ai/`.

[5] `https://en.ids-imaging.com/ensenso-n35.html`.

[6] `https://www.flir.co.uk/products/grasshopper3-usb3/`.

[7] Keerthy Kusumam, Tomáš Krajník, Simon Pearson, Tom Duckett, and Grzegorz Cielniak. 3d-vision based detection, localization, and sizing of broccoli heads in the field. *Journal of Field Robotics*, 34(8):1505–1518, 2017.

[8] `https://www.intelrealsense.com/depth-camera-d435/`.

[9] `https://www.loretech.io/products/ui-5280fa-c-hq`.

[10] Chengquan Zhou, Jun Hu, Zhifu Xu, Jibo Yue, Hongbao Ye, and Guijun Yang. A monitoring system for the segmentation and grading of broccoli head based on deep learning and neural networks. *Frontiers in Plant Science*, 11, 2020.