Assessing yield quality parameters in bush bean via RGB imagery

D. Jollet, U. Rascher, M. Müller-Linow[a]
Institute of Plant Sciences, Forschungszentrum Jülich, Jülich, Germany

Abstract

We observe an increasing demand for rapid and objective assessment of yield quality parameters in horticultural crops, which are featured by well-defined breeding targets and a multitude of parameters relevant for breeding processes. In this context, this paper addresses hardware and algorithmic requirements for a robust image-based assessment of bush bean traits according to the breeders' evaluation practice and demonstrates this with a new method for the estimation of pod length and caliber. To achieve this we built a customized acquisition box with a calibrated camera for maximally controlled imaging conditions. We trained a random-forest classifier with six different features extracted from 140 bean pod images for automatic identification of apical and basal pod ends including a check for the presence of peduncles, which together with the distal tips need to be excluded from the pod length measures with respect to marketing requirements. We validated estimations of metric pod length and caliber with manual measurements from an expert breeder and achieved in both cases a high concordance with correlations of 96%. The method developed in this study will offer new opportunities for automated sensor-based assessment of quality traits in bush beans with the potential to considerably reduce the breeding effort, in particular the manual labor-intensive scorings. The approach comes along with higher objectivity and an increased scoring resolution.

Keywords: horticulture, vegetables, image analysis, plant phenotyping, plant breeding, bush bean, shape analysis

INTRODUCTION

Breeding vegetables is a complex task, which is influenced by expectations of growers, consumers and distributors. Growers consider yield stability and quality, treatment tolerance, pathogen resistance and early vigor as important criteria (Witcombe, 2005). Buying behavior of consumers shows that they primarily pay attention to product freshness (Babicz-Zielinska and Zagorska, 1998; Jiménez-Guerrero et al., 2012; Steenkamp, 1997), a complex trait that is judged from the visual cues (mainly color-related like intensity and homogeneity) in the first place, followed by seasonality and regionality aspects (e.g. transport chains). Quality judgment from haptic and olfactory sensations plays a subordinate role. In addition, distributors consider logistic factors like packaging and stowing, and thus favor uniform products of defined size and shape. An aspect, which has been marginally studied, is the evaluation of these traits, called scorings (Simko and Hayes, 2018). Precise trait assessment requires many years of expert knowledge and efficient execution with one or two persons dedicated to one experiment or crop. This includes having scorings at the same day time and the additional use of well-known genotypes for self-calibration purposes. Despite these objectification measures several issues remain that may have an unfavorable impact on scorings: time pressure due to narrow time frames and increasingly larger experimental designs, variable individual form of each expert on the day, and increasingly little phenotypic discriminability of different cultivars. To meet the demand for faster and more objective screening methods affordable sensor-based solutions have the potential to tackle these problems and, in addition, to utilize new scoring categories

[a]E-Mail: m.mueller-linow@fz-juelich.de

and to increase scoring resolution. In this respect horticultural and especially vegetable breeding is payed little attention, i.e. new methods primarily focus on a few fruit vegetables like tomato and pepper, particularly in fruit detection (Raza et al., 2015; Schillaci et al., 2012; Song et al., 2014; Sun et al., 2018; Zhao et al., 2020) or disease detection and quantification (Liu and Wang, 2020; Raza et al., 2015; Wang et al., 2019; Wu et al., 2020). Most methods are not targeting an essential part of the breeding process, the large number of yield quality parameters, which are usually scored by hand. A first step in this direction was done by Polder et al. (1996), who developed a user-friendly software tool. CultivarJ is an ImageJ plugin that allows for the quantification of several shape parameters of legumes (pods), carrots (taproot) and onions (bulb, see also Heijden et al. (1996)). Torres et al. (2012) developed a camera-based approach with backlight, where bean pods appear as black silhouettes. However, these studies assume that cut bean samples always come with intact pod ends, displaying peduncles in particular, which is usually not the case. Besides, no validation was presented for neither of these developments.

In this manuscript we want to introduce a new approach to quantify pod parameters of legumes, which puts emphasis on the correct classification of different pod ends, a crucial step to calculate correct shape traits. In industrial processing the pods' distal tips and peduncle leftovers are removed. In addition, our approach includes an extensive validation of our results that was done in collaboration with a professional bean breeding company, which conducted hand measurements parallel to our computer vision measurements. Pod properties to be investigated with this setup are two standard traits during the breeding process of new bean varieties, namely pod length and caliber.

MATERIALS AND METHODS

Plant material and hand measurements

Several bush bean varieties were grown on an open-field site according to standard practice by *Van Waveren Saaten GmbH* at Rosdorf, 37124 Germany. 14 varieties that were phenotypically different in pod appearance were chosen for analysis. When the beans reached the state of green maturity, pods were harvested by hand. It was paid attention that there was a good mixture of pods with base and peduncles to get equally distributed classes for later training with the supervised approach. Varieties (ID: 1, 4, 8, 27, 32, 36, 69, 73, 78, 81, 82, 88, 119 and 156) were selected to cover a wide range of both color and shape features. 10 pods were analyzed per variety.

Hand measurements were conducted by an expert breeder of *Van Waveren Saaten GmbH*. First, it was determined if the bean had a peduncle or not. Furthermore, the length of the bean was measured with a ruler by unbending the bean as far as possible. Peduncles and tips were not considered.

Imaging setup

Bean pods were placed in a customized photo box with reflective walls as shown in Figure 1. This light-tight box design was chosen to ensure homogeneous illumination for a sufficiently large amount of bean pods. The top of the box contains a LED panel (*Effilux*) with integrated camera opening in the center. The base plate covers an area of 50 x 38 cm and consists of a blue robust material with low reflecting properties. The blue background facilitates pod detection in image processing. Images were acquired with a RGB camera *Manta G-507* (*Allied Vision*, 5MP) and a 12 mm Ricoh FL-CC1218-5MX lens. The camera was calibrated according to Bouguet (2005) to check the impact of radial distortions and to calculate a factor to convert the pixel to mm. The distance between camera lens and target area was approximately 80cm. At this distance no considerable distortions were observed and all further computations were conducted with the original image data.

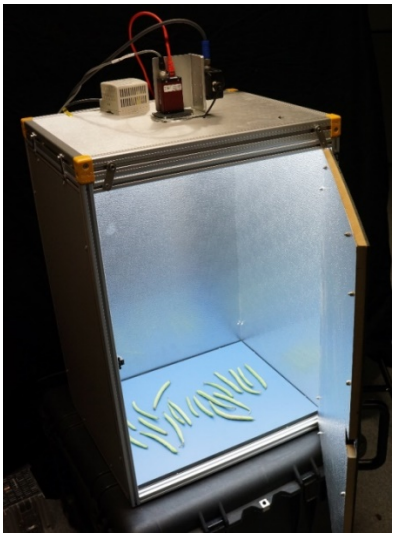

Figure 1. Photo Box in active state (door is closed during image acquisition)

Image processing and data analysis

We developed a new processing pipeline for a correct estimation of pod length and caliber and took special care that the way pods are harvested does not bias the results. Image processing was conducted with *Python* (using the *OpenCV*, *PlantCV* and *sklearn* libraries). Data analysis and visualization was done in *RStudio*. Pods were analyzed separately as follows. A *Random Forest* (RF) model was trained to classify both ends of each pod according to three categories, the pod's apical thin end (further called *tip*) and two kinds of basal ends, either a *peduncle* or not (further called *base*). RF models belong to the class of bootstrap aggregating methods, i.e. they combine several independent classification models, in this case multiple decision trees, where single outputs are finally combined to one decision (classification) according to the majority principle. Our RF classification is based on features that are derived from the change of pod width values x in a predefined range starting from each end of the pod's center line. In addition to the width data two transformations of the same are used for feature calculation, which are a moving average function $m(x)$ and a fitting function $w(x)$. The initial step is the computation of the pod's center line, which corresponds to a skeletonized binary pod mask. This is done by segmenting images in HSV (hue-saturation-value) color space with subsequent skeletonization and skeleton pruning. The hue channel of HSV was used to segment pods (composed of brown, orange, yellow, and green pixels) from the blue background such that pod pixels are represented by TRUE values and background pixels by FALSE values in the binary output. The binary mask was then skeletonized. Skeleton refinement was conducted with iterative pruning, which preserves the main branch of the skeleton. Hereby, branches up to a length of 50 pixels are subsequently removed, starting with shortest branches first. This threshold depends on image resolution and pod size and was chosen high enough (5-10% of the computed pod length) to remove larger side branches. Width values for corresponding center line pixels are then estimated by computing orthogonal lines for each center line section in a range of 10 pixels. Orthogonal lines were then compared with the binary pod representation. The length of each overlaying line segment represents the width of the pod at that specific position. The moving average function $m(x)$ was computed with a sliding frame of 50 pixels size. The fitting function $w(x)$ was fitted to the width values according to the following formula:

$$w(x) = c + \frac{(d-c)}{1 + e^{(s*(\ln(x)-\ln(p)))}}$$

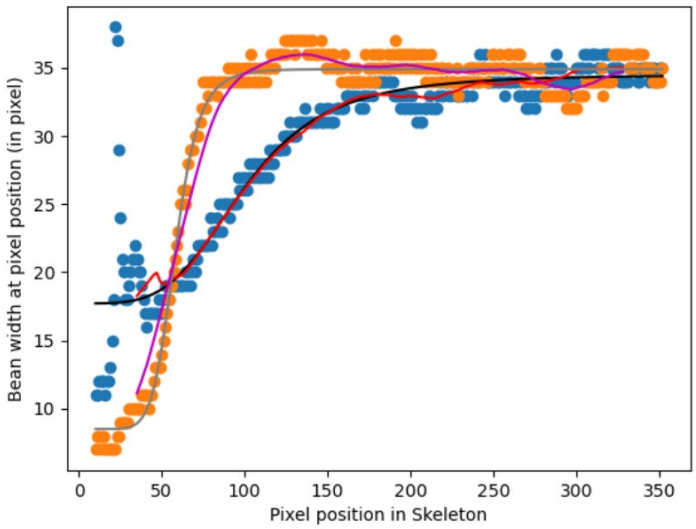

Figure 2. Pod width *x* at each pixel of the center line: Blue dots show the pod widths for one half of the bean. Orange dots show the bean widths for the second half. Black and gray lines indicate the 4-parametric log-logistic model fits *w(x)*. Red and magenta lines indicate the moving average functions *m(x)*.

In this log-logistic model parameter *x* denotes the pixel position on the pod's center line, other parameters are maximum width *c*, minimum width *d*, slope s and mirror point *p*. Figure 2 displays widths *x* and corresponding function values of *w(x)* and *m(x)* for one selected pod. Six parameters were derived from the first 70 values of *x*, *m(x) and w(x)*, respectively, and used as feature data in RF training: the minimum width in *x*, the minimum width *d* and the slope *s* estimated for *w(x)* and the coefficient *R* of a linear correlation between *x* and *w(x)*. Furthermore, two binary states, one for local decreases [i], one for outliers [ii], were calculated: [i] function *m(x)* was checked for local decreases by comparing each value in *m(x)* with the value three pixels apart. If the width of at least one position was greater than the width of the following, a local decrease was detected, resulting in a TRUE binary state for this feature. [ii] For outlier detection, the data was first normalized:

$$x' = x/w(x) \tag{1}$$

After this, outliers were detected as follows:

$$x' = \begin{cases} outlier, & x' > Q3 + 1.5 * IQR \\ outlier, & x' < Q1 - 1.5 * IQR \end{cases} \tag{2}$$

Q1 and Q3 denote lower and upper quartile, IQR denotes the inter quartile range. Again, the detection of at least one outlier resulted in a TRUE binary state. The following considerations led to the selection of the listed features. Local decreases, as displayed in Figure 2 on the red line, are typical for pod ends with a peduncle. Peduncles are also indicated by a weaker correlation between hand measured and modeled data and a smaller slope in comparison to tip ends. Also outliers are an indication of peduncles. In contrast, tips show a strong correlation

factor, a higher slope, no outliers, and no local decreases. In addition, low width minima (both modeled and measured) are characteristic for tips. Bases are normally characterized by higher minima in comparison to tips, but they also show strong correlations and no outliers and no local decreases. They also have a smaller slope, in comparison to tips.

For RF training an image data set with 120 beans was labeled and split up into 80% training data and 20% test data. The RF decision trees were trained with the feature data and parametrized as follows (see *sklearn* function for details*)*: *number* of decision trees: 100 and quality *criterion: 'entropy',* which controls the split at each node in the tree, affecting the overall tree structure. All other parameters were left unchanged to the default settings. The trained model was finally applied to the test data set. After classification tips and peduncles were cut off and excluded for the further analysis (Figure 3: red-colored) as follows: Pods' ends and consequently center line and corresponding width values were cropped at a position, where the moving average *m(x)* exceeded a particular width that gave the best results with respect to the reference measurement. Pod length was then calculated from the remaining center line with an *arcLength*-function (*OpenCV*). The caliber corresponds to the maximum width from the remaining width values *x*.

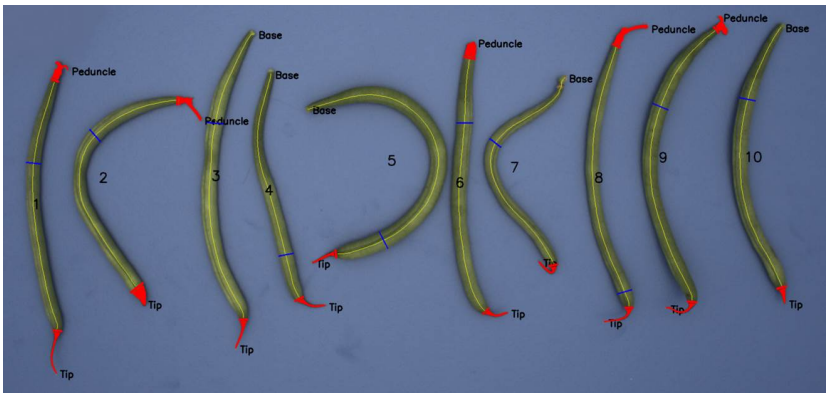

Figure 3. Processed image with beans of variety 32; red: detected tips and peduncles, excluded from further analysis; yellow: pod center line; blue: position of the caliber estimation.

RESULTS AND DISCUSSION

This experiment was conducted with 140 beans. Due to the time-consuming scorings and hand measurements a larger sample size was not possible. In 3 out of 140 beans two tips were detected (varieties 82 and 88), which is the reason, why they were excluded from further analysis. One bean from variety 81 was not considered due to a clearly wrong hand measurement. For the remaining beans the classifier reached an accuracy of 97% for the apical end (tip), 84% for the base and 85% for the peduncle, which gives a total accuracy of 85% for the classification of the basal end. Figure 3 shows one classification result: All tips at the apical end were classified correctly; the basal ends of bean 6 and 7 were misclassified. Detected peduncles and tips were cut off correctly (red regions). Misclassified peduncles (like bean 7) typically had no petioles in our experiment and were therefore contributing only small errors to the length estimation. Length and caliber values were computed and converted from pixel to metric values using a conversion factor of 0.2 mm / pixel. As illustrated in Figure 4, computer vision measurements of pod lengths correlate strongly with hand measurements, underlined by high accuracy and precision statistics with a correlation factor of $\rho = 0.96$ and a linear fit (blue line), which is almost identical to the identity line. In addition, the coefficients of variation are in

the same range for the computer vision and hand measurement data (Table 1). The findings are strengthened by an average percentage deviation between hand measurements and computer vision data of 0.49% (Table 1). Also the caliber measurements via computer vision delivered precise results and correlated strongly with the hand measurements ($\rho = 0.96$). However, we observed a systematic shift between identity line and prediction of ~1.5mm for the entire caliber range, which is caused by the background segmentation step, where some pixels that consist of both bean and blue background information (mixed pixels) have been attributed to the wrong class. The parametrization of the segmentation step could be improved to achieve a higher accuracy in the prediction of pod calibers. However, the precision of the estimations is high and the relation between computer vision estimations and the hand measurements can be used to make reliable predictions.

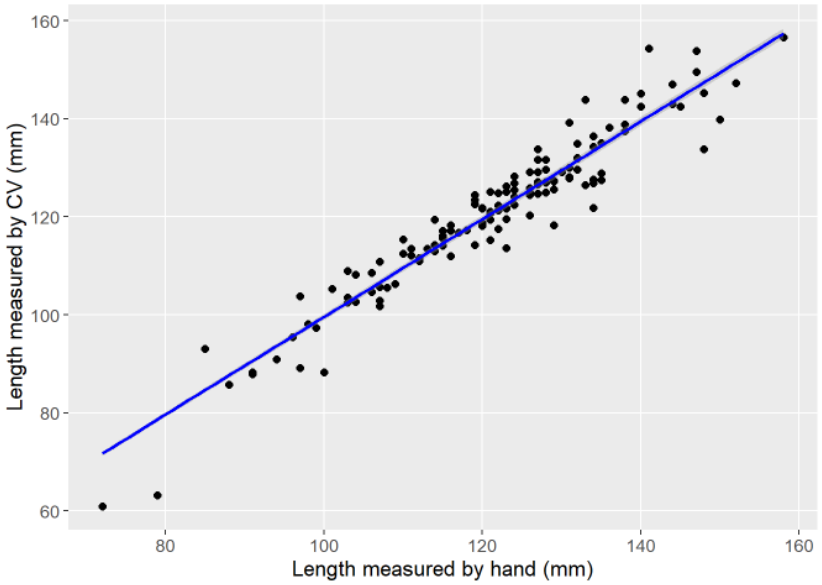

Figure 4. Comparison of length estimations via computer vision (CV) and hand-measured data. Correlation factor $\rho = 0.96$. Linear fit (blue line) is close to the identity line, indicating the high estimation quality.

CONCLUSION

We developed a customized hardware setup together with a new image processing software that allows for the simultaneous acquisition of bean pods. The LED panel and the reflective walls promote a pretty homogeneous illumination across the image area, which is important for image segmentation. In such a setup flat objects like the bean pods do not cast any shadows that are visible to the camera and which may cause problems during image processing. Simple operationality that accounts for the breeder's daily routine was one important aspect. Therefore we ensured that no particular pod preparation is required and the setup is easy to use. In our experimental study we used bean pods displaying a wide range of shapes and colors, i.e. different calibers, lengths, curvatures, and colors, and did not encounter any problem particularly related to a specific cultivar. Admittedly, our method is restricted to beans, but we assume that it is capable to work with a very wide range of bean phenotypes. Despite color traits were not examined in this study the imaging box also provides an optimal environment for color assessment. We consciously built the box without backlight that may facilitate shape detection (see e.g. Torres et al. (2012)), but complicates color recognition. In this study, an

amount of 10 bean pods per variety was analyzed simultaneously, but the size of the photo box allows the analysis of many times the amount as long as the pods are not in direct contact to each other. In any case the pod orientation is completely irrelevant. In our study the Random Forest Classifier reached a high precision of 97% for the detection of the apical end, but only 85% for the classification of the basal end. To improve this further, the classifier needs to be trained with a larger data set, which was not available for this study due to the limiting hand measurements. The image features used for classification may be improved further or completed by additional features like changes in color or brightness. We also observed a few particular cases, where the classifier failed, e.g. some tips came without pinnacle and some peduncles without petals giving them a smooth shape. This could also be compensated by a larger training set. Finally, both pod end classification and localization are typical application scenarios for deep learning approaches, which we will test on our data in the future.

Table 1. Summary table of hand-measured and computer vision data.

ID	Replicates	Peduncle number $[E]_a$	Peduncle number $[M]_b$	Mean pod length [E] (mm)	CV_c of pod length [E]	Mean pod length [M] (mm)	CV of pod length [M]	Length deviation [%]	Mean caliber [E] (mm)	Mean caliber [M] (mm)
1	10	2	3	125.28	0.08	125.60	0.08	-0.26	7.30	6.00
4	10	2	4	115.19	0.09	114.30	0.10	0.77	6.92	5.40
8	10	3	4	120.77	0.07	121.00	0.07	-0.19	7.26	5.80
27	10	5	5	100.30	0.17	104.40	0.11	-4.09	6.90	5.30
32	10	5	5	123.59	0.11	125.70	0.11	-1.71	7.62	6.10
36	10	3	2	122.21	0.04	124.10	0.05	-1.55	7.36	5.80
69	10	7	6	126.52	0.07	128.90	0.06	-1.88	8.36	7.05
73	10	9	10	124.45	0.07	125.10	0.07	-0.52	7.64	6.35
78	10	6	7	135.17	0.09	135.10	0.09	0.05	8.32	6.95
81	9	8	6	141.71	0.06	139.56	0.07	1.52	9.36	7.89
82	9	5	6	92.10	0.16	94.00	0.11	-2.06	8.49	7.00
88	8	4	4	125.42	0.12	125.75	0.10	-0.26	9.85	8.19
119	10	8	8	121.93	0.09	121.50	0.08	0.35	9.80	8.40
156	10	7	8	115.03	0.09	111.60	0.09	2.98	9.08	7.05

[a]M: manual measurements [b]E: estimated results from computer vision [c]CV: Coefficient of variation

ACKNOWLEDGMENTS

We gratefully thank the Arbeitsgemeinschaft industrieller Forschungsvereinigungen "Otto von Guericke" e.V. (AiF) and the German Federal Ministry for Economic Affairs and Energy (BMWI) for funding the project 'Shape & Color' (IGF-Vorhaben 20943 N). 'Shape & Color' is a collaboration between the Institute of Plant Sciences of Forschungszentrum Jülich and several plant breeding companies. The project deals in the application of mostly non-invasive imaging methods for the quantitative assessment of plant parameters during the selection of horticultural crops.

Literature cited

Babicz-Zielinska, E., and Zagorska, A. (1998). Factors affecting the preferences for vegetables and fruits. Polish Journal of Food and Nutrition Sciences *48*, 755-762.

Bouguet, J.-Y. (2005). Calibration toolbox for Matlab.

Heijden, G., Vossepoel, A., and Polder, G. (1996). Measuring onion cultivars with image analysis using inflection points. Euphytica *87*, 19-31.

Jiménez-Guerrero, J.F., Gázquez-Abad, J., Huertas-Garcia, R., and Mondéjar-Jiménez, J.-A. (2012). Estimating consumer preferences for extrinsic and intrinsic attributes of vegetables. A study of German consumers. Spanish Journal of Agricultural Research *10*, 539.

Liu, J., and Wang, X. (2020). Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network. Frontiers in Plant Science *11*, 12.

Polder, G., Blokker, G., and van der Heijden, G. (1996). An ImageJ plugin for plant variety testing. Paper presented at: Proceedings of the ImageJ User and Developer Conference.

Raza, S., Prince, G., Clarkson, J.P., and Rajpoot, N.M. (2015). Automatic Detection of Diseased Tomato Plants Using Thermal and Stereo Visible Light Images. Plos One *DOI:10.1371*, 20.

Schillaci, G., Pennisi, A., Franco, F., and Longo, D. (2012). Detecting tomato crops in greenhouses using a vision based method. In International Conference RAGUSA SHWA.

Simko, I., and Hayes, R. (2018). Accuracy, reliability, and timing of visual evaluations of decay in fresh-cut lettuce. PLoS ONE *13*.

Song, Y., Glasbey, C.A., Horgan, G.W., Polder, G., Dieleman, J.A., and van der Heijden, G.W.A.M. (2014). Automatic fruit recognition and counting from multiple images. Biosystems Engineering *118*, 203-215.

Steenkamp, J.E.M. (1997). Dynamics in consumer behavior with respect to agricultural and food products. In Agricultural Marketing and Consumer Behaviour in a Changing World, B. Wierenga, A. van Tilburg, K. Grunert, J.-B.E.M. Steenkamp, and M. Wedel, eds. (Kluwer Academic Publishers), pp. 143-188.

Sun, J., He, X., Ge, X., Wu, X., Shen, J., and Song, Y. (2018). Detection of Key Organs in Tomato Based on Deep Migration Learning in Complex Background. Agriculture *8*, 15.

Torres, C., Clement, A., Frison, E., Auperpin, E., Parmentier, P., and Feutry, A. (2012). Image analysis for characterising French bean (Phaseolus vulgaris L.) pods. Acta Horticulturae *934*, 145-150.

Wang, Q., Qi, F., Sun, M., Qu, J., and Xue, J. (2019). Identification of Tomato Disease Types and Detection of Infected Areas Based on Deep Convolutional Neural Networks and Object Detection Techniques. Computational Intelligence and Neuroscience, 1-15.

Witcombe, J. (2005). Crop Variety Improvement and its Effect on Productivity: the Impact of International Agricultural Research. Journal of Agricultural Science *143*, 109-109.

Wu, Q., Ji, M., and Deng, Z. (2020). Automatic Detection and Severity Assessment of Pepper Bacterial Spot Disease via MultiModels Based on Convolutional Neural Networks. International Journal of Agricultural and Environmental Information Systems *11*, 29-43.

Zhao, X., Li, H., Zhu, Q., Huang, M., Guo, Y., and Qin, J. (2020). Automatic sweet pepper detection based on point cloud images using subtractive clustering. International Journal of Agricultural and Biological Engineering *13*, 154-160.