Computers and Electronics in Agriculture 105 (2014) 111–120

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Stem localization of sweet-pepper plants using the support wire as a visual cue

C.W. Bac [a,b,*], J. Hemming [a], E.J. van Henten [a,b]

[a] Wageningen UR Greenhouse Horticulture, Wageningen University and Research Centre, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[b] Farm Technology Group, Wageningen University and Research Centre, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 28 October 2013
Received in revised form 24 March 2014
Accepted 12 April 2014

Keywords:
Stem localization
Small baseline stereo
Harvesting robots
Accuracy
Support wire

ABSTRACT

A robot arm should avoid collisions with the plant stem when it approaches a candidate sweet-pepper for harvesting. This study therefore aims at stem localization, a topic so far only studied under controlled lighting conditions. Objectives were to develop an algorithm capable of stem localization, using detection of the support wire that is twisted around the stem; to quantitatively evaluate performance of wire detection and stem localization under varying lighting conditions; to determine depth accuracy of stereo-vision under lab and greenhouse conditions. A single colour camera was mounted on a pneumatic slide to record image pairs with a small baseline of 1 cm. Artificial lighting was developed to mitigate disturbances caused by natural lighting conditions. An algorithm consisting of five steps was developed and includes novel components such as adaptive thresholding, use of support wires as a visual cue, use of object-based and 3D features and use of minimum expected stem distance. Wire detection rates (true-positive/scaled false-positive) were more favourable under moderate irradiance (94/5%) than under strong irradiance (74/26%). Error of stem localization was measured, in the horizontal plane, by Euclidean distance. Error was smaller for interpolated segments (0.8 cm), where a support wire was detected, than for extrapolated segments (1.5 cm), where a support wire was not detected. Error increased under strong irradiance. Accuracy of the stereo-vision system (±0.4 cm) met the requirements (±1 cm) in the lab, but not in the greenhouse (±4.5 cm) due to plant movement during recording. The algorithm is probably capable to construct a useful collision map for robotic harvesting, if the issue of inaccurate stereo-vision can be resolved by directions proposed for future work. This is the first study regarding stem localization under varying lighting conditions, and can be useful for future applications in crops that grow along a support wire.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

This research is part of a project in which a robot is developed to harvest sweet-pepper in a greenhouse (Hemming et al., 2011). The manipulator and end-effector of this harvesting robot should avoid obstacles during motion towards a target (fruit or peduncle). The motion planner requires locations of these obstacles. In our prior research, plant stems were obstacles more difficult to detect than fruit, leaves or petioles (Bac et al., 2013a, 2013b). This work therefore focuses on stem localization.

A low-cost sensor, a Red, Green, Blue (RGB) camera, was selected to fit economic feasibility requirements for the harvesting robot (Pekkeriet, 2011). Alternative sensors, such as LIDAR for detection of canopy structure in apple trees (Fleck et al., 2004) were considered to be too expensive. X-ray scanners were used for rose stem detection (Noordam et al., 2005), but are rather expensive and require the object to be placed between a source and receiver, which is a complicated configuration in a greenhouse environment. In our previous work, multi-spectral imaging was used (Bac et al., 2013b). But, we selected an RGB camera because the algorithm, described in this work, relies little on spectral features and uses mostly object-based features (size and shape). Such features can be extracted from RGB images as good as from multi-spectral images.

Stem detection and localization was studied under controlled lighting conditions (Paproki et al., 2012). Yet, we reviewed studies pertaining to our work that include experiments conducted under varying lighting conditions, and employ either multi-spectral imaging or colour imaging. Two studies describe classification of

* Corresponding author at: Wageningen UR Greenhouse Horticulture, Wageningen University and Research Centre, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. Tel.: +31 317 481880; fax: +31 317 418094.
E-mail address: wouter.bac@wur.nl (C.W. Bac).

http://dx.doi.org/10.1016/j.compag.2014.04.011
0168-1699/© 2014 Elsevier B.V. All rights reserved.

cucumber plant parts into leaves, stems and fruit: a study regarding a cucumber leaf picking robot (Van Henten et al., 2006) and a multi-spectral imaging study (Noble and Li, 2012). Lu et al. (2011) detected branches of citrus using multi-spectral imaging. Stems of lychee were detected using colour imaging (Deng et al., 2011). Branches and trunk of apple trees were detected using colour imaging (Jidong et al., 2012). Two articles describe classification of grape foliage into several plants parts: a study using RGB (Dey et al., 2012), and a study using multi-spectral imaging (Fernández et al., 2013). Our previous work also dealt with detection of several plants parts (Bac et al., 2013b). Although work exists regarding stem detection or fruit localization (Bac et al., 2014), to the best of our knowledge, only one article exists in which stem localization was briefly described as part of a leaf picking robot (Van Henten et al., 2006).

To localize the stem, we used stereo-vision. Accuracy of stereo matching has been thoroughly investigated (Scharstein and Szeliski, 2002), but depth accuracy of stereo-vision seems mostly qualitatively described for applications in a crop (Song et al., 2011; Van Der Heijden et al., 2012). To fill this gap, this study quantified depth accuracy and validated if the required accuracy (±1 cm) can be achieved, to localize obstacles for robotic harvesting (Hemming et al., 2011).

The approach included novel elements in terms of the baseline and algorithm. A small baseline of 1 cm was taken to improve matching score of stereo-vision and to decrease occlusion of the stem. Delon and Rougé (2007) note that few studies applied a small baseline so far and describe the advantages. Yet, a disadvantage is the difficulty to record images simultaneously. Regarding the algorithm, support wires were used as a visual cue to localize the plant stem because wires are twisted around the stem and can be distinguished from the vegetation. Support wires therefore approximate the location of the stem. Furthermore, the algorithm developed employed adaptive thresholding, object-based and 3D features, and filtering by minimum expected stem distance, to better handle varying lighting conditions.

Objectives were to (1) develop an algorithm capable of stem localization using detection of the support wire; (2) quantitatively evaluate performance of wire detection and stem localization under varying lighting conditions; (3) determine depth accuracy of stereo-vision under lab and greenhouse conditions.

This is the first study regarding stem localization and can be useful for future applications, to localize plant stems under varying lighting conditions. The algorithm and experimental set-up may not only be useful to localize obstacles for collision-free harvesting in sweet-pepper, but also in other crops that grow along a support wire, such as tomato, cucumber or egg-plant. The algorithm may furthermore fit for tasks other than harvesting, such as leaf picking, side shoot removal or plant phenotyping.

2. Image acquisition

Images of plants were recorded using an experimental set-up shown in Fig. 1. Plants were of the red sweet-pepper cultivar 'Waltz' and were cultivated in the V-system (Jovicich et al., 2004). A total of 151 stems were recorded in 38 scenes. Solar irradiance was measured and ranged from 140 to 880 W/m^2 for these recordings.

2.1. Camera and pneumatic slide

For stereo-vision, a camera was mounted on a pneumatic slide (Mini slide SLT; Festo AG & Co. KG, Germany). After recording the left image, the pneumatic slide was shifted to record the right image. Shifting took 0.4 s. The camera used was a 5 megapixel camera with a 2/3″ CCD (Prossilica GC2450C; Allied Vision Technologies GmbH, Germany). A low-distortion lens with 5 mm focal length (LM5JC10M; Kowa GmbH, Germany) was mounted on the camera. A digital laser rangefinder was used (PLR 50; Bosch GmbH, Germany) to validate calculated depth values of the stereo-vision system.

2.2. Artificial lighting

For illumination of the scene, 30 halogen lamps (230VAC, 50 W) were used. Six rows of lamps illuminated the vertical range of the image (2448 pixels). Distance between the rows was 15 cm. Each row consisted of five lamps to cover the horizontal range of the image (2050 pixels). Distance between lamps in a row was 14 cm. Similar to previous research (Bac et al., 2013b), each row was horizontally shifted (7 cm) with respect to the previous row to improve equal light distribution.

Lamps were equipped with a dichroic reflector (∅ 51 mm), to reduce strong reflections in the centre of the image. Two types of reflectors were used: a reflector causing a beam angle of 25° (GU10/50/Clear Prolite; Ritelite Ltd., UK) and a reflector causing a beam angle of 50° (HI-Spot ES50; Sylvania Europe Ltd., UK). Lamps (N = 18) emitting a beam angle of 25° were positioned at the edge of the lighting set-up, whereas lamps (N = 12) emitting a beam angle of 50° were positioned in the centre of the lighting set-up. As a result, light was more diffuse in the centre of the image than at the edge.

Fig. 1. Experimental set-up comprising a cart with a height-adjustable imaging set-up on top (left). Detailed view of the imaging set-up during recording in a row (right).

3. Algorithm

The algorithm to localize stems (Fig. 2) was developed in the image processing library HALCON© 11.0.1 (MVTec Software GmbH, Germany). Images were processed on a computer with an Intel Core i5 CPU 2.4 GHz Quad core processor with 4 GB of memory.

The algorithm starts with rectification and stereo-matching in Step 1, to obtain 3D information from the scene. In Step 2, support wires, which are twisted around the stem, are detected. After Step 2, it is yet unclear to which stem a detected wire belongs and some leaf edges are falsely detected as support wires. Steps 3 and 4 solve these issues by matching wire segments with a real stem and by removing false detections. In Step 5, (x, y, z) coordinates are extracted from the support wires to construct a sequence of coordinates that represent the stem.

3.1. Step 1: rectification and stereo-matching

Images were rectified using intrinsic camera parameters. Intrinsic camera parameters were determined by a calibration procedure (Steger et al., 2007) in HALCON that uses a calibration plate (30×30 cm) ordered from MVTec Software GmbH (Germany). The procedure estimated the centre pixel of the image, focal length, width and height of a pixel, width and height of the image, and one parameter ($\kappa = 278$) to compensate for lens distortion. Another calibration procedure (Steger et al., 2007) in HALCON was used to determine extrinsic camera parameters, i.e. the relative pose between left and right image. As a result, images satisfied the epipolar constraint (Brown et al., 2003), with a maximum measured error of 0.093 pixel. Baseline, i.e. linear displacement between left and right image, was 9.87 mm.

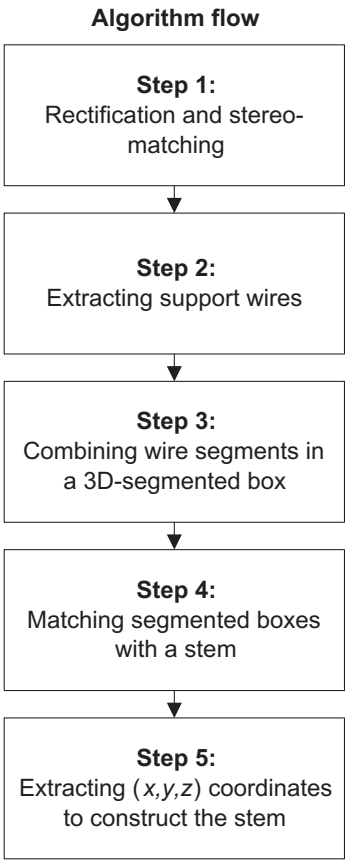

Fig. 2. Algorithm flow to localize the stem of sweet-pepper plant using stereo-RGB images.

After rectification, we applied a local stereo-matching approach using the Normalized Cross-Correlation (NCC) as block-matching method (Hannah, 1974; Brown et al., 2003). The matching method performed linear sub-pixel interpolation along the epipolar line. We used the green channel of the colour image to match vegetation pixels reflecting in this channel. A window of size 15×15 pixels was taken, which is about 1/3rd of the width of a stem, because smaller windows resulted in 'salt and pepper noise' and larger windows resulted in 'foreground fattening' (Hu and Mordohai, 2012). Minimum and maximum disparities were set corresponding to a depth range of 0.3–0.8 m because plant stems always appeared within this range. To further reduce the disparity range searched, two levels of image pyramids were used (Scharstein and Szeliski, 2002). As a result, computation time was saved. A left–right consistency check (Hu and Mordohai, 2012) was performed to filter out erroneous matches. To match objects of interest (stem and support wire) and remove other objects, a threshold was set on correlation score (>0.95). This value did not cause a false removal in the images tested. Output of Step 1 was a disparity map.

3.2. Step 2: extracting support wires

Support wires were extracted in six sub-steps: adaptive thresholding, intersection with stereo-matches, morphological image processing, thresholding on orientation, segmentation on roundness and removal of regions intersecting with the sky. We discuss each sub-step in the following paragraphs.

In Step 2a, wires – assumed as bright objects in the scene – were detected by a threshold on the R, G and B channel. Thresholds were adapted to handle illumination changes in the images. Maximum threshold was fixed at a grey value of 255, whereas minimum threshold $Threshold_{\min}$ was determined (Eq. (1)) for each channel. Eq. (1) ensured that thresholds were taken around the mean grey value of a channel and, in addition, shift from this mean was small for dark images and larger for bright images.

$$Threshold_{\min} = M_{match} + (255 - M_{match}) \cdot c \quad (-) \tag{1}$$

where M_{match} (–) mean grey value within stereo-matched region for either R, G or B channel; c (–) constant to control threshold shift from M_{match}.

Constant c was experimentally determined such that all wire segments were detected and few other objects were detected. c was 0.1 for the red channel, 0.1 for the green channel and 0.05 for the blue channel. These values were sufficient for images tested, but perturbations of 0.05 led to a worse detection performance.

In Step 2b, an intersection of four regions was taken to remove non-wire regions. These four regions comprised the three regions obtained in Step 2a and the region where a stereo-match occurred (Step 1).

In Step 2c, morphological top hat filtering was applied on the output region of Step 2b, to extract rectangular shapes that represent a wire. The rectangular element applied (width = 15 pixels and height = 2 pixels) assured that wire segments remained. As a final step, small regions (<30 pixels) were removed because these never included wire segments.

In Step 2d, more undesired regions were removed through a threshold on orientation of the region in the image (Haralick and Shapiro, 1992). Wires were twisted around the stem and therefore approximately vertically oriented. An orientation of −1.57 rad referred to an upright orientation and regions that were reasonably upright (orientation between −2.1 and −0.9 rad or between 0.9 and 2.1 rad) were kept, other regions were removed. These values were empirically determined.

In Step 2e, regions were filtered on 'roundness'. Roundness (Haralick and Shapiro, 1992) is a feature in HALCON and its values range between 0 (non-round regions) and 1 (round regions). Support wires are non-round regions with roundness between 0 and 0.47, values were empirically determined. Other regions (>0.47) were removed.

In Step 2f, regions intersecting with the sky were removed. These regions mostly occurred at the edge between a leaf and the sky and were falsely detected as a wire. To obtain the sky as region, an empirically determined threshold of >200 was applied on the blue channel of the image.

3.3. Step 3: combining wires in a 3D-segmented box

In this step, 3D features are used to combine wires in a box. Step 3 starts with intersecting detected wires (Step 2) with the disparity map (Step 1) to extract (x, y, z) coordinates by triangulation in Eqs. (2)–(4) (Rodriguez and Aggarwal, 1990).

$$x = \frac{b \cdot x_L}{d} \quad (\mathrm{m}) \tag{2}$$

$$y = \frac{b \cdot y_L}{d} \quad (\mathrm{m}) \tag{3}$$

$$z = \frac{f \cdot b}{d} \quad (\mathrm{m}) \tag{4}$$

where x is the direction of the plant row; y is the height-direction; z is depth; b (m) is the baseline; x_L (m) and y_L (m) represent the coordinate system of the left image; f (m) is focal length; d (m) is disparity.

The function '*auto_threshold*' (Steger et al., 2007), in HALCON, was applied to determine thresholds for histograms of the x-direction and z-direction. This function solved the issue of strongly differing histograms among scenes, which caused fixed thresholds to fail. The function smoothed the histogram using a Gaussian filter with a manually adjustable standard deviation. Subsequently, thresholds were extracted at minima of the smooth histogram, to generate regions. A higher standard deviation results in stronger smoothing and fewer regions. Sometimes smoothing was too strong to obtain regions. Therefore we ran the function in a loop starting at a standard deviation of 0.04 m and decreased it until '*auto_threshold*' returned regions. As a result, Step 3 generated boxes for matching with a stem in Step 4.

3.4. Step 4: matching segmented boxes with a stem

Step 4 uses expected stem distance as indicator to delete boxes probably not corresponding to a stem, i.e. to remove false detections (Fig. 3). Expected object distance has been used earlier to detect sugar beet in a row (Bontsema et al., 1991).

Three properties were used for each box: the region that covers all pixels in the box, area (pixels) of this region, and mean x-coordinate (m) of all pixels in the box. Three corresponding arrays were constructed, where number of array elements corresponds to the number of boxes. Elements were sorted based on mean x-coordinate, i.e. from small to large values, and x-displacement was computed between subsequent elements. If x-displacement was smaller than expected minimum distance of the stem, the element with the smallest number of pixels was removed. For minimum expected distance of the stem, we took a value of 10 cm because stems were mostly 10–30 cm apart.

```
% Inputs and Outputs
Boxes: array of regions that cover a box
x_coord: array of mean x-coordinates (m) for each box
Area: array of areas (pixels) for each box
% Parameters
min_displ: threshold for min expected stem displacement (m)
% Local variables
Indices: array of indices of array Boxes that should be kept
FP: boolean to indicate if there are any false-positive stem
detections
x_displ: displacement between two x-coordinates

FP:=true
while(FP)
   Count number of elements N in Boxes
   Generate Indices with values 0, 1,...N-1
   FP:=false
   for i:=0 to N-2 by 1
      x_displ:=x_coord[i]-x_coord[i-1]
      if (x_displ<min_displ)
         FP:=1
         if (Area[i]>Area[i-1])
            Delete element i-1 from Indices
         else
            Delete element i from Indices
         endif
         Create new array of Boxes with Indices of previous
         array of Boxes. Same for 'Area' and 'x_coord'.
         Break execution of for loop
      endif
   endfor
endwhile
```

Fig. 3. Pseudo-code to match segmented boxes with a stem, using minimum expected stem distance.

3.5. Step 5: extracting (x, y, z) coordinates to construct the stem

In Step 5, the algorithm extracts (x, y, z) coordinates, from pixels in a box, and connects the coordinates to represent a stem. More precisely, four operations were implemented. Firstly, pixels in a box were connected, to obtain blobs, and (x, y, z) coordinates were determined. Secondly, for each blob, the pixel with minimum y-coordinate and the pixel with maximum y-coordinate were determined. Thirdly, these two pixels were connected by a line to represent an interpolated wire segment. Fourthly, interpolated wire segments were connected by a line to represent extrapolated wire segments. In addition, extrapolated wire segments were added at the smallest and largest y-coordinate of pixels in a box. The smallest y-coordinate was connected to $y = 0$ m. Similarly, the greatest y-coordinate was connected to $y = 5$ m, which represents greenhouse height. Consequently, a localized stem consisted of interpolated wire segments and extrapolated wire segments.

4. Experiments

Four experiments were performed in this research. Firstly, performance of wire detection was visually and quantitatively assessed, to validate the algorithm (Section 4.1). Secondly, we determined the error between localized wire segments and a labelled stem, to validate accuracy of stem localization (Section 4.2). Thirdly, precision and accuracy of stereo-vision were determined, under lab conditions, to test if required accuracy (±1 cm) for obstacle localization (Hemming et al., 2011) can be achieved by stereo-vision (Section 4.3). Fourthly, accuracy was determined under greenhouse conditions as well, for comparison with lab conditions (Section 4.4).

4.1. Wire detection

Wire detection was qualitatively assessed by visualizing the output, for each of the five steps in the algorithm. Furthermore, wire detection was quantitatively assessed using the four elements of a confusion matrix: true-positive (TP), false-positive (FP), false-negative (FN) and true-negative (TN). A wire was counted as detected (TP) if at least one of its interpolated segments was actually on a real stem. If only an extrapolated wire segment, or no segment at all, intersected with the real stem, it was counted as a false detection (FP). If the algorithm did not detect any wire segment on the stem, it was counted as a missed detection (FN). Finally, a correctly removed detection (TN) was counted if a box was successfully removed in Step 4 of the algorithm (Section 3.4).

True-positive rate (TPR), (TP/(TP + FN), and scaled false-positive rate (SFPR), (FP/(TP + FN), were calculated to demonstrate performance of wire detection on 151 stems that were visible in 38 scenes. These scenes were split into two subsets based on measured outdoor solar irradiance, to test the effect of irradiance on detection rate. A subset of 28 scenes, recorded under moderate irradiance of 140–300 W/m^2, was compared with a subset of 10 scenes, recorded under strong irradiance of 300–880 W/m^2. Average grey-value of stem pixels, recorded in the green channel, ranged between 14 and 26 for scenes recorded under moderate irradiance and between 27 and 49 for scenes recorded under strong irradiance.

4.2. Error of stem localization

A stem location was approximated by connecting (x, y, z) coordinates extracted from wire segments (Section 3.5). The error of this approximation was determined by comparing localized wire segments with labelled pixels on the actual stem. We labelled visible parts of the stem by drawing a line through the vertical axis of the stem (Fig. 4). Subsequently, (x, y, z) coordinates of localized wire segments were compared with (x, y, z) coordinates of the labelled stem, obtained by intersecting the labelled stem with the disparity map and applying Eqs. (2)–(4). Thus, this experiment involved a stereo-to-stereo comparison.

To compare more coordinates than the two coordinates extracted from a localized wire segment (Section 3.5), intermediate (x, y, z) coordinates (Fig. 4) were generated through interpolation. We interpolated these coordinates such that y-coordinates were aligned with y-coordinates of the labelled stem. As a result, we were able to calculate errors for a large number of coordinates per wire segment. To express error in the horizontal plane, errors in x- and z-direction were used to calculate Euclidean distance (Eq. (5)), for each labelled pixel (Fig. 4).

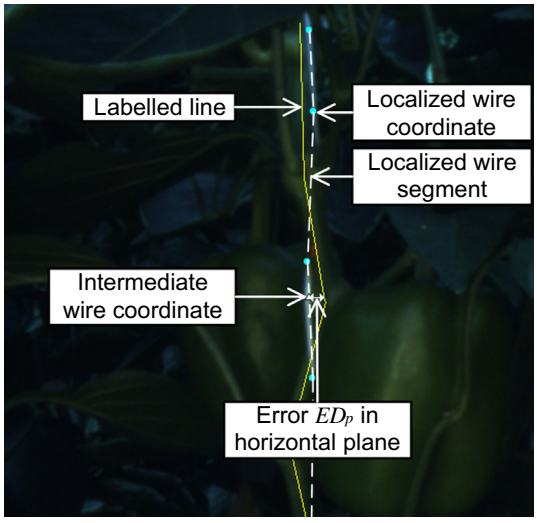

Fig. 4. The labelled line (stem) was used to calculate error ED_p relative to a localized wire segment. Only visible parts of the stem were labelled.

$$ED_p = \sqrt{(Lab_x - Loc_x)^2 + (Lab_z - Loc_z)^2} \quad \text{(cm)} \tag{5}$$

where ED_p (cm) Euclidean distance between a labelled and localized pixel; Lab_x (cm) x-coordinate of a labelled pixel; Loc_x (cm) x-coordinate of a localized pixel; Lab_z (cm) z-coordinate of a labelled pixel; Loc_z (cm) z-coordinate of a localized pixel.

Finally, Euclidean distance of a wire segment, to the stem, was calculated by averaging ED_p of all pixels in the wire segment.

We calculated mean and sample standard deviation (SD) for wire segments grouped by two variables. The first variable, type of wire segment, validated the hypothesis that Euclidean distance is smaller for interpolated segments than for extrapolated segments. We expected larger errors for extrapolated segments because of a greater distance to the stem. The second variable, irradiance, validated if Euclidean distance was greater for wires recorded under strong irradiance (300–880 W/m^2) than for wires recorded under moderate irradiance (140–300 W/m^2). For strong irradiance, 30 wires were analysed that were composed of 446 segments: 208 interpolated and 238 extrapolated. Due to complete occlusion of some of these segments, we calculated Euclidean distance for only 101 interpolated and only 130 extrapolated wire segments. For moderate irradiance, 41 wires were analysed that were composed of 415 segments: 187 interpolated and 228 extrapolated. Due to complete occlusion of some of these segments, we calculated Euclidean distance for only 126 interpolated and only 138 extrapolated wire segments.

4.3. Accuracy of stereo-vision under lab conditions

Precision was tested to evaluate consistency of stereo-matching. Subsequently, accuracy was tested to evaluate both consistency of stereo-matching and accuracy of calibration (Section 3.1).

Precision was tested by recording 100 image pairs of a constant scene (Fig. 5), at a distance of 54 cm. To compare precision for two different objects, a black dot of a calibration plate and a support wire segment of a sweet-pepper plant were detected, using a threshold in a manually drawn region-of-interest around these objects. Subsequently, a mean (x, y, z) coordinate was calculated over matched pixels in the object detected, using matching settings and triangulation equations of the algorithm (Section 3). Finally, we calculated precision, separately for x, y and z-direction, by taking population standard deviation (σ) over these 100 mean coordinates.

Whereas precision was calculated in three directions, accuracy was tested in only the z-direction (Eq. (6)) because literature indicated most inaccuracy can be expected in the z-direction (Van Henten et al., 2002) and this statement was supported by our results of precision.

$$Accuracy = \frac{\sum_{i=0}^{N} |z(Stereo)_i - z(GT)_i|}{N} \quad \text{(m)} \tag{6}$$

where $z(Stereo)_i$ (m) depth measurement i of the stereo-vision system; $z(GT)_i$ (m) depth measurement i of the ground-truth measurement device.

We recorded the same scene as for precision (Fig. 5), but camera-dot distance was varied in a range of 30–120 cm with steps of 1.2 cm. Hence, 76 image pairs were recorded.

To measure ground-truth in depth direction, a laser rangefinder was pointed at the calibration plate. Specs of the laser rangefinder report an accuracy of ±1 mm for depths less than 1 m. In addition,

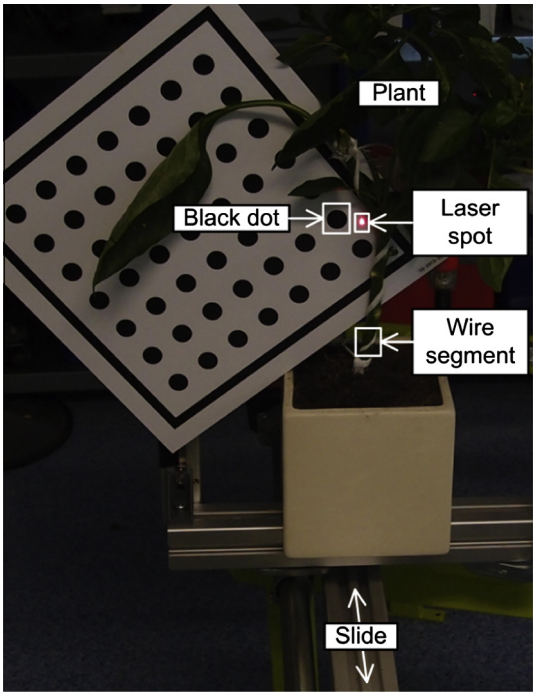

Fig. 5. Set-up used for accuracy validation of stereo-vision. Distance between camera and black dot was varied using the slide. The laser measurement served as ground truth of distance. The stereo-vision system determined distance of both the black dot and the wire segment indicated.

inaccuracy of ground truth slightly aggravated with distance due to misalignment. To align camera and laser rangefinder, we rotated the laser rangefinder until (x, y) coordinate of the localised black dot was constant for the depth range investigated. Similarly, we aligned slide and laser rangefinder by translating the slide until the laser spot appeared at the same location at the calibration plate for the depth range investigated. Such alignment is not perfectly accurate and we therefore assume a maximum offset of 3 cm in (x, y) direction at a distance of 1 m, which corresponds to an error of 0.45 mm in depth. Summing the errors (misalignment and laser range finder) led to an accuracy of ±1.1 mm at a distance of 0.3 m and ±1.5 mm at a distance of 1.2 m. Hence, we considered laser measurement a proper ground-truth because its accuracy (±1.5 mm) is greater than theoretical depth resolution of stereo-vision (10.1 mm) calculated by Eq. (7) (Rodriguez and Aggarwal, 1990).

$$\Delta z = \frac{-z^2 \cdot p \cdot s}{b \cdot f + z \cdot p \cdot s} \quad \text{(m)} \tag{7}$$

Δz (m) is the depth resolution; p pixel size was $3.45 \cdot 10^{-6}$ m; s precision of sub-pixel interpolation was assumed 1/10; b ($9.87 \cdot 10^{-3}$ m) is the baseline; f (0.005 m) is the focal length; z is the depth taken at 1.2 m.

To determine the optical camera centre, we aligned the front side of the laser rangefinder and the front side of the lens. Subsequently, mean offset between depth measurements ($N = 76$) of laser rangefinder and stereo-vision were calculated for the black dot. This offset represented the distance between the optical camera centre and the front of the lens. In addition, we added 8 mm to depth measurements of the wire to compensate for the measured depth offset between wire and calibration plate (Fig. 5).

Precision of sub-pixel interpolation s was determined for the average and maximum error measured, using Eq. (8) (Brown et al., 2003). To be able to assess s, we first checked if mismatches of one or more pixels occurred.

$$s = \frac{b \cdot f \cdot E_z}{p \cdot z^2 - p \cdot z \cdot E_z} \quad (-) \tag{8}$$

E_z (m) is the error measured in the depth direction.

4.4. Accuracy of stereo-vision under greenhouse conditions

For depth validation in the greenhouse, the laser beam was pointed at a stem and an image pair was recorded. Subsequently, the depth measurement of the laser rangefinder was read from the display and reported. There was a time lag of about 20 s between recording of the image pair and reading the depth measurement. In this time period, stems sometimes moved due to a breeze in the greenhouse and therefore we expect the error of the ground-truth to be about ±1 cm. In addition, we quantified the effect of stem movement on accuracy of stereo-vision.

To compare distance of stereo-vision with laser distance, the pixel containing the laser spot was manually selected. The laser spot did not disturb matching because the red spot was not visible in the green channel used for stereo-matching. In case no disparity was found on the laser spot, the nearest matched pixel was taken. In total, 91 stems were measured: 19 spots appeared at a wire, 72 spots appeared at a stem. In lab tests we did not find any difference in distance measurements between spots appearing at the wire and spots appearing at a stem.

5. Results

The following sub-sections correspond to experiments described in sub-sections of Section 4.

5.1. Wire detection

Output for each of the five steps in the algorithm (Section 3) is visualized in Fig. 6.

Regarding performance of wire detection for scenes recorded under moderate irradiance (113 stems), true-positive rate was 94% and scaled false-positive rate was 5%. For scenes recorded under strong irradiance (38 stems), true-positive rate was 74% and scaled false-positive rate was 26%. Hence, strong irradiance deteriorated performance. In addition, under strong irradiance, constructed stems sometimes contained a combination of correct detections of wire segments and of incorrect detections of leaf edges and fruit detected as wire segment. As a result, constructed stems contained bends that increased error of stem localization (Section 5.2). Yet, this effect was hardly observed in images recorded under moderate irradiance.

5.2. Error of stem localization

Euclidean distance in the horizontal plane was smaller for interpolated wire segments (0.8 cm; 2.9 cm) than for extrapolated wire segments (1.5 cm; 3.9 cm), both under moderate and strong irradiance (Table 1). Furthermore, for interpolated wire segments, Euclidean distance was almost four times greater under strong irradiance (2.9 cm) than under moderate irradiance (0.8 cm). Also, for extrapolated wire segments, Euclidean distance was two times greater under strong irradiance (3.9 cm) than under moderate irradiance (1.5 cm).

5.3. Accuracy of stereo-vision under lab conditions

Precision of stereo-vision for the black dot ($N = 100$) detected was 1.2 mm in x-direction, 0.1 mm in y-direction and 2.7 mm in z-direction. For the detected wire segment ($N = 100$), precision was similar: 1.0 mm in x-direction, 0.5 mm in y-direction and

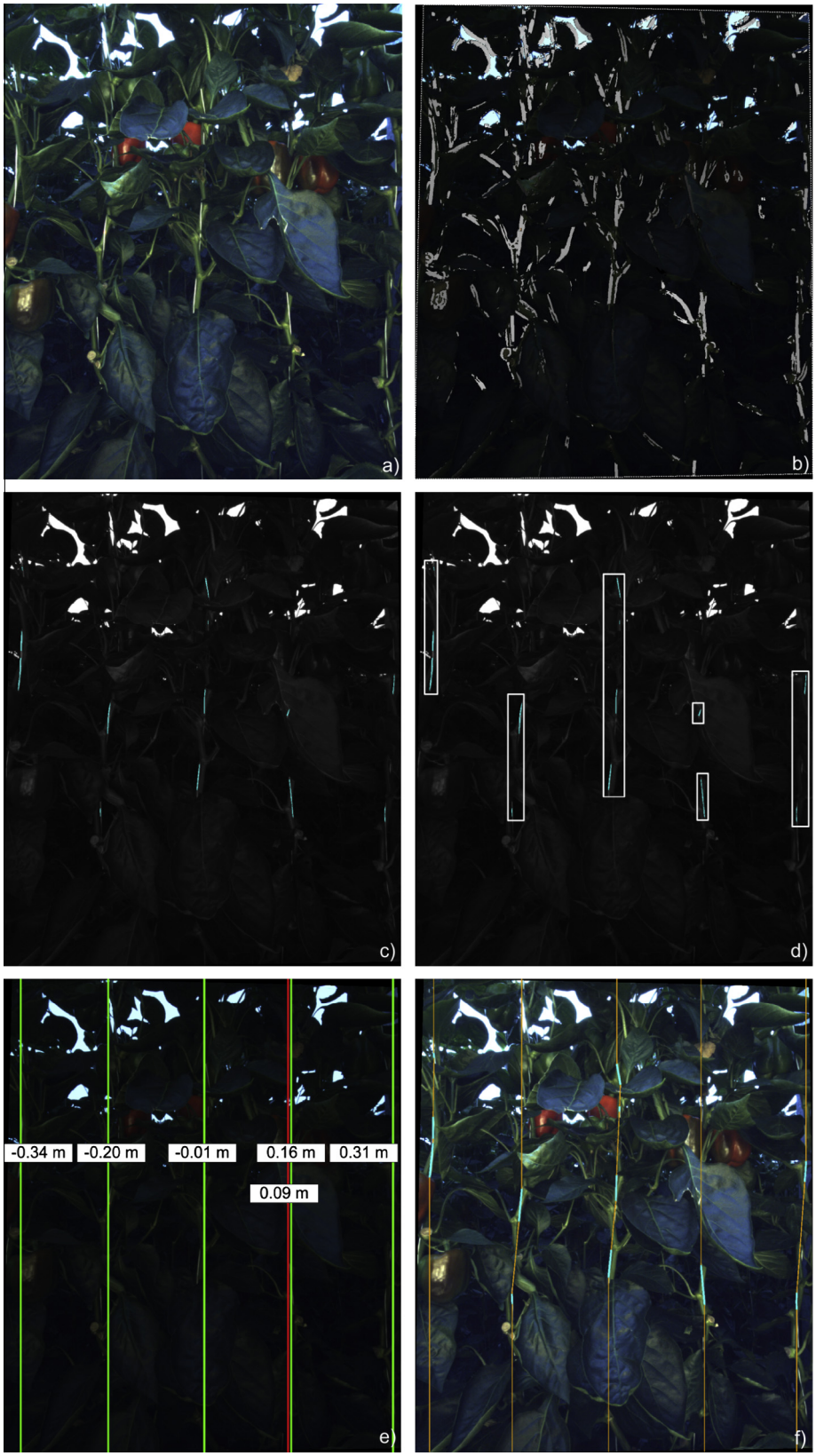

Fig. 6. Results of stem localization algorithm that consists of Steps 1–5. (a) Left frame of recorded RGB frame (histogram was scaled for better appearance). (b) Disparity image, overlayed on the rectified left frame (Step 1). (c) Extracted support wires, indicated in cyan (Step 2). (d) Wires combined in six 3D-segmented boxes (Step 3), where regions belonging to a box are indicated by a white rectangle. (e) Mean x-value (m) of pixels in each box after matching (vertical bars are for visualization of x-value). The red bar ($x = 0.09$ m) indicates a correctly rejected box. The other five green bars are correct matches of a box with a stem (Step 4). (f) Constructed stems that consist of interpolated segments, indicated in cyan, and extrapolated segments, indicated in orange (Step 5; histogram was scaled for better appearance). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Mean (SD) Euclidean distance (cm), in the horizontal plane, of wire segments to the actual stem. Data is shown for scenes recorded under moderate or strong irradiance.

Moderate irradiance (140–300 W/m^2)		Strong irradiance (300–880 W/m^2)	
Interpolated wire segments (N = 126)	Extrapolated wire segments (N = 138)	Interpolated wire segments (N = 101)	Extrapolated wire segments (N = 130)
0.8 (1.2)	1.5 (1.6)	2.9 (4.6)	3.9 (4.3)

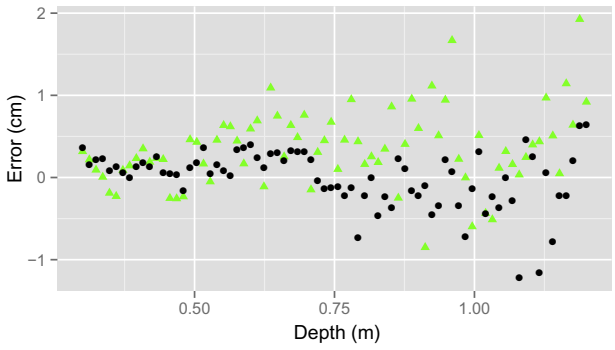

Fig. 7. Error of depth for the stereo-vision system, measured under lab conditions. Objects localized were a black dot of a calibration plate (●) and a wire segment (▲). The error increases with depth.

2.5 mm in z-direction. Since precision in x-direction is relatively high (<1.2 mm), we conclude that shifts of the pneumatic slide were accurate and did not cause an inconsistent baseline among scenes.

Error of stereo-vision for a distance of 0.3–1.2 m is in Fig. 7. The optical camera centre was 9.1 mm behind the front surface of the lens.

For the range in which stems occur (0.3–0.8 m), depth measurements of the black dot (±0.2 cm) were slightly more accurate than the wire segment (±0.4 cm). For a longer range (>0.8 m), accuracy decreased for both black dot (±0.3 cm) and wire segment (±0.6 cm). Consequently, under lab conditions, accuracy of wire localization (±0.4 cm) fits the requirement (±1 cm) (Hemming et al., 2011).

A mismatch did probably not occur because a mismatch of one pixel would cause an error of 0.6 cm at a depth of 30 cm and an error of 9.5 cm at a depth of 120 cm. Such large errors were not observed (Fig. 7). Therefore we were able to assess precision of sub-pixel interpolation: 1/11 for an average error (±0.4 cm) and 1/5 for the maximum error observed (1.1 cm).

5.4. Accuracy of stereo-vision under greenhouse conditions

Accuracy of stereo-vision was ±4.5 cm and was based on 91 measurements. This accuracy deviates from the accuracy (±1 cm) required for harvesting (Hemming et al., 2011). Analysis of a scene with a large error (−8.7 cm) shows that the pixel was correctly matched (Fig. 8).

Whereas we did not observe a mismatch of pixels (Fig. 8), movement of the plant or cart during recording might explain the large error (−8.7 cm). Probably about one cm of this error can be explained by inaccuracy of stereo-vision or ground-truth. The remaining error corresponds to a disparity offset of three pixels at a laser-stem distance of 66.5 cm. If a plant would have moved horizontally by three pixels, speed of plant movement would have been 3.0 mm/s. We calculated this speed by dividing real-world width of three pixels (1.2 mm), at a depth of 66.5 cm, by the duration of the slide movement (0.4 s). Whereas we did not measure speed of plant movement, a speed of 3.0 mm/s seems likely to have occurred because of the breeze we observed in the greenhouse.

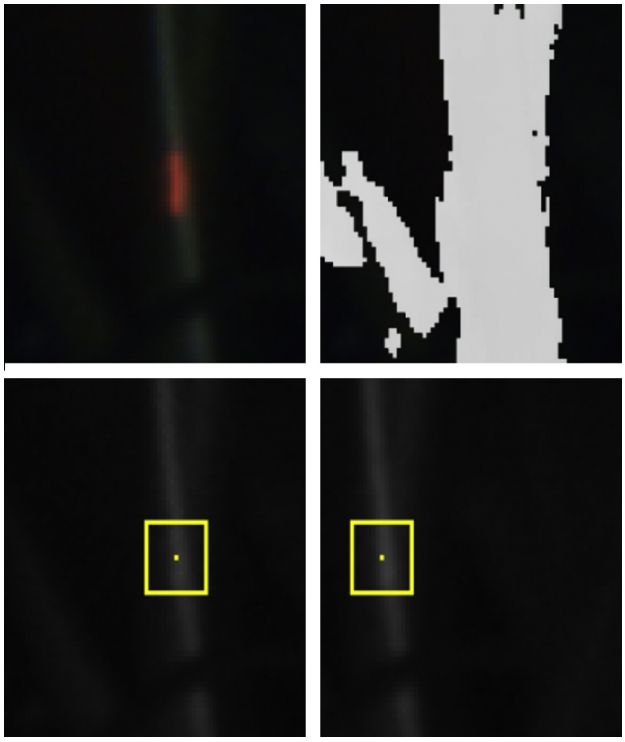

Fig. 8. An example of a stereo-match on the laser spot. RGB image with the laser spot visible on a wire (top-left). The disparity map (top-right) obtained after stereo-matching of the left frame (bottom-left) and right frame (bottom-right). The mask (15 × 15 pixels) is displayed in the left and right frame. Also note that the laser spot was not visible in the green channel used for stereo-matching.

6. Discussion

We discuss each experiment in the following paragraphs.

Wire detection yielded a true-positive rate (TPR) of 94% and a scaled false-positive rate (SFPR) of 5% under moderate irradiance. To avoid the performance drop under strong irradiance (TPR of 74%; SFPR of 26%), exposure time of the camera can be adapted and algorithm parameters can be optimized in future work. We were unable to compare performance with related work regarding stem detection (Bac et al., 2013b; Fernández et al., 2013) because these studies report performance on a pixel basis, whereas this study reports performance on a blob basis. Therefore we compared performance with the state-of-the-art in green fruit detection. For green apple detection, Linker et al. (2012) achieved similar performance: a TPR of 95% and a SFPR of 4% under diffuse natural lighting and a TPR of 88% and a SFPR of 25% under intense natural lighting.

Errors of stem localization under strong irradiance (2.9 cm) were mainly due to leaf edges and fruit falsely being detected as wire. Adapting exposure time may decrease these errors because, in darker images, leaf edges and fruit appeared darker than wires and were therefore not detected as wire. To compensate for errors, an additional danger zone can be taken into account during planning of collision-free motions. Planning such motions may improve harvest success and decrease damages to the plant compared with current harvesting robots that mostly do not consider obstacles

(Bac et al., 2014). However, adding danger zones may reduce the chance to find a collision-free motion and might increase calculation time needed for motion planning.

Error of stereo-vision (±0.4 cm) under lab conditions can be explained by inaccurate calibration, inaccuracy of ground-truth (±1.1–1.6 mm) and varying level of sub-pixel interpolation (±2.6 mm) due to illumination change (Brown et al., 2003). Sub-pixel interpolation achieved was about 1/11 and this value fits with Brown et al. (2003) who indicate it is difficult to obtain sub-pixel interpolation better than 1/10.

Comparing accuracy of stereo-vision under greenhouse conditions (±4.6 cm), with the literature, was hard because little work exists. In a related study sweet-pepper fruit were localized under greenhouse conditions using stereo-vision and time-of-flight images (Song et al., 2011; Van Der Heijden et al., 2012). Although these authors do not report accuracy, they indicate pixel values were not reliable for matching due to changes of perspective, lighting and noise. Such changes may partly explain why accuracy was much worse under greenhouse conditions (±4.6 cm) than under lab conditions (±0.4 cm). Yet, the main reason of poor accuracy is probably movement of the plant or cart during recording. Due to the poor accuracy achieved, we were unable to separately assess the effect of perspective, lighting and noise because of the strong effect of plant movement on error of depth measurements. For the same reason we were unable to assess precision of sub-pixel interpolation. Therefore simultaneous acquisition of the left and right image is critical for accurate localization in future work, as also indicated by others (Biskup et al., 2007). Simultaneous acquisition is only possible using a wider baseline than the one chosen in this research (1 cm) because the width of the current camera restricts minimum baseline to 3.5 cm. A drawback of such wider baseline is less similar images (more occlusion, more noise, larger geometrical deformations) and causes less accurate disparity measurements due to a more difficult matching process (Delon and Rougé, 2007). In addition, performance of the wire detection algorithm might drop because support wires can become more occluded by leaves. Therefore we suggest to use two baselines: a small baseline of 1 cm to detect stems and a baseline of 3.5 cm to accurately localize as many of the detected pixels as possible. Alternatively a solution with a mirror and a beam splitter can be used (Pachidis and Lygouras, 2005), or a combination of RGB and time-of-flight images.

7. Conclusion

The algorithm developed is capable of stem localization using the support wire as a visual cue and using stereo-images with a small baseline (1 cm). Novel components of the algorithm include adaptive thresholding, use of support wires as a visual cue, use of object-based and 3D features and use of minimum expected stem distance.

Wires were detected with a true-positive rate (TPR) of 94% and a scaled false-positive rate (SFPR) of 5%, under moderate irradiance. Although related work that includes detection performance does not exist, this performance is comparable with state-of-the-art performance of green fruit detection. The algorithm, however, suffers from strong irradiance because detection rate dropped to a TPR 74% and an SFPR of 26%. Also, error of stem localization suffers from strong irradiance because error was two times greater for interpolated wire segments and four times greater for extrapolated wire segments. Adapting exposure time for irradiance is therefore a task for future work.

The algorithm is probably capable to construct a useful collision map for robotic harvesting, if the issue of inaccurate localization can be resolved in future work. Accuracy of stereo-vision meets the required accuracy (±1 cm) under lab conditions (±0.4 cm), but not under greenhouse conditions (±4.5 cm). Plant movement seems the major cause of this decreasing accuracy and simultaneous acquisition of the left and right frame is needed to avoid errors caused by plant movement. A possible direction for future work is to investigate double-baseline stereo, a combination with time-of-flight images, or a solution with mirrors and a beam splitter.

Acknowledgements

We thank Yael Edan and Bart van Tuijl for their contributions to the experimental set-up and experiments. We are grateful to growers Ted and Robert Vollebregt and Cees and Rolf Vijverberg who allowed us to perform experiments in their greenhouses. This research was funded by the European Commission in the 7th Framework Programme (CROPS GA No. 246252) and by the Dutch Horticultural Product Board (PT No. 14555).

References

Bac, C.W., Hemming, J., Van Henten, E.J., 2013a. Pixel classification and post-processing of plant parts using multi-spectral images of sweet-pepper. In: IFAC Biorobotics Conference, Sakai, Japan, 27–29 March 2013, pp. 150–155.

Bac, C.W., Hemming, J., Van Henten, E.J., 2013b. Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper. Comput. Electron. Agric. 96, 148–162.

Bac, C.W., Van Henten, E.J., Hemming, J., Edan, Y., 2014. Harvesting robots for high-value crops: state-of-the-art review and challenges ahead. J. Field Robot. http://dx.doi.org/10.1002/rob.21525.

Biskup, B., Scharr, H., Schurr, U., Rascher, U., 2007. A stereo imaging system for measuring structural parameters of plant canopies. Plant, Cell Environ. 30, 1299–1308.

Bontsema, J., Grift, T.E., Pleijsier, K., 1991. Mechanical weed control in sugar beet growing: the detection of a plant in a row. In: Hashimoto, Y., Day, W. (Eds.), Mathematical and Control Applications in Agriculture and Horticulture. Matsuyama, Japan, pp. 207–212.

Brown, M.Z., Burschka, D., Hager, G.D., 2003. Advances in computational stereo. IEEE Trans. Pattern Anal. Mach. Intell. 25, 993–1008.

Delon, J., Rougé, B., 2007. Small baseline stereovision. J. Math. Imag. Vis. 28, 209–223.

Deng, J., Li, J., Zou, X., 2011. Extraction of litchi stem based on computer vision under natural scene. In: Proceedings – International Conference on Computer Distributed Control and Intelligent Environmental Monitoring, CDCIEM, 2011, pp. 832–835.

Dey, D., Mummert, L., Sukthankar, R., 2012. Classification of plant structures from uncalibrated image sequences. Proceedings of IEEE Workshop on Applications of Computer Vision, pp. 329–336.

Fernández, R., Montes, H., Salinas, C., Sarria, J., Armada, M., 2013. Combination of RGB and multispectral imagery for discrimination of Cabernet Sauvignon grapevine elements. Sensors (Switzerland) 13, 7838–7859.

Fleck, S., Van Der Zande, D., Schmidt, M., Coppin, P., 2004. Reconstructions of tree structure from laser-scans and their use to predict physiological properties and processes in canopies. Int. Arch. Photogramm., Remote Sens. Spatial Inform. Sci. 36, 119–123.

Hannah, M.J., 1974. Computer Matching of Areas in Stereo Images. PhD Thesis, Stanford University, CA, USA.

Haralick, R.M., Shapiro, L.G., 1992. Computer and Robot Vision. Addison-Wesley, Boston, USA.

Hemming, J., Bac, C.W., Tuijl, B.A.J., 2011. CROPS project Deliverable 5.1: Report with Design Objectives and Requirements for Sweet-pepper Harvesting. Wageningen UR Greenhouse Horticulture, Wageningen, The Netherlands.

Hu, X., Mordohai, P., 2012. A quantitative evaluation of confidence measures for stereo vision. IEEE Trans. Pattern Anal. Mach. Intell. 34, 2121–2133.

Jidong, L., De-An, Z., Wei, J., Yu, C., Ying, Z., 2012. Research on trunk and branch recognition method of apple harvesting robot. In: International Conference on Measurement, Information and Control (MIC), 2012, pp. 474–478.

Jovicich, E., Cnatliffe, D.J., Sargent, S.A., Osborne, L.S., 2004. Production of Greenhouse-Grown Peppers in Florida. University of Florida, IFAS Extension, Gainesville, FL.

Linker, R., Cohen, O., Naor, A., 2012. Determination of the number of green apples in RGB images recorded in orchards. Comput. Electron. Agric. 81, 45–57.

Lu, Q., Tang, M., Cai, J., 2011. Obstacle recognition using multi-spectral imaging for citrus picking robot. In: Proceedings – PACCS 2011: 2011 3rd Pacific–Asia Conference on Circuits, Communications and System, Wuhan, China, pp. 1–5.

Noble, S., Li, D., 2012. Segmentation of greenhouse cucumber plants in multi-spectral imagery. In: International Conference of Agricultural Engineering, CIGR-Ageng, Valencia, Spain, pp. 1–5.

Noordam, J.C., Hemming, J., van Heerde, C., Golbach, F., van Soest, R., Wekking, E., 2005. Automated rose cutting in Greenhouses with 3D vision and robotics: analysis of 3D vision techniques for stem detection. Acta Hort. (ISHS) 691, 885–892.

Pachidis, T.P., Lygouras, J.N., 2005. Pseudo-stereo vision system: a detailed study. J. Intell. Robot. Syst.: Theor. Appl. 42, 135–167.

Paproki, A., Sirault, X., Berry, S., Furbank, R., Fripp, J., 2012. A novel mesh processing based technique for 3D plant analysis. BMC Plant Biol. 12, 63.

Pekkeriet, E.J., 2011. CROPS Project Deliverable 12.1: Economic Viability for Each Application. Wageningen UR Greenhouse Horticulture, Wageningen, The Netherlands.

Rodriguez, J.J., Aggarwal, J.K., 1990. Stochastic analysis of stereo quantization error. IEEE Trans. Pattern Anal. Mach. Intell. 12, 467–470. http://dx.doi.org/10.1109/TPAMI.1982.4767278.

Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vis. 47, 7–42.

Song, Y., Glasbey, C.A., Van Der Heijden, G.W.A.M., Polder, G., Dieleman, J.A., 2011. Combining stereo and time-of-flight images with application to automatic plant phenotyping. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 467–478.

Steger, C., Ulrich, M., Wiedemann, C., 2007. Machine Vision Algorithms, Machine Vision Algorithms and Applications. Wiley-VCH, Weinheim, Germany (Chapter 3).

Van Der Heijden, G., Song, Y., Horgan, G., Polder, G., Dieleman, A., Bink, M., Palloix, A., Van Eeuwijk, F., Glasbey, C., 2012. SPICY: towards automated phenotyping of large pepper plants in the greenhouse. Funct. Plant Biol. 39, 870–877.

Van Henten, E.J., Hemming, J., Van Tuijl, B.A.J., Kornet, J.G., Meuleman, J., Bontsema, J., van Os, E.A., 2002. An autonomous robot for harvesting cucumbers in greenhouses. Auton. Robot. 13, 241–258.

Van Henten, E.J., Van Tuijl, B.A.J., Hoogakker, G.J., Van Der Weerd, M.J., Hemming, J., Kornet, J.G., Bontsema, J., 2006. An autonomous robot for de-leafing cucumber plants grown in a high-wire cultivation system. Biosyst. Eng. 94, 317–323.