



## Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper



C.W. Bac<sup>a,b,\*</sup>, J. Hemming<sup>a</sup>, E.J. van Henten<sup>a,b</sup>

<sup>a</sup> Wageningen UR Greenhouse Horticulture, Wageningen University and Research Centre, Droevedaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>b</sup> Farm Technology Group, Wageningen University and Research Centre, Droevedaalsesteeg 1, 6708 PB Wageningen, The Netherlands

### ARTICLE INFO

#### Article history:

Received 23 November 2012

Received in revised form 29 April 2013

Accepted 4 May 2013

### ABSTRACT

Sweet-pepper plant parts should be distinguished to construct an obstacle map to plan collision-free motion for a harvesting manipulator. Objectives were to segment vegetation from the background; to segment non-vegetation objects; to construct a classifier robust to variation among scenes; and to classify vegetation primarily into soft (top of a leaf, bottom of leaf and petiole) and hard obstacles (stem and fruit) and secondarily into five plant parts: stem, top of a leaf, bottom of a leaf, fruit and petiole. A multi-spectral system with artificial lighting was developed to mitigate disturbances caused by natural lighting conditions. The background was successfully segmented from vegetation using a threshold in a near-infrared wavelength (>900 nm). Non-vegetation objects occurring in the scene, including drippers, pots, sticks, construction elements and support wires, were removed using a threshold in the blue wavelength (447 nm). Vegetation was classified, using a Classification and Regression Trees (CART) classifier trained with 46 pixel-based features. The Normalized Difference Index features were the strongest as selected by a Sequential Floating Forward Selection algorithm. A new robust-and-balanced accuracy performance measure  $P_{Rob}$  was introduced for CART pruning and feature selection. Use of  $P_{Rob}$  rendered the classifier more robust to variation among scenes because standard deviation among scenes reduced 59% for hard obstacles and 43% for soft obstacles compared with balanced accuracy. Two approaches were derived to classify vegetation: Approach A was based on hard vs. soft obstacle classification and Approach B was based on separability of classes. Approach A ( $P_{Rob} = 58.9$ ) performed slightly better than Approach B ( $P_{Rob} = 56.1$ ). For Approach A, mean true-positive detection rate (standard deviation) among scenes was 59.2 (7.1)% for hard obstacles, 91.5 (4.0)% for soft obstacles, 40.0 (12.4)% for stems, 78.7 (16.0)% for top of a leaf, 68.5 (11.4)% for bottom of a leaf, 54.5 (9.9)% for fruit and 49.5 (13.6)% for petiole. These results are insufficient to construct an accurate obstacle map and suggestions for improvements are described. Nevertheless, this is the first study that reports quantitative performance for classification of several plant parts under varying lighting conditions.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

This research is part of the EU funded CROPS project, 'Clever Robots for Crops', in which a sweet-pepper harvesting robot will be developed (Hemming et al., 2011). A harvesting robot includes a manipulator and end-effector for which collision-free motions should be planned to be able to reach a target (fruit or peduncle). The motion planner therefore requires locations of obstacles. But, before obstacle locations can be determined, obstacles must be de-

tected. Consequently, obstacle detection is the focus of this research.

We decided to separate obstacles in *hard obstacles* and *soft obstacles* because the dense obstacle map requires the manipulator to push some obstacles aside to reach the target. Soft obstacles (leaves and petioles) can be pushed aside because damage to these plant parts would hardly harm the physiological status of the plant. Collisions with hard obstacles (stems, fruits, support wires, construction elements) are, however, critical. For instance, a damage to the stem may strongly limit future fruit set. Similarly, a small scratch in the fruit skin will reduce the market value of a fruit. Support wires are twisted around the stem to guide plant growth upward. A broken support wire would cause the plant to drop on the floor and such an event would severely damage the plant. A collision with a greenhouse construction element can

\* Corresponding author at: Wageningen UR Greenhouse Horticulture, Wageningen University and Research Centre, Droevedaalsesteeg 1, 6708 PB Wageningen, The Netherlands. Tel.: +31 317 481880; fax: +31 317 418094.

E-mail address: [wouter.bac@wur.nl](mailto:wouter.bac@wur.nl) (C.W. Bac).

cause damage to the greenhouse construction, manipulator or end-effector.

Moreover, leaves were classified in the top side of a leaf (adaxial) and the bottom side of a leaf (abaxial) because this information may be required if either an upward or downward leaf motion is desired. Generally, leaves will move upward when pushed from the bottom side and move downward when pushed from the top side. In some motion planning cases it can be useful to know whether a leaf will move upward or downward when pushed. For instance, a leaf on top of a ripe fruit should be moved upward by the gripper to avoid leaf damage during fruit detachment.

A low-cost sensing solution, multi-spectral imaging, was selected to improve economic feasibility of the harvesting robot. Alternative sensors used in related applications, such as LIDAR for detection of canopy structure in apple trees (Fleck et al., 2004) and X-ray for rose stem detection (Noordam et al., 2005), were considered to be too expensive. We used multi-spectral imaging instead of colour imaging to use the spectral information present in both visible and near-infrared wavelengths. Two near-infrared wavelengths were selected: one within the red edge (696–736 nm) (Filella and Penuelas, 1994) and one within a water absorption band (900–1000 nm) (Center, 2008). Combinations of a wavelength in visible and near-infrared spectrum turned out to be useful features in this research.

Little research has been performed regarding obstacle detection for fruit harvesting (Bac et al., submitted for publication) and quantitative performance was not reported. We reviewed studies performed under varying lighting conditions using either multi-spectral imaging or colour imaging. Two studies describe classification of cucumber plant parts into leaves, stems and fruits: a study regarding a cucumber leaf picking robot using two near-infrared wavelengths (Van Henten et al., 2006) and a multi-spectral imaging study in which several wavelengths and sensors were compared (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). Unfortunately, these studies lack quantitative classification performance. The article most closely related to the work presented here is classification of grape foliage into leaves, branches and fruits (green or coloured) using RGB cameras with a Support Vector Machine classifier. For green grapes, true-positive rate was 91.9% with a false-positive rate of 2.7%. Performance for branches and leaves was not reported (Dey et al., 2012). To the best of our knowledge, quantitative performance of plant part classification under varying lighting conditions was only reported for fruit detection (Jiménez et al., 2000), whereas this study reports quantitative performance for five plant parts.

A Classification And Regression Trees (CART) classifier was used for classification (Breiman et al., 1984). CART is a variant of a decision tree classifier and performs similar to variant C4.5 (Unay et al., 2006). Decision trees were hardly used in related classification studies, whereas decision trees can reach classification accuracies similar to other classifiers (Kavdir and Guyer, 2004) and fit the requirements listed in this article.

Two literature reviews served as inspiration for the approach taken. Firstly, McCarthy et al. (2010) discuss how disturbances by natural lighting can be addressed upon design of the vision system. Based on that, we assembled intense lighting to mitigate natural lighting disturbances. Moreover, a new classification performance measure, *robust-and-balanced accuracy*  $P_{Rob}$ , is introduced that considers both balanced accuracy and robust classification performance among scenes. Robustness is defined as the ability of a system to resist change without adapting its initial stable configuration (Wieland and Wallenburg, 2012). We show the advantage of  $P_{Rob}$  compared with balanced accuracy as performance measure. Secondly, Kapach et al. (2012) show that few studies use visual cues other than colour, i.e. shape, size and especially texture, for agricultural applications. We therefore evaluated spectral and texture features to classify pixels.

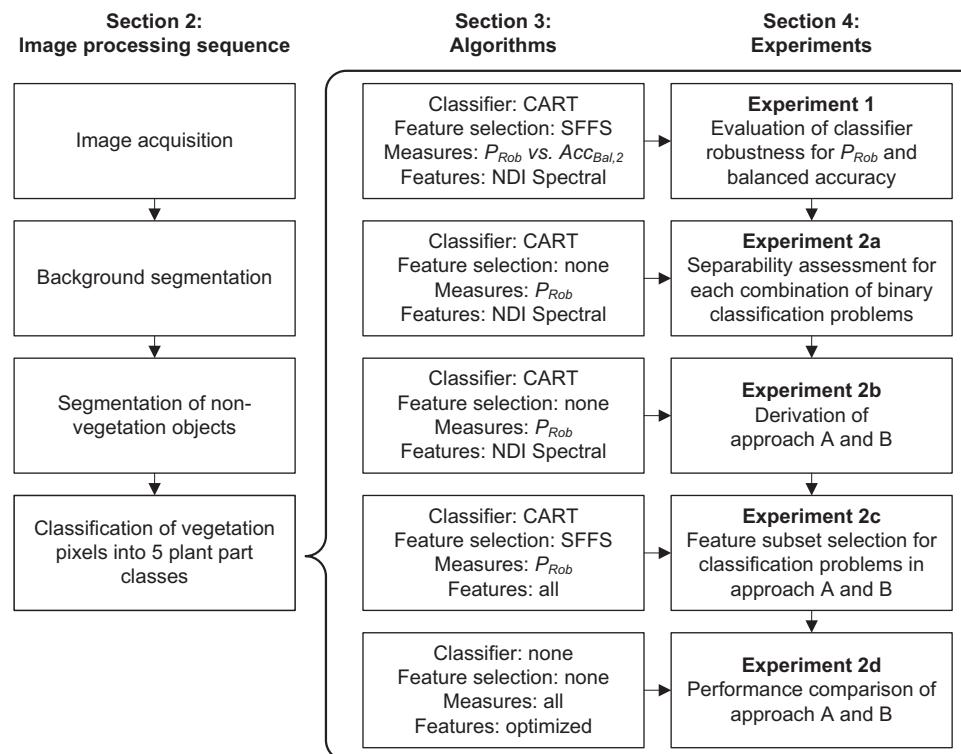
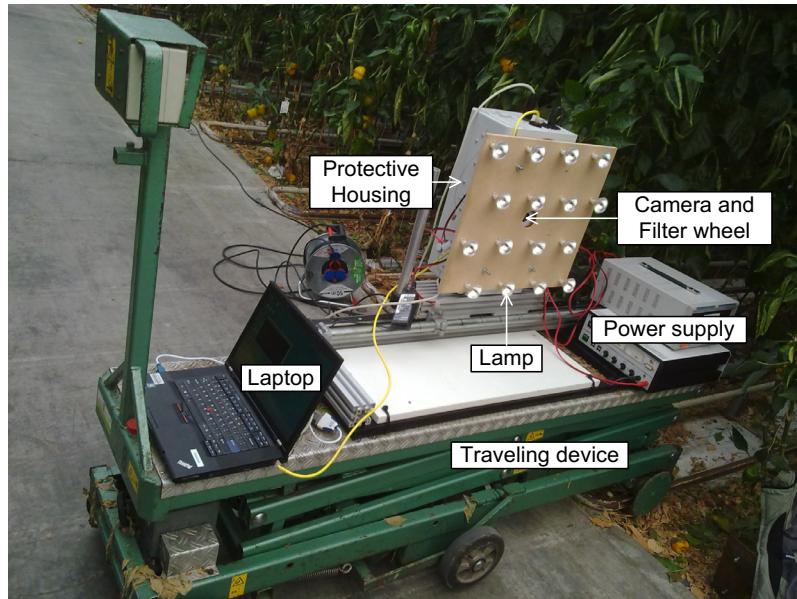


Fig. 1. Overview of the image processing sequence and of algorithms used for each experiment.



**Fig. 2.** Overview of the experimental set-up in a greenhouse.

Objectives were to (1) detect plant vegetation in a crop row; (2) segment non-vegetation objects; (3) prune a decision tree and select features such that the classifier is robust to variation among scenes; and (4) classify vegetation primarily into hard and soft obstacles and, secondarily, into stems, top of leaves, bottom of leaves, green fruits and petioles.

The experimental set-up, experiments and new performance measure presented in this article may contribute tools to perform plant part classification under varying lighting conditions.

## 2. Image processing sequence

**Fig. 1** presents an overview of this article. The image processing sequence to classify vegetation is described in Section 2. Algorithms (performance measures, classifier, feature selection algorithm and features) for vegetation classification are described in Section 3. These algorithms were used in Experiments 1 and 2, described in Section 4. Experiment 1 relates to the third objective and Experiment 2 relates to the fourth objective.

The five-class classification problem, in Experiment 2, was split into four binary classification problems because Kavdir and Guyer, 2004 demonstrated that accuracy for binary classification problems was greater than for multi-class classification problems.

### 2.1. Image acquisition

The multispectral imaging set-up developed consisted of a monochrome camera, housing around the camera and a filter wheel (Section 2.1.1) holding six filters (Section 2.1.2). The scene was illuminated by continuous lighting (Section 2.1.3). An overview of the experimental set-up is in **Fig. 2**.

#### 2.1.1. Camera and filter wheel

The monochrome camera used was a 5 megapixel camera with a 2/3" CCD (Manta G504B; Allied Vision Technologies GmbH, Germany). A low-distortion lens with 8 mm focal length (LM8JCM; Kowa GmbH, Germany) was mounted on the camera. Additionally, a filter wheel (Stock No. NT56-658; Edmund Optics Ltd., UK), holding six Ø25 mm filters, was placed in front of the lens. Exposure time was set such that less than 5% of the pixels were overexposed.

After determining this time for each filter, exposure time was set constant during recording.

We assembled a motorized filter wheel because an off-the-shelf motorized filter wheel was too thick (>15 mm) and would occlude the edges of the image. The assembled filter wheel comprised a manually rotatable filter wheel (thickness 6.35 mm) with a stepper motor (QMOT QSH4218; Trinamic Motion Control GmbH & Co. KG, Germany) and a belt. Recording of six images, including wheel rotation, took about 2 s. Misalignment of the image was 0–3 pixels vertically and 0–1 pixel horizontally, with respect to the first image taken in a wheel rotation. We consider this misalignment to be of negligible influence to the results because labelled ground-truth regions were well within the edge of plant parts.

Image processing was performed on a computer with an Intel Core i5 CPU 2.4 GHz Quad core processor with 4 GB memory.

#### 2.1.2. Wavelength selection

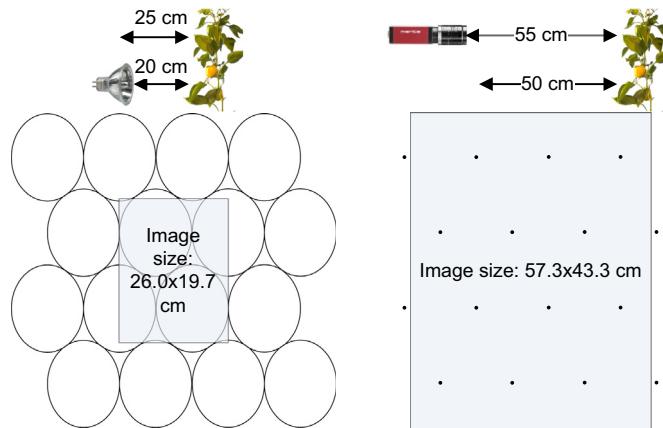
Six wavelengths were selected to distinguish between plant parts. Selection of these wavelengths was based on unpublished work (Fransen, 2011) in which stem, peduncle, fruit, top of a leaf and bottom of a leaf were recorded with a hyperspectral camera – Specim ImSpector V10E (Spectral Imaging Ltd., Finland). Hence, six filters were used (Edmund Optics Ltd., UK):

- 447 nm – bandwidth 60 nm (Stock No. NT48-074).
- 562 nm – bandwidth 40 nm (Stock No. NT48-085).
- 624 nm – bandwidth 40 nm (Stock No. NT48-087).
- 692 nm – bandwidth 40 nm (Stock No. NT48-148).
- 716 nm – bandwidth 40 nm (Stock No. NT67-039).
- >900 nm – longpass (Stock No. NT66-237).

Note that measured bandwidth of the >900 nm longpass filter is limited to 1000 nm due to limited sensitivity of the CCD chip in the camera.

#### 2.1.3. Artificial lighting

For illumination of the scene, 16 halogen lamps (12VDC, 75 W, OSRAM GmbH, Germany) were used. Lamps were placed 5 cm in front of the camera. Each lamp was equipped with a dichroic reflector (Ø51 mm) for equal light distribution. During measurements in a dark room it turned out a lamp emitted light horizon-



**Fig. 3.** Ovals display lighting distribution, at a camera–canopy distance of 25 cm (left). The grey rectangle displays the image size at a camera–canopy distance of 55 cm (right).

tally at an angle of 36° and vertically at an angle of 43°. Based on these measured light emission angles, 16 lamps were distributed over four rows. The horizontal distance between lamps, in a row, was 13 cm. The vertical distance between rows was 13.6 cm. Each row was horizontally shifted (6.5 cm) with respect to the previous row. As a result, equal lighting distribution was ensured for a camera–canopy distance of 25 until 55 cm (Fig. 3). At distances closer than 25 cm, the scene was not equally illuminated. At distances farther than 55 cm, lighting was less intense at edges of the image than at the centre because light spots overlapped at the centre, whereas light spots did not overlap at edges.

## 2.2. Environment and objects

Plants were of the red sweet-pepper cultivar ‘Viper’ and were cultivated in the V-system (Jovicich et al., 2004). Plants (height = 1.5 m) were grown in an experimental greenhouse in Wageningen, The Netherlands. Images were recorded from 11:00 am until 11:30 am under a clear sunny sky. Outdoor solar irradiance was measured during image recording and varied between 374 and 435 W/m<sup>2</sup>. Plant(stem)–camera distance was on average 80 cm and varied in a range of 63–109 cm among scenes. Camera–plant(leaf) distance was mostly at least 25 cm and the lighting (Section 2.1.3) was therefore fit for this distance.

We classified objects, which occurred in the scene, for later use in a motion planning problem (Table 1). Only green unripe fruits occurred in recorded scenes.

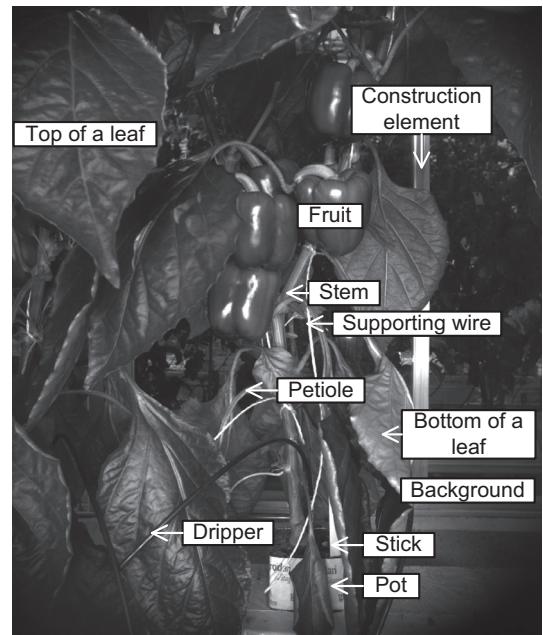
An example of objects occurring in a scene are displayed in Fig. 4.

Construction elements in the scene were aluminium window frames. Other construction elements such as support poles did not occur in the recorded scenes.

**Table 1**

Objects that occurred in recorded scenes. Each object was classified for motion planning by an image processing operation.

Object type	Classified for motion planning as	Operation used for classification	Abbreviation
Objects with distance >1.2 m	Background	Threshold (>900 nm)	Back
Support wire	Hard obstacle	Threshold (447 nm) and area-based segmentation	SW
Stick, dripper and pot	Hard obstacle	Threshold (447 nm) and area-based segmentation	Pot
Construction elements	Hard obstacle	Threshold (447 nm) and area-based segmentation	Const
Stem	Hard obstacle	Pixel-based classification	Stem
Green fruit	Hard obstacle	Pixel-based classification	Fruit
Petiole	Soft obstacle	Pixel-based classification	Pet
Top of a leaf	Soft obstacle	Pixel-based classification	TL
Bottom of a leaf	Soft obstacle	Pixel-based classification	BL



**Fig. 4.** Overview of objects that occurred in an arbitrary scene.

## 2.3. Background segmentation

After image acquisition, the background was segmented using a grey-value threshold, which was empirically determined in HALCON© 10.0.1 (MVTec GmbH, Germany). The threshold was applied on the >900 nm image because the background hardly reflected in this spectral range due to low solar irradiance in the range of 925–975 nm (Center, 2008). In other words, the scene was mainly illuminated by artificial lighting and therefore hardly any light reflected from distances farther than 1.2 m. As a result, the background occurred dark and was easily segmented. Remaining holes in the background were filled with a morphological ‘fill up’ operation.

## 2.4. Segmentation of non-vegetation objects

Objects other than vegetation, i.e. support wire, construction elements, stick, pot and dripper, were removed by a grey-value threshold, which was empirically determined in HALCON©.

A blue wavelength (447 nm) was selected to segment non-vegetation by this threshold because green vegetation hardly reflected, whereas non-vegetation objects strongly reflected. Segmented regions varied in size and an area-based threshold was used to assign small regions to the background and larger regions to hard obstacles. Larger regions were enlarged with a dilation (3 × 3 mask) operation, to include object edges.

### 3. Algorithms

Performance measures, the classifier, the feature selection algorithm and features used for classification of vegetation pixels are described in the following sections.

#### 3.1. Performance measures

Only binary classification problems were performed in this research. Standard performance measures for these problems are described in Section 3.1.1 and a new measure is introduced in Section 3.1.2. After four binary classification problems were performed, five classes remained and five-class performance measures were used to assess overall performance (Section 3.1.3).

##### 3.1.1. Two classes: standard measures

A confusion matrix (Table 2) was used to assess classification performance. Labels 'Object I' and 'Object II' were substituted for combinations of plant part classes 'TL', 'BL', 'Fruit', 'Stem', 'Pet' throughout this article. In addition, label 'Object I' was substituted for 'Hard', which indicates union of two hard obstacle classes (Stem  $\cup$  Fruit), and label 'Object II' was substituted for 'Soft', which indicates union of three soft obstacle classes (TL  $\cup$  BL  $\cup$  Pet).

The confusion matrix was used to evaluate classification performance in terms of: total classification accuracy, true-positive detections for each class, false-positive detections for each class and balanced accuracy (Eqs. (1)–(6)). The number '2' in the symbols refers to a binary problem.

$$Acc2_{Tot} = \frac{100 \cdot TP_I + TP_{II}}{TP_I + TP_{II} + FP_I + FP_{II}} (\%) \quad (1)$$

$$TPR2(I) = \frac{100 \cdot TP_I}{TP_I + FP_{II}} (\%) \quad (2)$$

$$TPR2(II) = \frac{100 \cdot TP_{II}}{TP_{II} + FP_I} (\%) \quad (3)$$

$$FPR2(I) = \frac{100 \cdot FP_I}{FP_I + TP_{II}} (\%) \quad (4)$$

$$FPR2(II) = \frac{100 \cdot FP_{II}}{FP_{II} + TP_I} (\%) \quad (5)$$

$$Acc2_{Bal} = 0.5 \cdot (TPR2(I) + TPR2(II)) (\%) \quad (6)$$

where  $Acc2_{\{\text{Tot}\}}$  (%) total classification accuracy;  $TPR2(I)$  (%) true-positive detection rate of object I;  $TPR2(II)$  (%) true-positive detection rate of object II;  $FPR2(I)$  (%) false-positive detection rate of object I;  $FPR2(II)$  (%) false-positive detection rate of object II;  $Acc2_{Bal}$  (%) balanced accuracy.

In the literature, researchers mostly use total classification accuracy as a performance measure to evaluate algorithms. A

drawback of this measure is that each class needs to be equal in sample size, otherwise classification accuracy will approximate true-positive detection rate of the class containing most samples. A workaround would be to calculate mean true-positive rate for two classes, i.e. balanced accuracy. Still then, such a performance measure does not consider classification accuracy variation among scenes. Classification accuracy variation among scenes can be large due to three reasons: different lighting conditions, different camera-object distances and the varying shape, colour, size, texture and position of objects. As robustness to these three disturbances is currently not assessed, we introduce a new performance measure.

##### 3.1.2. Two classes: new measure

We established a new performance measure *robust-and-balanced accuracy*  $P_{Rob}$ , which compensates for class size differences and considers classification robustness among scenes. This new measure was inspired by the  $d'$  measure, commonly used in signal detection theory (Wickens, 2001) and was applied to distinguish between orange fruits and leaves (Bulanon et al., 2010). In contrast to the  $d'$  measure,  $P_{Rob}$  considers variation among scenes

$$P_{Rob} = \frac{Rob_{Mit} \cdot 0.5 \cdot (M_{TPR2(I)} + M_{TPR2(II)})}{0.5 \cdot (SD_{TPR2(I)} + SD_{TPR2(II)}) + Rob_{Mit}} (-) \quad (7)$$

where  $P_{Rob}$  (-) robust-and-balanced accuracy;  $M_{TPR2(I)}$  (%) mean of multiple scenes with true-positive detection rate of object I;  $M_{TPR2(II)}$  (%) mean of multiple scenes with true-positive detection rate of object II;  $SD_{TPR2(I)}$  (%) standard deviation of multiple scenes<sup>1</sup> with true-positive detection rate of object I;  $SD_{TPR2(II)}$  (%) standard deviation of multiple scenes with true-positive detection rate of object II;  $Rob_{Mit}$  (%) robustness mitigation factor, must be >0. Value was set to 20%.

$P_{Rob}$  is a division of the mean TPR among scenes by its sample standard deviation (SD). An increasing SD will punish  $P_{Rob}$ , whereas an increasing mean TPR will reward  $P_{Rob}$ . We use  $P_{Rob}$  to determine the pruning level of the classifier and to select features.

SD is a measure to quantify the effect of disturbances on consistent classification accuracy among scenes. By incorporating SD, the classifier becomes more robust to disturbances. Sometimes, additional robustness of the classifier may come at a cost of accuracy loss. To vary the weight of robustness vs. accuracy, a robustness mitigation factor  $Rob_{Mit}$  was added. As such,  $P_{Rob}$  has similarity with the coefficient of variation, but weight on mean (accuracy) vs. SD (robustness) can be adjusted through  $Rob_{Mit}$ . The user must set  $Rob_{Mit}$  a priori based on required robustness in his application. A user can calculate  $Rob_{Mit}$  if allowable average standard deviations are known for which performance  $P_{Rob}$  should drop by 50% because, by definition,  $P_{Rob}$  drops by 50% when  $Rob_{Mit} = 0.5 \cdot (SD_{TPR2(I)} + SD_{TPR2(II)})$ . In this application, the exact required robustness of the classifier was unknown. But,  $Rob_{Mit}$  was empirically set to 20 to assure a stronger punishment of  $P_{Rob}$  by an increase of 0.5 · ( $SD_{TPR2(I)} + SD_{TPR2(II)}$ ) than a similar decrease of 0.5 · ( $M_{TPR2(I)} + M_{TPR2(II)}$ ).

##### 3.1.3. Five classes

To assess combined performance of four binary problems, true-and false-positive detection rates, total accuracy and balanced accuracy were calculated for a 5 by 5 confusion matrix. Similar to measures for two classes (Section 3.1.1), total accuracy for five classes  $Acc5_{Tot}$  is the division of the summed confusion matrix diagonal by the sum of all confusion matrix elements. Mean

<sup>1</sup> Note that standard deviation  $SD_{FPR2(I)}$  is equal to  $SD_{TPR2(II)}$ , and similarly  $SD_{FPR2(II)}$  is equal to  $SD_{TPR2(I)}$ .

**Table 2**  
Confusion matrix.

		Actual class	
		Object I	Object II
Classified class	Object I	$TP_I$	$FP_I$
	Object II	$FP_{II}$	$TP_{II}$

Where  $TP_I$  (-) true-positive detection of object I;  $TP_{II}$  (-) true-positive detection of object II;  $FP_I$  (-) false-positive detection of object I;  $FP_{II}$  (-) false-positive detection of object II.

true-positive detection rates, taken over multiple scenes, of the stem ( $M_{TPRS(Stem)}$ ), top of leaf ( $M_{TPRS(TL)}$ ), bottom of a leaf ( $M_{TPRS(BL)}$ ), fruit ( $M_{TPRS(Fruit)}$ ) and petiole ( $M_{TPRS(Pet)}$ ) are calculated as the ratio between correctly classified pixels of a class and the total actual pixels of that class. Balanced accuracy  $Acc5_{Bal}$  is the average of these five mean true-positive detection rates. Each class includes four false-positive detection rates, i.e. 20 values in total.

### 3.2. Classifier

Three requirements resulted in the choice for a decision tree classifier: (1) the classifier must consume little computational load because the application requires an overall harvesting cycle time of 6 s per fruit; (2) the classifier must handle data which are non-normally distributed because Q-Q plots of the grey-value distributions per class and per image revealed that data were non-normally distributed; and (3) the classifier should be easily transferable to another programming language. Non-parametric classifiers, i.e. decision tree, artificial neural network and  $k$ -nearest neighbour, do not assume a mathematical distribution and are preferred. Linker et al. (2012) selected a non-parametric classifier (nearest neighbor) for similar reasons. A decision tree requires less computational load than an artificial neural network and decision trees can be easily implemented in any programming language by a set of if-then-else statements. Hence, a Classification And Regression Trees (CART) classifier (Breiman et al., 1984) was implemented in MATLAB® 2007b.

Moisen (2008) explains basics of CART and discusses which properties should be set for CART training. The Gini Diversity Index (GDI) is a commonly used splitting criterion and was used in this research as well because other splitting criteria (twoing and deviance reduction) performed equally or worse. After training, the tree was pruned to avoid overfitting (Section 3.2.1). Other relevant parameters for decision tree training are initial probabilities and weights (Section 3.2.2) and equalization of the training samples in a class (Section 3.2.3).

#### 3.2.1. Pruning

As Moisen (2008) points out, pruning can be performed by  $n$ -fold cross-validation, the 1-SE rule, or maximizing performance on an independent test set. We optimized pruning on an independent test set. To evaluate effects of the new performance measure,  $P_{Rob}$  and  $Acc2_{Bal}$  were each used to prune the decision tree

$$prune_{opt} = \max(P_{Rob} \text{ or } Acc2_{Bal})(-) \quad (8)$$

where  $prune_{opt}$  (-) optimal prune level, which determines tree length.

Note that performance is always reported for testing data, i.e. the independent test set, throughout the article.

#### 3.2.2. Initial probabilities and weights

Initial probabilities set the likeliness an arbitrary pixel belongs to a certain class. We set initial probabilities to equal values because it turned out that for other probabilities large classes were favoured over small classes and, as a result, performance  $P_{Rob}$  decreased.

Weights are another option to influence decision tree training. A weight  $w_{i,j}$ , for  $i \neq j$ , is the cost of pixel of class  $i$  to be misclassified as class  $j$ . In a brute force simulation of all possible weight combinations it turned out that weights, other than equal weights, reduced total classification accuracy. We therefore set weights to an equal value ( $w_{i,j} = 1$ ).

#### 3.2.3. Equalization of training set size per class

CART has the tendency to favour classes with most training samples (Cieslak and Chawla, 2008). Training sets were therefore reduced to the size of the smallest training class such that training sets were balanced.

### 3.3. Feature selection algorithm

As Blum and Langley (1997) indicated, two critical issues in machine learning are training data selection and feature selection. Training samples were abundantly available and therefore efforts were put on feature selection only. A suboptimal Sequential Forward Floating Selection (SFFS) feature selection method was chosen. This 'floating search method' is, compared with the optimal branch-and-bound method, less computationally intensive, able to handle problems that do not require the monotonicity condition and approaches optimality closely (Jain, 1997). The best feature set is selected based on classification performance in terms of  $P_{Rob}$  or  $Acc2_{Bal}$ . The SFFS method was also used for classification of phalaenopsis stem and root (Huang and Lee, 2008) and for apple stem and calyx detection (Unay et al., 2006). An improved version of the SFFS method has been developed as well (Nakariyakul and Casasent, 2009), but its computational load is at least four times greater than the SFFS method. Given that feature selection with SFFS already lasted three days per binary problem, we used the SFFS method for feature selection.

We also tested Principal Component Analysis (PCA), but results of PCA are not reported because the first four components (98% of correlation) performed worse than feature selection applied to features in Section 3.4.

### 3.4. Pixel-based features

An overview of features used is given in the following sections. In total, 46 features ( $6 + 6 + 15 + 15 + 2 + 2$ ) were calculated for each pixel. Except for entropy, features were calculated in MATLAB®.

#### 3.4.1. Raw data

Merely the reflectance at six wavelengths, i.e. six intensity values per pixel.

#### 3.4.2. Entropy

An entropy operation (mask size =  $3 \times 3$ ) was applied on the raw data using HALCON® and yielded six values per pixel. Entropy is a method to calculate texture within the specified mask. The relative frequencies of grey-values within the mask is used as an input to calculate entropy of the centre pixel in the mask.

#### 3.4.3. Normalized Difference Index (NDI)

The Normalized Difference Index (NDI) is a measure which describes the normalized difference between two different wavelength channels (Davies, 2009). An advantage of this index is that lighting influences are reduced through a division by two channels. NDI is commonly applied to raw data but here we tested its effect on entropy as well. In total 15 possible combinations of two wavelengths from a set of six wavelengths exist. That is, applied for raw data and entropy, 30 values were calculated per pixel.

#### 3.4.4. Spectral Angle Mapper (SAM)

The spectral angle mapper is a commonly used feature for hyperspectral or multi-spectral analysis. It measures similarity of an image pixel spectrum with respect to a reference spectrum (Yang et al., 2008). One spectral angle was calculated per class, i.e. each pixel yielded two values.

### 3.4.5. Mahalanobis distance

The Mahalanobis distance algorithm is an improvement of the commonly known Euclidean distance algorithm because correlations of wavelengths are considered (DeVries, 2005).

An average intensity vector was calculated from labelled ground-truth regions in the training image. As for the SAM, each pixel yielded two values: one value per class.

## 4. Experiments

Ground truth labelling, assigning training and testing data and Experiments 1 and 2 are explained in the following sections.

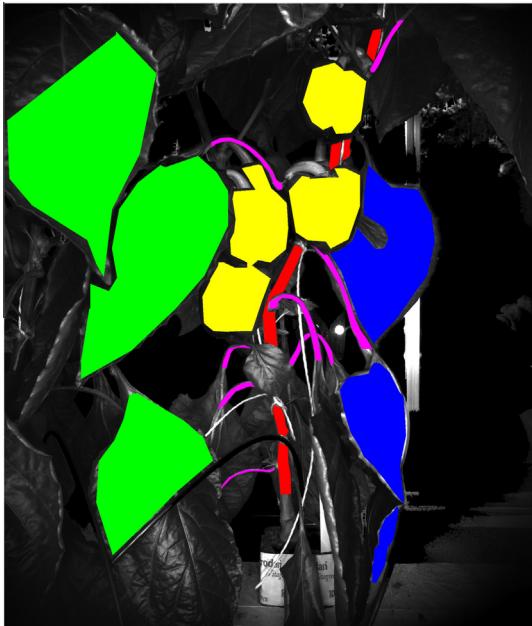
### 4.1. Ground truth labelling

In total, 12 scenes were recorded using the experimental set-up, one plant per scene. Ground truth labelling was manually performed, using HALCON®, for the five plant parts to be classified. A result of labelling is in Fig. 5.

In total,  $14.6 \times 10^6$  pixels were labelled and this number comprised 29.4% of the vegetation present in the 12 scenes. The number of labelled pixels varied among scenes because visibility of plant parts also varied among scenes. Most of labelled pixels were leaves – TL (54.6%), BL (22.4%) – followed by fruits (15.6%), stems (3.7%) and petioles (3.7%). Vegetation pixels which were not labelled (70.6%) were mostly leaves and these were not labelled because we assumed labelled leaves already represented the majority of the leaf variation occurring in the scene. In addition, labelling all samples would increase the computational load dramatically, especially during feature selection.

### 4.2. Training and testing data

We evaluated the number of scenes required for training by evaluating all possible combinations of selecting this number of scenes from the population of 12 scenes (Table 3). We performed hard vs. soft obstacle classification using 15 NDI spectral features



**Fig. 5.** Example of manually drawn ground truth labels for five classes: stem (red), top of a leaf (green), bottom of a leaf (blue), green fruit (yellow) and petiole (purple). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**

Effect of different combinations and number of training scenes on classification performance.

Scenes used for training vs. testing	Combinations evaluated (#)	Mean (SD) $P_{Rob}$ (–) among combinations	Performance increase of mean $P_{Rob}$ (–)
1 vs. 11	12	44.4 (7.7)	
2 vs. 10	66	49.6 (5.2)	5.2
3 vs. 9	220	52.3 (4.1)	2.7
4 vs. 8	495	54.7 (3.4)	2.4

as input. Only NDI spectral features were used because these features were strong and therefore sufficient to assess the size of the training set.

We selected two scenes for training and ten for testing because performance increase was greater (5.2) than the increase to three training scenes (2.7) or four training scenes (2.4). Performance increase for three or four training scenes may have been partially due to the decreasing number of testing scenes. Regarding the combination selected out of 66 possible combinations, a combination that performed on average ( $P_{Rob} = 49.4$ ) was selected to perform Experiments 1 and 2. These two scenes comprised  $2.9 \times 10^6$  labelled pixels; ten scenes for testing comprised  $11.7 \times 10^6$  labelled pixels.

### 4.3. Experiment 1: evaluation of classifier robustness

The effect of the new measure  $P_{Rob}$  vs.  $Acc2_{Bal}$  was separately evaluated in terms of accuracy and consistent classification among scenes, i.e. robustness. Each measure was used for decision tree pruning and for selection of the optimal feature subset, using 15 NDI spectral features. The best measure was  $P_{Rob}$  and was therefore used for pruning and feature selection in Experiment 2.

### 4.4. Experiment 2a: separability assessment for each combination of binary classification problems

As mentioned earlier, the five plant part classes were classified by an approach of four binary classification problems. To define such an approach, the separability of all possible binary combinations were evaluated. In total,  $\binom{5}{2} + \binom{5}{1} = 15$  combinations exist to divide five classes into a binary problem. Hence, unions of 2 and 3 classes (10 combinations) or 1 and 4 classes (5 combinations) were taken.

Similar to the evaluation of training data set sizes (Section 4.2), a fixed set of 15 NDI spectral features was used to calculate the separability for each of the 15 combinations. Feature selection was therefore only performed in Experiment 2c to avoid a computation time of 45 days (3 days \* 15 combinations). And, we doubt whether feature selection would lead to a different best separable combination.

### 4.5. Experiment 2b: derivation of Approach A and B

There are many approaches possible to reduce a 5-class classification problem into four binary classification problems, based on results from Experiment 2a. We investigated two different approaches. Firstly, Approach A was inspired by the application. Here, hard (Stem  $\cup$  Fruit) and soft (TL  $\cup$  BL  $\cup$  Pet) obstacles were classified first. Secondly, Approach B was inspired by the separability of classes. Here, the best separable classes, identified in Experiment 2a, were classified first. Once both approaches were defined, their four binary classification problems were optimized using feature selection (Section 4.6).

#### 4.6. Experiment 2c: feature subset selection

Feature selection was performed for each binary classification problem of Approach A and B, using the feature selection algorithm and pixel-based features described in Section 3. To reduce the computational load of feature selection for a set of 46 features, feature selection was first performed per type of feature (Section 3.4). Secondly, these selected sets (7) were clustered in a ‘Cluster’ set for which feature selection was applied.

#### 4.7. Experiment 2d: performance comparison of Approach A and B

Approach A and B were compared in terms of robust-and-balanced accuracy of hard and soft obstacle classification and in terms of 5-class performance measures. Hard and soft obstacle classification performance is more critical for motion planning than the classification performance for five classes. The best approach was therefore chosen based on robust-and-balanced accuracy of hard and soft obstacle classification.

### 5. Results

#### 5.1. Background segmentation

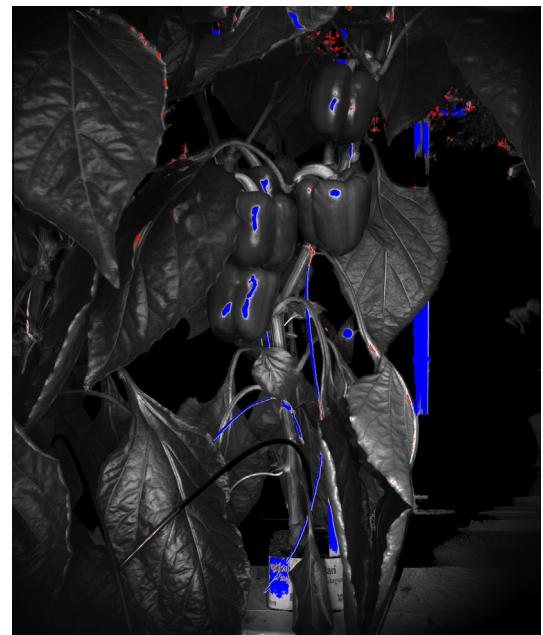
To segment the background from other objects in the scene, a threshold was set to  $\leq 27$  ( $>900$  nm image; 8 bit). Subsequently holes were filled (Fig. 6). Calculation took 7 ms.

Vegetation area decreased by 2.7% if the grey-value threshold was set to 30, i.e. an increase of 11%. Hence, the threshold value has an effect on classifiable vegetation area and had to be selected carefully.

#### 5.2. Segmentation of non-vegetation objects

To segment all non-vegetation objects, a threshold was set to  $>139$  (447 nm image; 8 bit). Segmented regions with an area larger than or equal to 300 pixels included construction elements, stick, pot, and supporting wires. These hard obstacles were therefore removed before classification of vegetation. Segmented regions with an area smaller than 300 pixels were assigned to the background class (Fig. 7). Calculation took 6 ms.

Some vegetation pixels, about 1% of total vegetation area, were also part of the segmented regions (Fig. 7) and were therefore not classified. These vegetation pixels were mainly fruits and because fruits are hard obstacles it is not problematic that segmented regions included some vegetation pixels. These fruit segments were



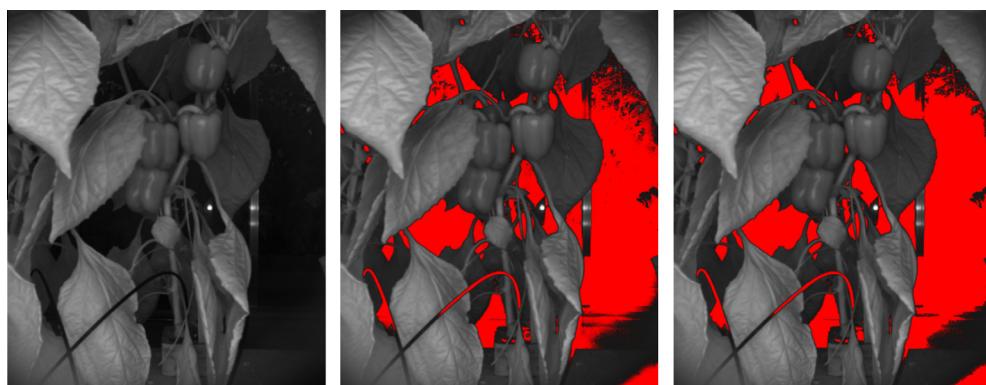
**Fig. 7.** Segmented regions, mostly non-vegetation objects, with an area smaller than 300 pixels (red) and segmented regions with an area larger than or equal to 300 pixels (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

overexposed due to direct light reflection on the glossy fruit surface.

The pot, support wire, construction elements and stick were removed by this threshold operation. Drippers were not detected due to their black colour and use of white drippers may be a solution to detect drippers as a hard obstacle. Nevertheless drippers were classified as background (Fig. 6) and a scene remained that contained only the five plant parts to be classified. All non-vegetation objects were removed from the scene.

#### 5.3. Experiment 1: evaluation of classifier robustness

Due to the use of the new performance measure  $P_{Rob}$ , we observed that feature selection resulted in different optimal feature sets compared with feature selection based on only balanced accuracy as measure. As an example, Table 4 shows that accuracy is slightly greater for the balanced accuracy performance measure



**Fig. 6.** The background of the image (left) was segmented (centre) from other objects in the scene using a grey value threshold ( $\leq 27$ ). Subsequently holes were filled (right). Unfortunately, part of the dripper is classified as background whereas it should have been classified as a hard obstacle.

**Table 4**

Comparison between balanced accuracy and robust-and-balanced accuracy as performance measure for decision tree pruning and feature selection. Feature selection was applied to NDI spectral features for classification problem A1: hard vs. soft obstacle classification.

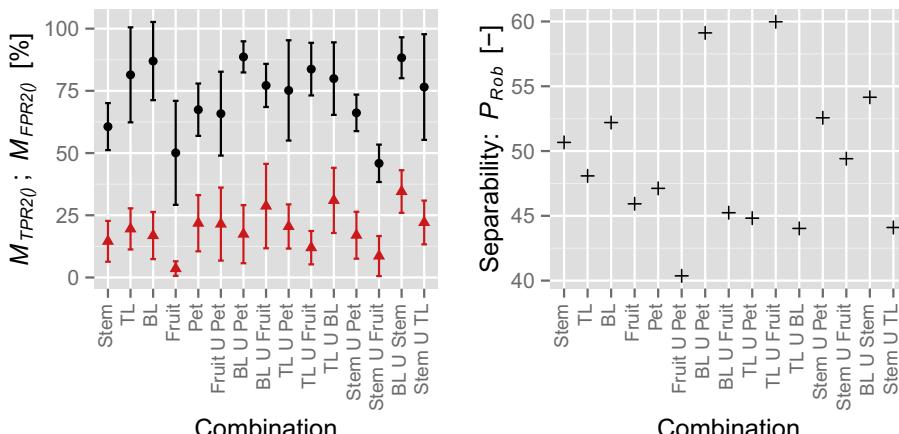
	Performance measure used	
	Balanced accuracy: $Acc2_{Bal}$	Robust-and- balanced accuracy: $P_{Rob}$
Features (NDI spectral) in the pruned decision tree; ordered on occurrence.	562&900; 692&716; 692&900; 562&716; 624&692; 562&624	447&624; 624&900; 692&716; 562&624
Tree length before pruning (nodes)	9427	9287
Tree length after pruning (nodes)	67	9
Balanced accuracy $Acc2_{Bal}$ (%)	77.1	75.4
Robust-and-balanced accuracy $P_{Rob}$ (-)	47.7	58.9
$Acc2_{Tot}$ (SD) (%)	84.2 (6.0)	84.5 (5.0)
$M_{TPR2(hard)}(SD_{TPR2(hard)})$ (%)	66.5 (17.2)	59.2 (7.1)
$M_{TPR2(soft)}(SD_{TPR2(soft)})$ (%)	87.4 (7.0)	91.5 (4.0)

(77.1) than for the robust-and-balanced accuracy performance measure (75.4). But, for balanced accuracy, the standard deviation on the true-positive-rates is about two times greater. Hence, the classifier is more robust to variations among scenes when  $P_{Rob}$  was used instead of balanced accuracy.

Table 4 shows that, for  $P_{Rob}$ , two features (447&624, 624&900) were selected that were not selected when  $Acc2_{Bal}$  was taken as performance measure. One may conclude these two features are less sensitive to influences of lighting variations and plant-camera distance variations.

#### 5.4. Experiment 2a: separability assessment for each combination of binary classification problems

Fig. 8 demonstrates the separability of the 15 combinations possible. Union of top of a leaf (TL) and fruit is the best separable combination ( $P_{Rob} = 60.0$ ). This combination was therefore classified first in Approach B. Separability of hard (Stem  $\cup$  Fruit) vs. soft (TL  $\cup$  BL  $\cup$  Pet) obstacles ( $P_{Rob} = 49.4$ ) is average compared with the other 15 combinations.



**Fig. 8.** Separability for 15 combinations, each combination represents a binary classification problem. Classes in each combination were taken vs. all remaining classes. Mean and SD ( $N = 10$ ) of true-positive detection rates  $M_{TPR20}$  (●) and false-positive detection rates  $M_{FPR20}$  (▲) (left) were used as input to calculate the separability performance measure  $P_{Rob}$  (+) (right). Combination 'TL  $\cup$  Fruit' ( $P_{Rob} = 60.0$ ) is best separable.

#### 5.5. Experiment 2b: derivation of Approach A and B

An overview of Approach A is shown in Fig. 9. This approach was based on relevance for the application, i.e. hard vs. soft obstacle classification first.

Three different options were available to divide three classes for classification problem A3. In a classification where the 15 NDI spectral features were fed to the classifier, it turned out that TL and union of BL and Pet ( $P_{Rob} = 57.7$ ) resulted in better separability than BL and union of TL and Pet ( $P_{Rob} = 49.1$ ) or Pet and union of BL and TL ( $P_{Rob} = 45.8$ ). Consequently, TL and union of BL and Pet were taken as classes in classification problem A3 and finally BL and Pet were classified in problem A4.

An overview of Approach B is shown in Fig. 10. This approach was based on separability of classes, i.e. the easiest classification problem first.

Three different options were available to divide three classes for problem B3. In a classification where the 15 NDI spectral features were fed to the classifier, it turned out that stem and union of BL and Pet ( $P_{Rob} = 53.3$ ) resulted in better separability than Pet and union of Stem and BL ( $P_{Rob} = 41.3$ ) or BL and union of Stem and Pet ( $P_{Rob} = 35.0$ ). Consequently, Stem and union of BL and Pet were taken as classes in problem B3 and finally BL and Pet were classified in problem B4.

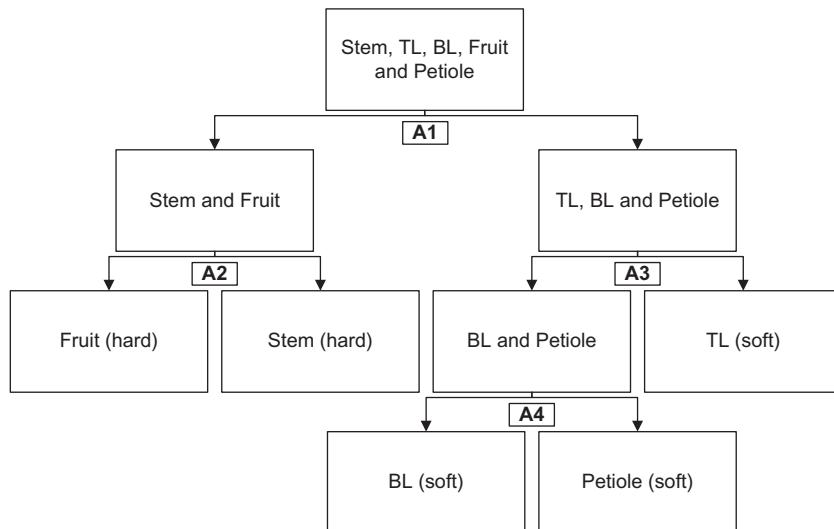
Section 5.6 describes the result of feature selection on each of the four binary problems, for Approach A (A1–A4) and for Approach B (B1–B4).

#### 5.6. Experiment 2c: feature subset selection

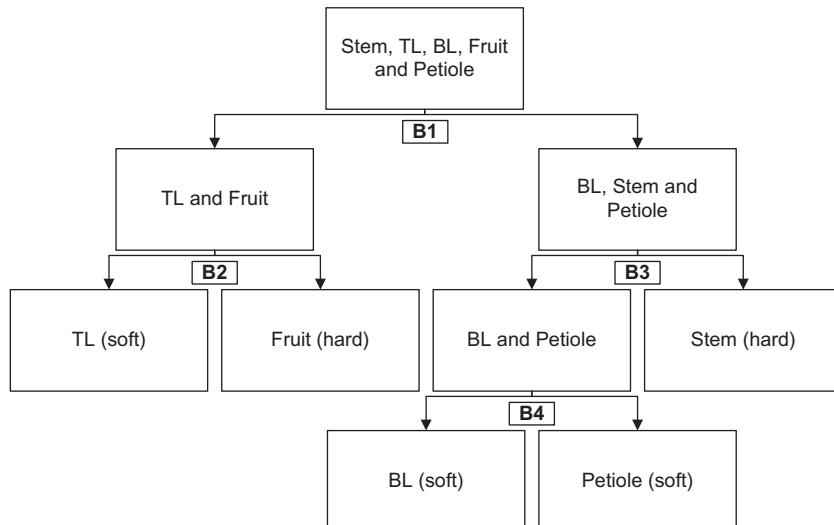
To assess feature strength and the effectiveness of feature selection, we show detailed results of classification problem A1. Such detailed results are shown only for A1 to avoid repetition of similar findings in other classification problems.

Fig. 11 demonstrates the result of feature selection for features listed in Section 3.4. Details of feature selection and performance are in Table 5. Only selected raw spectral (3), raw entropy (4), NDI spectral (4) and NDI entropy (3) were combined in the 'Cluster' set (Fig. 11; Table 5). SAM and Mahalanobis distance were not added to the 'Cluster' feature set because, due to their poor performance, deteriorated performance of the 'Cluster' set. Furthermore, they were not investigated in successive classification problems.

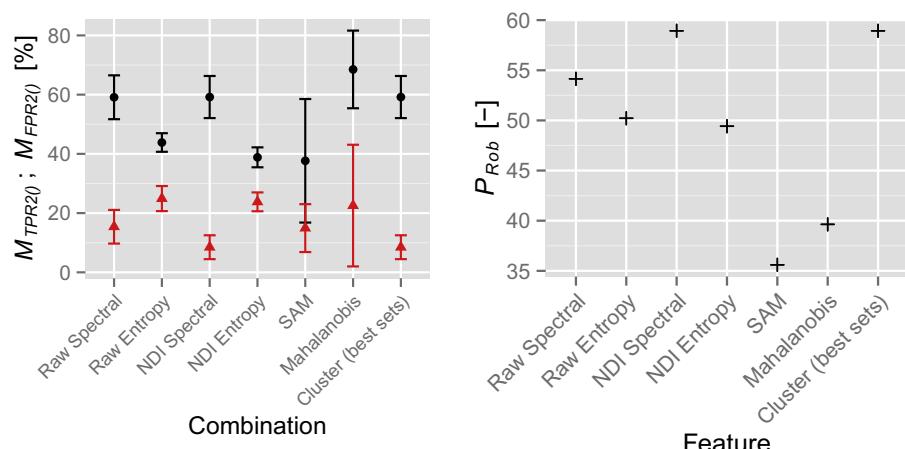
Fusion of the best four selected sets did not result in better performance. The NDI spectral features were again selected from the



**Fig. 9.** Approach A to reduce a five-class classification problem into four binary problems. The classification sequence is based on the relevance to the application, i.e. hard vs. soft obstacle classification first.



**Fig. 10.** Approach B to reduce a five-class classification problem into four binary problems. The classification sequence is based on separability of the classes, i.e. best separable first.

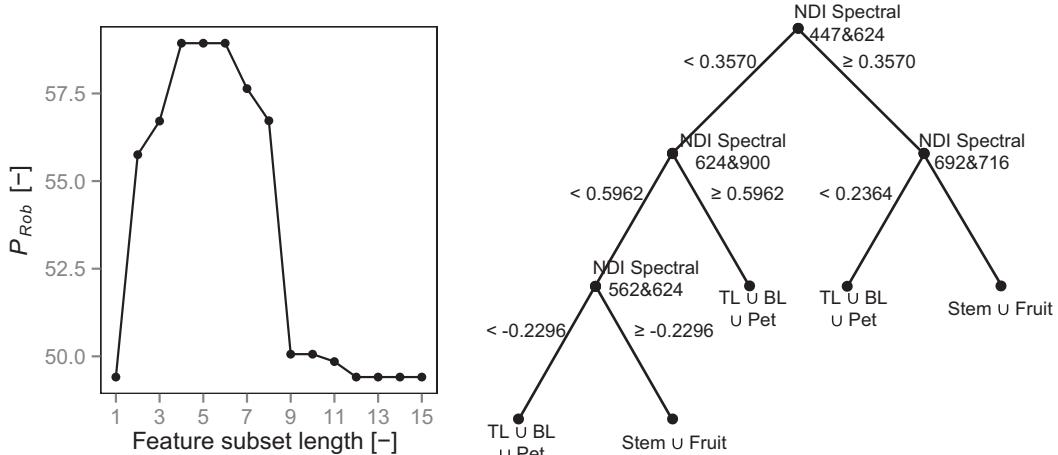


**Fig. 11.** Performance per feature after feature selection, using the SFFS algorithm. Results are for classification problem A1: hard vs. soft obstacles. Mean and SD ( $N = 10$ ) of true-positive detection rates  $M_{TPR20}$  (●) and false-positive detection rates  $M_{FPR20}$  (▲) per feature (left) were used to calculate the performance measure  $P_{Rob}$  (+) per feature (right).

**Table 5**

Classification performance results and classifier characteristics per feature type, for problem A1 (hard vs. soft obstacles).

	Raw spectral	Raw entropy	NDI spectral	NDI entropy	SAM	Mahalanobis distance	Cluster (best sets)
Features before selection (#)	6	6	15	15	2	2	14
Feature subset length after selection (−)/(features)	3/447; 624; 716	4/447; 562; 624; 692	4/447&624; 624&900; 692&716; 562&624; 624&900	3/447&562; 624&716; 624&900	2/SAM (hard); SAM (soft)	1/Mahalanobis (hard)	4/NDI spectral: 447&624; 624&900; 692&716; 562&624;
Tree length before pruning (nodes)	4469	13,247	9287	32,995	39,645	60,823	9287
Tree length after pruning (nodes)	105	27	9	19	3	3	9
$P_{Rob}$ (−)	54.1	50.2	58.9	49.4	35.6	39.6	58.9
$Acc_{Tot,2}$ (SD)	79.0 (4.8)	68.4 (6.0)	84.5 (5.0)	68.2 (5.5)	75.7 (10.0)	75.6 (15.0)	84.5 (5.0)
$M_{TPR2(hard)}(SD_{TPR2(hard)})$ (%)	59.1 (7.4)	43.8 (3.1)	59.2 (7.1)	38.8 (3.4)	37.7 (20.9)	68.5 (13.1)	59.2 (7.1)
$M_{TPR2(soft)}(SD_{TPR2(soft)})$ (%)	84.6 (5.7)	75.1 (4.2)	91.5 (4.0)	76.2 (3.2)	85.1 (8.1)	77.5 (20.5)	91.5 (4.0)

**Fig. 12.** Classification performance of optimized feature subset lengths by SFFS, for NDI spectral features (left). Results are for classification problem A1: hard vs. soft obstacles. The decision tree for the optimal feature subset with length = 4 (right).

'Cluster' set ( $P_{Rob} = 58.9$ ). Despite the suggested potential of texture features (entropy) in the literature (Kapach et al., 2012), the best set only contained NDI spectral features.

In general, feature selection improved classification results for all feature types. As an example, Fig. 12 (left) demonstrates how the feature subset strongly influences classification performance.

Clearly, the feature subset is optimal at feature subset length = 4. For lengths 5 and 6, CART is still able to find the optimized tree. Once the feature subset becomes longer than 6 features, performance decreases. This process of decreasing performance is caused by the suboptimal nature of classifiers and aggravates with high-dimensional feature sets, an effect which is often observed in machine learning (Jain, 1997).

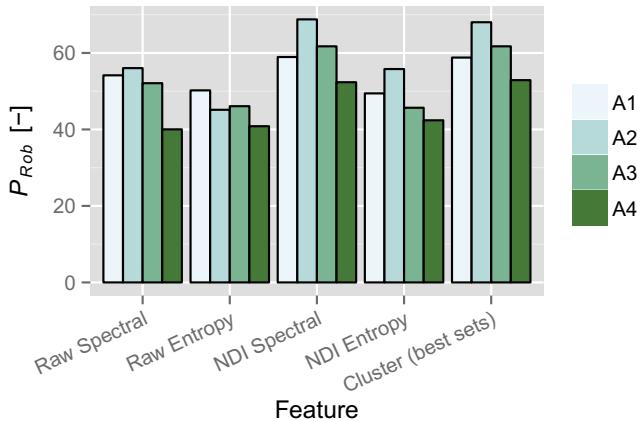
#### 5.6.1. Performance of A1–A4

Fig. 11 already shows how means and standard deviations relate to the performance measure  $P_{Rob}$ . Therefore, only the

performance  $P_{Rob}$  was shown in Fig. 13, for the most useful features per classification problem.

Classification problem A2 (stem vs. fruit) was clearly the easiest among the classification problems. Apparently the stem and fruit are easier to separate than other classes. Especially problem A4 is hard (BL vs. Pet).

In general, NDI spectral features were the strongest features in all classification problems. Features are given in parenthesis and are ordered on occurrence in the pruned tree: A1 comprised 4 NDI spectral features (447&624, 624&900, 692&716, 562&624); A2 comprised 4 NDI spectral features (562&624, 692&716, 447&692, 447&716); A3 comprised 5 NDI spectral features (692&900, 447&562, 624&716, 447&900, 562&716); A4 comprised 4 NDI spectral features (692&716, 562&900, 562&624, 624&900) and 1 raw spectral feature (624). All six wavelengths were used. Texture features were never selected and one can conclude that texture features did not contribute relevant information in Approach A.



**Fig. 13.** Performance of classification problems A1 ( $\text{Stem} \cup \text{Fruit}$  vs.  $\text{TL} \cup \text{BL} \cup \text{Pet}$ ), A2 ( $\text{Stem}$  vs.  $\text{Fruit}$ ), A3 ( $\text{TL}$  vs.  $\text{BL} \cup \text{Pet}$ ) and A4 ( $\text{BL}$  vs.  $\text{Pet}$ ) per feature. NDI Spectral was the strongest feature in A1, A2 and A3. The combination of the best sets was the strongest feature in A4. A2 was the easiest classification problem.

#### 5.6.2. Performance of B1–B4

The performance  $P_{Rob}$ , for the most useful features per classification problem, is shown in Fig. 14.

Classification problem B1 ( $\text{TL} \cup \text{Fruit}$  vs.  $\text{TL} \cup \text{BL} \cup \text{Pet}$ ) was clearly the easiest among the classification problems.

In general, NDI spectral features were the strongest features in all classification problems. Features are given in parenthesis and are ordered on occurrence in the pruned tree: B1 comprised 4 NDI spectral features (562&900, 624&716, 447&624, 562&624); B2 comprised 3 NDI spectral features (447&624, 447&900, 692&716); B3 comprised 4 NDI spectral features (692&716, 716&900, 562&624, 447&900) and three raw spectral features (716, 900, 692); B4 comprised 4 NDI spectral features (692&716, 562&900, 562&624, 624&900) and 1 raw spectral feature (624). All six wavelengths were used. Texture features were not selected in Approach B and one can conclude that texture features did not contribute relevant information in Approach B.

#### 5.7. Experiment 2d: comparison of Approach A and B

The choice for the best approach was based on performance of hard vs. soft obstacle classification (Section 5.7.1). Furthermore,

**Table 6**  
Classification performance of hard vs. soft obstacle classification by Approach A and B.

Performance measure	Approach A	Approach B
$M_{TPR2(\text{hard})}(SD_{TPR2(\text{hard})}) (\%)$	59.2 (7.1)	61.6 (8.2)
$M_{TPR2(\text{soft})}(SD_{TPR2(\text{soft})}) (\%)$	91.5 (4.0)	89.4 (5.6)
Robust-and-balanced accuracy $P_{Rob}$ (-)	58.9	56.1
$Acc2_{\text{Tot}}$ (SD) (%)	84.5 (5.0)	83.1 (6.3)

Section 5.7.2 describes the classification performance among the five plant part classes.

#### 5.7.1. Performance of hard vs. soft obstacle

The performance of hard ( $\text{Stem} \cup \text{Fruit}$ ) vs. soft ( $\text{TL} \cup \text{BL} \cup \text{Pet}$ ) obstacle classification are in Table 6.

Robust-and-balanced accuracy was greater for Approach A ( $P_{Rob} = 58.9$ ) than for Approach B ( $P_{Rob} = 56.1$ ), which is caused by lower standard deviations. Hence, Approach A resulted in a slightly better performance.

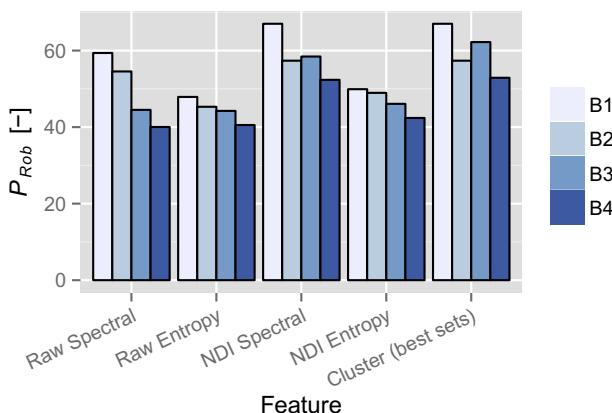
#### 5.7.2. Performance of five plant part classes

Fig. 15 displays classification performance per class, for both approaches.

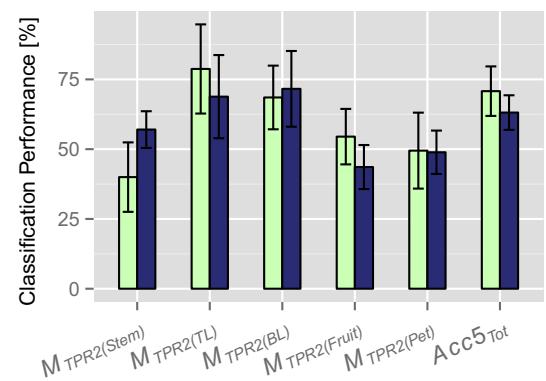
Approach A and Approach B do not seem to differ strongly in performance. For Approach A, total accuracy (70.8) is greater than Approach B (63.1%) and one may conclude that Approach A is better. But, TPRs differ strongly per class. For instance,  $M_{TPR5(\text{Stem})}$  is greater in Approach B (57.0%) than Approach A (40.0%). The average of mean TPRs,  $Acc5_{\text{Bal}}$ , is similar for Approach A (58.2%) and Approach B (58.0%).

Moreover, Fig. 15 clearly shows low TPRs for stem (<59%), fruit (<55%) and petiole (<50%). Such low detection rates are not problematic as long as FPRs are also low. Fig. 16 therefore displays how labelled ground truth regions were incorrectly classified.

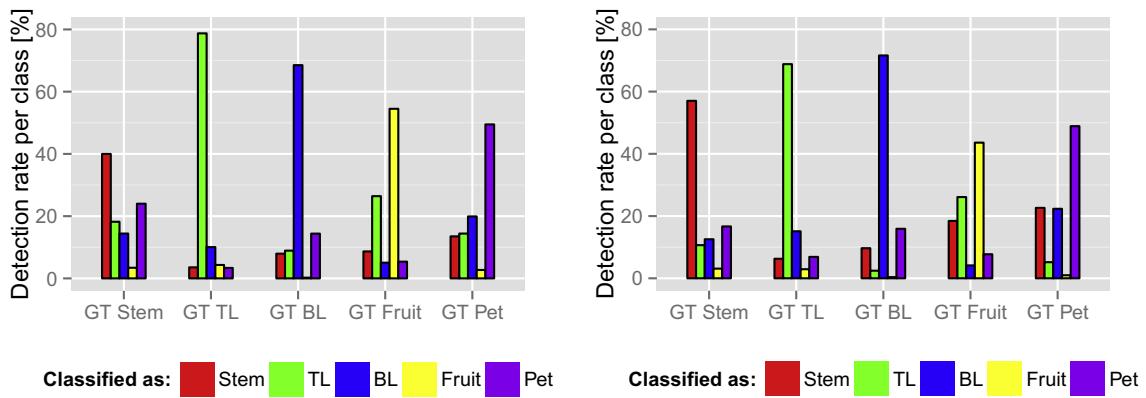
For both approaches, the highest detection rate of a class occurred in the corresponding labelled class. This finding indicates that post-processing can be promising because false-positive detections might be removed by morphological image processing such as opening or erosion. In conclusion, Fig. 16 shows in which plant parts false-positive detections occur and, as a result, a strategy for post-processing can be established. For instance, many stem and petiole pixels were falsely assigned to TL and BL classes and one might consider post-processing techniques that convert false TL and BL detections into true stem and petiole detections.



**Fig. 14.** Performance of classification problems B1 ( $\text{TL} \cup \text{Fruit}$  vs.  $\text{Stem} \cup \text{BL} \cup \text{Pet}$ ), B2 ( $\text{TL}$  vs.  $\text{Fruit}$ ), B3 ( $\text{Stem}$  vs.  $\text{BL} \cup \text{Pet}$ ) and B4 ( $\text{BL}$  vs.  $\text{Pet}$ ) per feature. NDI Spectral was the strongest feature in B1 and B2. The combination of the best sets was the strongest feature in B3 and B4. B1 was the easiest classification problem.



**Fig. 15.** Classification performance of Stem, top of a leaf (TL), bottom of a leaf (BL), Fruit and Petiole (Pet) for Approach A (light green) and B (dark blue). Performance shown in terms of mean and SD ( $N = 10$ ) of true-positive detection rate  $M_{TPR5}$  and total accuracy  $Acc5_{\text{Tot}}$ . Total accuracy for Approach A (70.8%) is greater than for Approach B (63.1%).

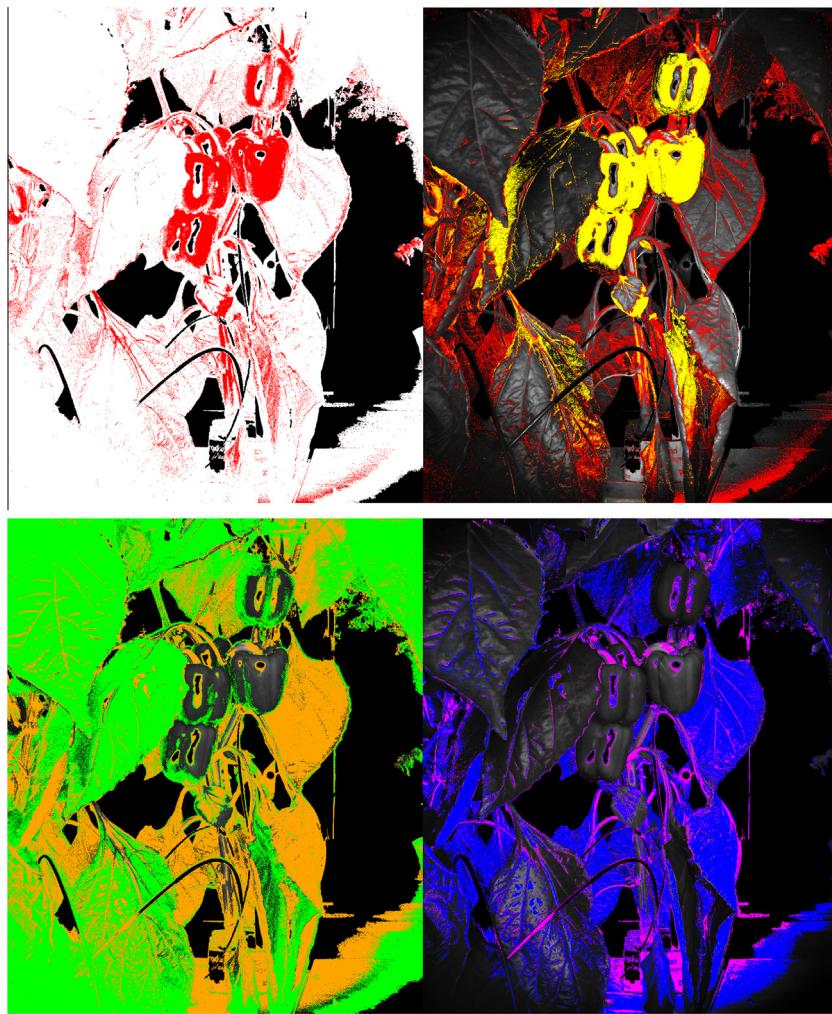


**Fig. 16.** Labelled ground truth (GT) regions were classified into 5 classes by Approach A (left) and Approach B (right). In both approaches, the highest mean ( $N = 10$ ) detection rate of a class occurred in the corresponding labelled class. For instance, stem detection was highest in GT Stem.

### 5.8. Example of a classified scene

Results of classification by Approach A are in Figs. 17 and 18.

Total execution time for one scene in HALCON® was 2.3 s: 1.5 s for calculation of 12 NDI spectral features, 0.23 s for A1, 0.14 s for A2, 0.18 s for A3 and 0.29 s for A4.



**Fig. 17.** Result of pixel-based classification by Approach A. A1 (top-left) represents hard (red) vs. soft obstacles (white). A2 (top-right) represents stem (red) vs. fruit (yellow). A3 (bottom-left) represents top of a leaf (green) vs. bottom of a leaf  $\cup$  petiole (orange). A4 (bottom-right) represents bottom of a leaf (blue) vs. petiole (magenta).



**Fig. 18.** Full classification of vegetation into five classes: stem (red), top of a leaf (green), bottom of a leaf (blue), fruit (yellow) and petiole (magenta). Black parts are either background or segmented non-vegetation objects.

Fig. 18 demonstrates that many false-positive detections occur in the scene.

## 6. Discussion

Classification performance was too low to build a reliable obstacle map for motion planning. Mean true-positive detection rate remains limited to 59.1% for hard obstacles and 91.5% for soft obstacles. As a result, 8.5% of the soft obstacles are falsely classified as hard obstacle and these false detections block a collision-free path for the harvesting manipulator. We were not able to remove these false detections, by a morphological ‘opening’ operation (Bac et al., 2013), without removing less than 50% of true-positive detections. An option to investigate in future work is to fuse Approach A and B in a judges-based approach.

Classification performance of related work conducted under varying lighting conditions is unknown and we only found performance of a study conducted indoor, under controlled lighting conditions. Humphries and Simonton (1993) performed pixel-based classification of geranium plant parts and report true-positive detection rates of 85% for leaves, 74% for stems and 21% for petioles. These slightly greater detection rates are probably due to controlled lighting conditions, and perhaps more constant camera-object distances. Humphries and Simonton (1993) improved detection rates to 97% for leaves, 95% for stems and 93% for petioles after addition of object-based features. Such an increase after addition of object-based features is also demonstrated by others (Wang et al., 2004) and addition of object-based features is therefore a task for future work.

Two causes may elucidate why classification performance was low. Firstly, the strong camera-object distance variation among scenes and within scenes resulted in illumination differences and, as a result, misclassifications. For instance, leaves located in front of a stem mainly received artificial illumination, whereas leaves behind a stem were further away and received a mixture of artificial illumination and natural lighting. Secondly, natural lighting varied (374–435 W/m<sup>2</sup>) and, in addition, light incidence

on objects varied because light penetration is influenced by the varying shape and size of plant parts and of non-vegetation objects. Effects of both issues on classification accuracy was not studied and is an objective for future work. Two solutions can be investigated: use of a reference card to compensate for temporal variation in lighting conditions (Ting et al., 2012) and use of distance information to compensate for spatial variation in lighting conditions.

## 7. Conclusion

Unfortunately, classification performance was too low to build a reliable obstacle map for motion planning. Nevertheless, sweet-pepper plant vegetation was successfully segmented from the background using a threshold in a near-infrared wavelength (>900 nm). Subsequently non-vegetation objects, which included drippers, pots, sticks, construction elements and support wires, were removed by a threshold in the blue wavelength (447 nm). Remaining plant vegetation was classified into five classes using a CART decision tree with 46 features. For classification of hard (Stem ∪ Fruit) vs. soft (TL ∪ BL ∪ Pet) obstacles, NDI spectral features ( $P_{Rob} = 58.9$ ) are stronger than raw spectral ( $P_{Rob} = 54.1$ ), raw entropy ( $P_{Rob} = 50.2$ ), NDI entropy ( $P_{Rob} = 49.4$ ), SAM ( $P_{Rob} = 35.6$ ), Mahalanobis distance ( $P_{Rob} = 39.6$ ) or the clustered sets ( $P_{Rob} = 58.9$ ).

Two approaches were derived to reduce the five-class classification problem into four consecutive binary problems: Approach A was based on the problem nature and Approach B was based on separability of classes. Approach A is slightly better than Approach B because, on hard vs. soft obstacle classification, Approach A ( $P_{Rob} = 58.9$ ) is equally accurate but more robust than Approach B ( $P_{Rob} = 56.1$ ).

For Approach A, mean true-positive detection rate (standard deviation) among scenes was 59.2 (7.1)% for hard obstacles, 91.5 (4.0)% for soft obstacles, 40.0 (12.4)% for stems, 78.7 (16.0)% for top of a leaf, 68.5 (11.4)% for bottom of a leaf, 54.5 (9.9)% for fruit and 49.5 (13.6)% for petiole. These performance values are low and post-processing techniques will be investigated in future research. This research also shows benefits of the new performance measure  $P_{Rob}$ , which considers both balanced accuracy and classification variation among scenes. Use of  $P_{Rob}$  rendered the classifier more robust to variation among scenes because standard deviation among scenes reduced 59% for hard obstacles and 43% for soft obstacles compared with balanced accuracy. This improvement is achieved, interestingly, because features probably less sensitive to lighting variations and varying plant–camera distances are selected.

## Acknowledgements

We thank Yael Edan and Efi Vitzrabin for their useful comments and suggestion to improve this manuscript. We thank Anja Dielman who provided the experimental greenhouse in Wageningen. Bart van Tuijl is acknowledged for his contributions to the experimental set-up. 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, 27–29 March 2013, Sakai, Japan.
- Bac, C.W., Van Henten, E.J., Hemming, J., Edan, Y., submitted for publication. Harvesting Robots for High-value Crops: State-of-the-art Review and Challenges Ahead. Wageningen University and Research Centre, Wageningen, The Netherlands.
- Blum, A.L., Langley, P., 1997. Selection of relevant features and examples in machine learning. *Artificial Intelligence* 97, 245–271.

- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman & Hall/CRC press, Boca Raton, FL.
- Bulanon, D.M., Burks, T.F., Alchanatis, V., 2010. A multispectral imaging analysis for enhancing citrus fruit detection. *Environmental Control in Biology* 48, 81–91.
- Center, S.E., 2008. Solar Radiation Hand Book. MNRE, Indian Metrological Department, India.
- Cieslak, D.A., Chawla, N.V., 2008. Learning decision trees for unbalanced data. *Lecture Notes in Computer Science*, 241–256 (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*).
- Davies, E.R., 2009. The application of machine vision to food and agriculture: a review. *Imaging Science Journal* 57, 197–217.
- Deng, J., Li, J., Zou, X., 2011. Extraction of litchi stem based on computer vision under natural scene. *Proceedings - International Conference on Computer Distributed Control and Intelligent Environmental Monitoring, CDCIEM 2011*, 832–835.
- DeVries, R.J., 2005. Spatial modelling using the Mahalanobis statistic: two examples from the discipline of plant geography. In: MODSIM05 – International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making, Proceedings, pp. 1368–1374.
- 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.
- Filella, I., Penuelas, J., 1994. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing* 15, 1459–1470.
- 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. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36, 119–123.
- Fransen, T.S., 2011. Determination of Specific Reflectance Curves for Different Parts of the Sweet-pepper Plant. BSc Thesis, Farm Technology Group, Wageningen University, Wageningen, The Netherlands. Open-access via: <<http://edepot.wur.nl/176380>>.
- 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.
- Huang, Y.J., Lee, F.F., 2008. Classification of phalaenopsis plantlet parts and identification of suitable grasping point for automatic transplanting using machine vision. *Applied Engineering in Agriculture* 24, 89–99.
- Humphries, S., Simonton, W., 1993. Identification of plant parts using color and geometric image data. *Transaction of the ASABE* 36, 1493–1500.
- Jain, A., 1997. Feature selection: evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 153–158.
- Jiménez, A.R., Ceres, R., Pons, J.L., 2000. A survey of computer vision methods for locating fruit on trees. *Transactions of the American Society of Agricultural Engineers* 43, 1911–1920.
- 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.
- Kapach, K., Barnea, E., Mairon, R., Edan, Y., Ben-Shahar, O., 2012. Computer vision for fruit harvesting robots – state of the art and challenges ahead. *International Journal of Computational Vision and Robotics* 3, 4–34.
- Kavdir, I., Guyer, D.E., 2004. Comparison of artificial neural networks and statistical classifiers in apple sorting using textural features. *Biosystems Engineering* 89, 331–344.
- Linker, R., Cohen, O., Naor, A., 2012. Determination of the number of green apples in RGB images recorded in orchards. *Computers and Electronics in Agriculture* 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.
- McCarthy, C.L., Hancock, N.H., Raine, S.R., 2010. Applied machine vision of plants: a review with implications for field deployment in automated farming operations. *Intelligent Service Robotics* 3, 209–217.
- Moisen, G.G., 2008. Classification and Regression Trees, *Encyclopedia of Ecology*. Academic Press, Oxford, pp. 582–588.
- Nakariyakul, S., Casasent, D.P., 2009. An improvement on floating search algorithms for feature subset selection. *Pattern Recognition* 42, 1932–1940.
- 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 Horticulturae (ISHS)* 691, 885–892.
- Ting, Y., Kondo, N., Wei, L., 2012. Sunlight fluctuation compensation for tomato flower detection using web camera. *Procedia Engineering* 29, 4343–4347.
- Unay, D., Gosselin, B., Debeir, O., 2006. Apple stem and calyx recognition by decision trees. In: *Proceedings of the 6th IASTED International Conference on Visualization, Imaging, and Image Processing, VIIP 2006*, pp. 549–552.
- 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-leaving cucumber plants grown in a high-wire cultivation system. *Biosystems Engineering* 94, 317–323.
- Wang, L., Sousa, W.P., Gong, P., 2004. Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *International Journal of Remote Sensing* 25, 5655–5668.
- Wickens, T.D., 2001. The Equal-variance Gaussian Model, *Elementary Signal Detection Theory*. Oxford University Press, USA (Chapter 2).
- Wieland, A., Wallenburg, C.M., 2012. Dealing with supply chain risks: linking risk management practices and strategies to performance. *International Journal of Physical Distribution and Logistics Management* 42, 887–905.
- Yang, C., Everitt, J.H., Bradford, J.M., 2008. Yield estimation from hyperspectral imagery using spectral angle mapper (SAM). *Transactions of the ASABE* 51, 729–737.