Effective Vision-Based Classification for Separating Sugar Beets and Weeds for Precision Farming

Philipp Lottes
Department of Photogrammetry
University of Bonn
Nussallee 15, 53115 Bonn
philipp.lottes@igg.uni-bonn.de

Markus Höferlin
Deepfield Robotics
Robert Bosch Start-up GmbH
Benzstr. 56, 71272 Renningen
markus.hoeferlin@de.bosch.com

Slawomir Sander
Deepfield Robotics
Robert Bosch Start-up GmbH
Benzstr. 56, 71272 Renningen
slawomir.sander@de.bosch.com

Cyrill Stachniss
Department of Photogrammetry
University of Bonn
Nussallee 15, 53115 Bonn
cyrill.stachniss@igg.uni-bonn.de

Abstract

Using robots in precision farming has the potential to reduce the reliance on herbicides and pesticides through selectively spraying individual plants or through manual weed removal. A prerequisite for that is the ability of the robot to separate and identify the value crops and the weeds on the field. Based on the output of the robot's perception system, it can trigger the actuators for spraying or removal. In this paper, we address the problem of detecting the sugar beet plants as well as weeds using a camera installed on a mobile field robot. We propose a system that performs vegetation detection, local as well as object-based feature extraction, random forest classification, and smoothing through a Markov random field to obtain an accurate estimate of the crops and weeds. We implemented and thoroughly evaluated our system on a real farm robot on different sugar beet fields and illustrate that our approach allows for accurately identifying the weed on the field.

1 Introduction

One target of sustainable farming is to increase yield while reducing reliance on herbicides and pesticides. Precision farming techniques seek to address this challenge by monitoring key indicators of crop health and targeting treatment only to plants that need it. Doing this manually, is a time consuming and expensive activity. There has, however, been a great progress on autonomous farming robots that target to automate the work on the field.

In order to build autonomous robots for farming applications, several challenges need to be addressed. These challenges include robust perception, fast and effective actuators, rough terrain navigation, long-term autonomy, and several others. In this paper, we investigate the first of those challenges, namely a part of the perception problem. Our aim is to develop an effective, vision-based perception system that can identify the value crop and distinguish it from weeds growing on the field. By automatically separating both classes of plants, we enable the robot to mechanically remove the weed or to perform spraying actions on a per-plant basis. An illustration of our field robot and an example classification results is depicted in Figure 1.

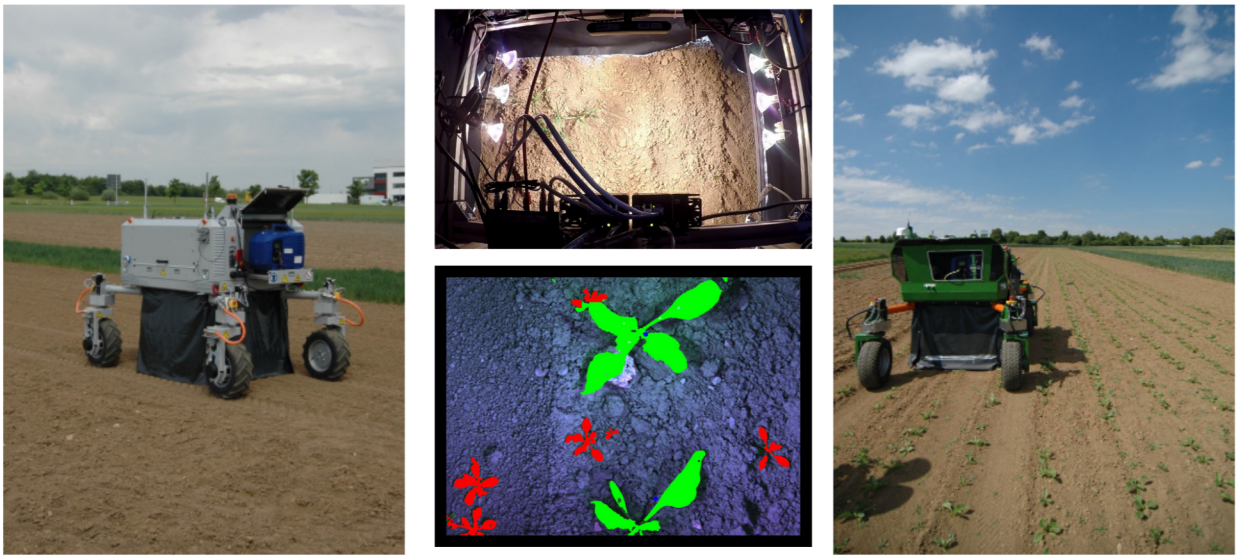

Figure 1: Left: Bonirob V3 robot operating on a sugar beet field. Middle: Downward looking camera capturing images of object space illuminated by halogen spots and example image with sugar beet/weed classification illustrated through green (crop) and red (weed) labels. Right: Bonirob V2 robot.

The contribution of this paper is a vision-based classification system for mobile robots to separate value crops from weeds. The crop under consideration here are sugar beets, an important crop in Germany and other countries in Northern Europe. To perform this classification task on the data obtained with out field robot, we rely on a processing pipeline that executes the following steps. We first preprocess each image to obtain a normalized average intensity for each channel and to separate the vegetation from the remaining parts of the image, i.e., the image background, which is mostly soil. We then compute a series of features in the image regions that correspond to vegetation and exploit a random forest for performing the classification of the image. In this paper, we consider two variants of this classification problem. The first variant computes *local features for keypoints* and classifies the area around each keypoint. The second variant is an *object- or segment-based classification*, which makes the decision for all pixels in an segment of vegetation pixels. After random forest classification, we take the neighborhood information between individual predicted keypoints into account. We achieve this through Markov random fields and in this way improve the individually predicted class labels of the random forest. Our approach can also exploit spatial priors, for example, if value crops were roughly planted at a known distance. We implemented the proposed system as ROS modules and evaluated them on different real field robots. We developed computationally demanding tasks on a GPU using CUDA to achieve classification results for online precision farming applications such as selective spraying or mechanical weed treatment.

For our experiments, we used different versions of BOSCH's Bonirob system and the sensor for classification is a single 4-channel JAI camera. We evaluated our approach on sugar beet plants at different growth stages and weed plants that grew on fields near Stuttgart, Germany. As the evaluations suggest, our system provides accurate classification results in a comparably short amount of time through the combination of the keypoint-based and the object-based approach. In our precision farming scenario, it is important to keep the number of false negatives, i.e., the number of sugar beet plants that are classified as weeds, small. This type of misclassification should be avoided as this would lead to the elimination of the value crop by the robot. In contrast, not detecting a weed is less critical. The evaluation of our approach suggests that the majority of weed plants get correctly classified while the number of false positives stays small.

2 Related Work

Extracting semantic information about the environment is a relevant topic in robotics (Ranganathan and Dellaert, 2007; Stachniss et al., 2005; Müter et al., 2013). In the context of agricultural applications, several vision-based crop and weed detection approaches for specific plants have been proposed. Whereas traditional methods on plant phenotyping typically bring the plant into a specialized, static sensor array, several innovative solutions have been developed for on-field operation. For example, the work by Müter et al. (2013), which focuses on the manual removal of weeds through the design and control of a mechanism for intra-row weeding. Related to that, Nieuwenhuizen (2009) presents an approach on the automated detection and control of volunteer potato plants.

Borregaard et al. (2000) perform crop versus weed classification using narrowband reflectance at 694 nm and 970 nm. Feyaerts and van Gool (2001) conducted a multi-spectral machine vision study with the aim to design an online weed detection system for selective spraying. They collected multi-spectral images using six channels with different wave length (441, 446, 459, 883, 924 and 988 nm) in the field. They report crop versus weed classification rates of 80% for sugar beet plants and 91% for weeds. Related to our method, Haug et al. (2014) present a method to classify carrot plants and weeds in RGB- and NIR-images without needing a presegmentation of the scenes into agglutinative objects. They achieve an average accuracy of 94% for carrot plants on an evaluation set of 70 images where both, intra- and inter-row overlap is present.

Other researchers have investigated the use of texture computed from grayscale and color images to identify plant species. Shearer and Holmes (1990) used gray level co-occurrence matrices in the hue-saturation-value (HSV) color space. Related to that, Burks et al. (2000) evaluated color texture classification of different weed species using a neural network classifier. Both works report that using statistical parameters extracted from co-occurrence matrices provide high discriminative power to identify or separate plants but accentuate that more research is needed for testing under uncontrolled field conditions. Latte et al. (2015) use features based on color space and gray level co-occurrence matrices to classify crop field images containing 8 different types of crop. They apply an artificial neural network classifier and achieve an average classification accuracy of 84% when using both the HSV based and gray level co-occurrence matrices based features. Their results show, that using statistical moments of the HSV distribution improves the overall classification performance.

Several works have been conducted in the context of leaf image classification and segmentation (Wang et al., 2008; Kumar et al., 2012; Cerutti et al., 2013; Elhariri et al., 2014). In the work by Wang et al. (2008), leaf images are segmented using morphological operators and shape features are extracted and used in a moving center hypersphere classifier to infer plant species. Kumar et al. (2012) start from segmented images of leaves using a binary classifier on global image signatures as a validity test and curvature features compared with a given database to extract the best match. To cover a variety of leaf shapes, also deformable leaf models and morphology descriptors have been exploited by Cerutti et al. (2013). Elhariri et al. (2014) compared a Random Forest classifier and a linear discriminant analysis based approach in their study for classifying 15 plant species through leaf images. They exploit HSV color space of leave images as well as gray level co-occurrence matrices to extract shape-, color- and vein features. Hall et al. (2015) conducted a study on features for leave classification. They compared classification performance on different feature types like typical handcrafted and ConvNet features using a random forest classifier. Furthermore, they evaluate the robustness of those features under simulated varying conditions on the public Flavia leaf dataset.

Tellaeche et al. (2008) present a vision-based approach for selective weed spraying. They capture images inclined downwards with respect to the horizontal plane of field scenes and subdivide them in to grid cells. For each cell a decision is made based on structural and area features using Bayesian decision theory. A further cell-based approach by Aitkenhead et al. (2003) fragments images in a top-down fashion containing seedlings of crop and weeds into 16 cells and classify each of them using a self organized neural network. They attain a classification performance close to 80%, but as in case of Tellaeche et al. (2008) at a comparably low resolution of the cells. In contrast, we provide labels for the full image resolution to allow a high precision treatment in object space.

Figure 2: From left to right: Raw input RGB and NIR image and processed NDVI image.

Hemming and Rath (2001) propose a vision based system which distinguishes carrots, cabbage and weeds using a fuzzy logic classifier. For leaf classification, they perform a presegmentation into individual plants to extract shape and color features per segment. They evaluate their approach in open field experiments and achieve classification accuracies of 72% up to 88%, but report that presegmentation of plants entails problems and embodies a limiting factor.

The contribution of this work is a visual 4-channel detection system for mobile robots operating on the field that allows for separating weeds from a value crop, here done for sugar beets. The proposed system performs vegetation detection, feature extraction, random forest classification, and smoothing through a Markov Random Field to obtain an accurate estimate of the crops and weed. This submission is an extension of a recent conference paper (Lottes et al., 2016). We extended our paper on several dimensions, which are (i) the combination of keypoint-based and object-based classification, (ii) a substantially extended experimental evaluation, (iii) the usage of a larger set of features, and (iv) an improved runtime of the overall approach.

3 Vision-Based Plant Classification

The primary objective of our proposed plant classification system is to enable mobile field robots to distinguish crops and weeds in agricultural field environments. Here, we consider sugar beets, a popular value crop in Northern Europe, especially in Germany. The classification is performed on a mobile robot, see Figure 1, which perceives the field using a 4-channel camera that offers in addition to RGB information also one near-infra-red (NIR) intensity measurement per pixel, i.e., RGB+NIR. See Figure 2 for example images. The NIR information is especially useful for separating the vegetation from the soil and other background due to the high reflectivity of chlorophyll and thus (healthy) plants in the NIR spectrum.

The main goal of the proposed system is identifying crop and weeds on a per-pixel basis in the RGB+NIR camera images. Our overall pipeline works in four steps: First, we identify the vegetation using the NIR information, which leads to a vegetation mask $\mathcal{I}_{\mathcal{V}}$, see Figure 3 (right) for an example. This step is highly effective as it allows us to compute the features on the subsequent processing steps only for the regions that correspond to vegetation. Second, we compute a set of features for the image regions that correspond to vegetation and classify them using random forests, which yields a probability distribution representing the fact that the area under consideration corresponds to our crop or to weed.

Here, we follow two different approaches: (i) a keypoint-based approach and (ii) a object-based approach. The keypoint-based approach computes local features on a dense grid of keypoints and performs the classification for each keypoint. This is computationally demanding but allows to handle situation in which two plants are closed to each other or overlap. The object-based approach performs only one classification per segment and thus is substantially faster to compute but cannot handle overlapping plants well. Finally, we improve the classification for the keypoint-based approach through exploiting neighborhood information using a Markov random field. This leads to a spatially smoothed labeling and and reduces the number of wrongly classified

Figure 3: Left: Histogram of $\mathcal{I}_{NDVI}$ and threshold (red) for separating vegetation. Middle: masked NDVI after threshold operation containing errors. Right: final vegetation mask after optimization.

keypoint. In the subsequent Subsections 3.1-3.5, we provide a detailed description of these steps. Finally, we discuss how to combine keypoint-based and object-based classification.

3.1 Vegetation Detection

Before any image operations are performed, we perform a local normalization of intensity values to obtain images with a normalized average intensity on a global scale. This is a standard procedure in most vision-based classification approaches and we implemented this step on a GPU and thus has only minimal impact on the overall runtime.

The goal of vegetation detection is to eliminate the irrelevant background from the image $\mathcal{I}$ so that the subsequent classification task operates on regions that correspond to vegetation. Due to the high reflectivity of chlorophyll in the NIR spectrum (Rouse et al., 1974), it is comparably easy to separate vegetation from soil or other objects. We compute a vegetation mask

$$\mathcal{I}_{\mathcal{V}}(i,j) = \begin{cases} 1, & \text{if } \mathcal{I}(i,j) \in \text{vegetation} \\ 0, & \text{otherwise} \end{cases}, \tag{1}$$

with the pixel location (i, j). To separate the vegetation, we exploit specific reflectance of healthy vegetation using the normalized difference vegetation index (NDVI) according to Rouse et al. (1974) using the NIR channel $\mathcal{I}_{NIR}$ and the red channel $\mathcal{I}_R$ on a per-pixel basis:

$$\mathcal{I}_{NDVI}(i,j) = \frac{\mathcal{I}_{NIR}(i,j) + \mathcal{I}_R(i,j)}{\mathcal{I}_{NIR}(i,j) - \mathcal{I}_R(i,j)} \tag{2}$$

Figure 2 (right) shows an example of a NDVI image $\mathcal{I}_{NDVI}$ for sugar beet plants and weeds. On the field, the reflectivity of chlorophyll typically leads to a bimodal intensity distribution in $\mathcal{I}_{NDVI}$ for healthy vegetation and allows us to perform a threshold-based classification on the $\mathcal{I}_{NDVI}$ information for every pixel. Figure 3 (left) depicts the NDVI intensity distribution for an example image.

A threshold-based classification based on the $\mathcal{I}_{NDVI}$ may lead to small residual errors. Examples for such small errors are visible on the top right area of the middle image in Figure 3. These effects are often caused by lens errors, especially chromatic aberration, resulting in slightly different mappings of the red and the near-infra-red light from the work space to pixels on the chip. Most of the residual errors can be eliminated through basic image processing techniques such as (i) requiring a minimum brightness in $\mathcal{I}_{NIR}$, (ii) using morphological opening and closing to fill gaps and to remove noise at contours, and (iii) removing regions corresponding a few pixels only. Figure 3 (right) depicts the application of the vegetation mask $\mathcal{I}_{\mathcal{V}}$ on the $\mathcal{I}_{NDVI}$ image.

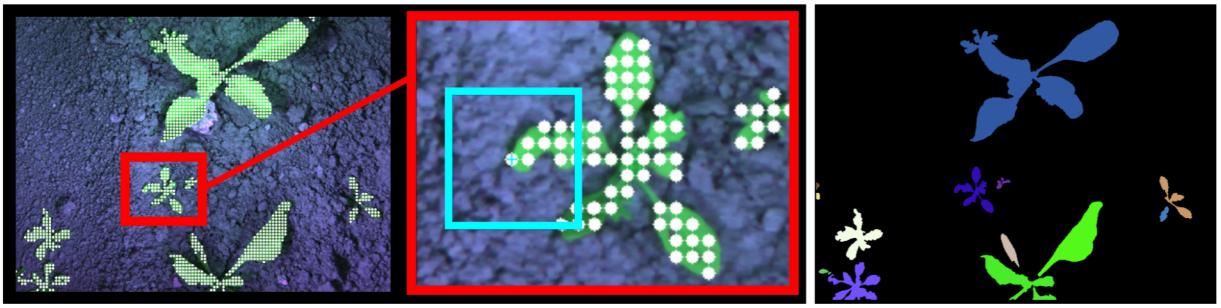

Figure 4: Left: Keypoints $\mathcal{K}$ (white) for classification at a 3 mm distance on the object. Middle: Zoomed view depicting the neighborhood $\mathcal{P}(\mathcal{K})$ of a keypoint $\mathcal{K}$ (blue) representing the region that is considered for the local feature computation for that keypoint. Right: Segmented vegetation. The segments define individual objects are used for the object-based classification. Here, we compute the features for each segment globally and perform a single classification per segment and not per keypoint.

3.2 Keypoint-Based vs. Object-Based Classification

There are two different ways to address the feature-extraction for our classification problem. First, we can compute *features for each keypoint* and perform the classification for each keypoint individually. Second, we can perform an object-based approach. Here, we define objects as segments of the vegetation pixels through connected components and perform only *one classification per object*. See Figure 4 for an example.

3.2.1 Keypoint-Based Approach

The keypoint-based approach computes features and perform the classification for each keypoint individually. This was the approach we used in our conference publication (Lottes et al., 2016). It has the advantage that it can deal with plants that overlap but at the cost of being computationally expensive. In our current implementation, keypoints are spaced 10 pixel by 10 pixel apart, which corresponds for our setup to a size of 3 mm by 3 mm in object space. To extract information about the class label of each keypoint, we use a fixed sized neighborhood to compute the features. In our current implementation, the neighborhood $\mathcal{P}(\mathcal{K})$ of a keypoint $\mathcal{K}$ has a size of 80 pixel by 80 pixel. Figure 4 (left, middle) illustrates the arrangement of the keypoints on an image and including the neighborhood for feature extraction. Details on the used features are given in Sec. 3.3.

3.2.2 Object-Based Approach

Alternatively, we can perform an object-based approach. Each object $\mathcal{O}$ is given through a connected component of the vegetation pixels in the image. Thus, we can perform one classification per object. Figure 4 (right) depicts examples of found objects. This approach has the advantage that it is substantially faster than the keypoint-based approach but suffers from situations in which weed and value crop overlap.

3.3 Feature Extraction

Both approaches share the concept of partitioning vegetation into parts which are classified separately and thus lead into the same feature extraction procedure, except that the areas in which the features are computed differ. We extract a set of features $\mathcal{F}$ for each approach, either for $\mathcal{P}(\mathcal{K})$ of for the whole object $\mathcal{O}$. We categorize the features into three groups, which are explained in the remainder of this section.

3.3.1 Statistical features

Our set $\mathcal{F}_{St}$ of statistical features includes the following parameters for describing the distribution of the inputs. Here, we use: min, max, range, mean, standard deviation, median, skewness, kurtosis, and entropy. These statistical features are computed on different input sources, which are given by different input channels of our image data as well as gradients, texture information, etc.

The input sources are S are defines by

$$S := \{\mathcal{I}_x, \nabla\mathcal{I}_x, \Delta\mathcal{I}_x, pLBP(\mathcal{I}_x), pLBP(\nabla\mathcal{I}_x), pLBP(\Delta\mathcal{I}_x)\}, \tag{3}$$

where $\mathcal{I}_x$ are different channels of the image data $\nabla\mathcal{I}_x$ its gradients, $\Delta\mathcal{I}_x$ the Laplacians, and $pLBP$ are locally binary patterns encoding texture information. All these quantities are defined in more detail in the remainder of this section.

First, we convert our four raw input channels R, G, B, NIR into the following six channels

$$\mathcal{I}_x \quad \text{with} \quad x = \{NDVI, G, B, H, S, L\}. \tag{4}$$

Here, NDVI is the normalized difference vegetation index as defined in Eq. (2), while G and B are the green channels of our images. H, S, and L refers to the Hue-Saturation-Lightness (HSL), which is a variant of the HSV color space, which is frequently used for plant and leaf classification, e.g. (Shearer and Holmes, 1990; Latte et al., 2015; Elhariri et al., 2014).

The $HSL(I_1, I_2, I_3)$ color space represents the three input channels I_1, I_2, I_3 as cylindric coordinates and separates intensity from color information. The dimension L called lightness is defined through

$$L = \frac{\max(I_1, I_2, I_3) - \min(I_1, I_2, I_3)}{2}. \tag{5}$$

We use the HSL space instead of the HSV space (lightness L instead of value V) because it is related to the average range of the input intensities and therefore more robust concerning biased intensities. Throughout this work, we define the HSL channels as

$$HSL = \text{HSL}(\mathcal{I}_{NDVI}, \mathcal{I}_G, \mathcal{I}_B). \tag{6}$$

For each input source $\mathcal{I}_x$, we consider also the gradients $\nabla\mathcal{I}_x$ and the Laplacian (second order gradients) $\Delta\mathcal{I}_x$. The magnitudes of $\nabla\mathcal{I}_x$ and $\Delta\mathcal{I}_x$ provide information about structure and homogeneous regions and are computed by:

$$\nabla\mathcal{I}_x = \left|\frac{\partial\mathcal{I}_x}{\partial i}\right| + \left|\frac{\partial\mathcal{I}_x}{\partial j}\right| \tag{7}$$

and

$$\Delta\mathcal{I}_x = \left|\frac{\partial^2\mathcal{I}}{\partial i^2} + \frac{\partial^2\mathcal{I}}{\partial j^2}\right| \tag{8}$$

Finally, we take into account distributions of texture information and contrast. The distributions are based in local binary patterns (LBP) according to Ojala and Pietikinen (1999). The LBP operator performs thresholding operations within a 8-connected neighborhood based on the value of its center pixel and converts this pattern as binary number. Figure 5 illustrates the computation of a LBP number and the associated contrast measure C for a pixel. The final distribution for an image or a local area I is defined as

$$pLBP(I) = p(LBP(I), C(I)). \tag{9}$$

The computation of our nine statistical features $\mathcal{F}_{St}(S)$ on all input sources S leads to 324 statistical features. A summary of the statistical features and inputs sources is given in Table 1.

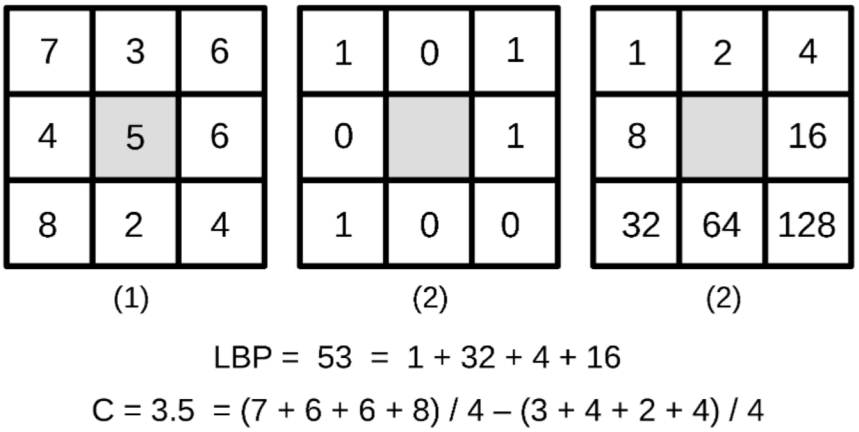

Figure 5: Example for the computation of a LBP number and the corresponding contrast measure C for a pixel given its 8-connected neighborhood. (1) input, (2) threshold operation by value of center pixel, and (3) binomial weights according to Ojala and Pietikinen (1999).

3.3.2 Shape Features

The next set of features describes different aspects of the shape of the plant. As before, the features are computed either in local neighborhood $\mathcal{P}(\mathcal{K})$ of a keypoint or for an object $\mathcal{O}$. The shape features $\mathcal{F}_{Sh}$ only need to be computed on the vegetation mask $\mathcal{I}_{\mathcal{V}}$, which is a binary image. We consider the following features describing contours, relations to geometric primitives and geometrical ratios:

- *Rectangularity* of the contour using its major a and minor b axes of the minimum enclosing oriented ellipse.

$$\mathcal{F}_{13} = \frac{\text{area}}{a\,b}, \tag{10}$$

  where area refers to the area covered by vegetation within the patch or the size of the object.

- *Aspect ratio* of the major a and minor b axes of the minimum enclosing oriented ellipse of the contour:

$$\mathcal{F}_{14} = \frac{a}{b} \tag{11}$$

- *Area change under smoothing*, which describes how the area of vegetation changes due to a smoothing with different sized Gaussian kernels G and is given by the ratio of the areas:

$$\mathcal{F}_{15} = \frac{\text{area}(G_{\mathcal{I}_{\mathcal{V}}}(\sigma))}{\text{area}(G_{\mathcal{I}_{\mathcal{V}}}(2\sigma))} \tag{12}$$

- *Form factor* F, which provides a measure of the shape of an object

$$\mathcal{F}_{16} = \frac{4\,\pi\,\text{area}}{\text{perimeter}^2}, \tag{13}$$

where *perimeter* is the perimeter of the area that is covered by the vegetation. We additionally exploit the convexity, compactness and solidity feature as described in Haug et al. (2014). Table 1 gives a summary of all features used in our classification system and described on which inputs they are computed.

Table 1: Features used by our classification system

Nr.	Feature Set
$\downarrow$	***Statistical Features*** $\mathcal{F}_{St}(S)$
$\mathcal{F}_1$	min
$\mathcal{F}_2$	max
$\mathcal{F}_3$	range
$\mathcal{F}_4$	mean
$\mathcal{F}_5$	standard deviation
$\mathcal{F}_6$	median
$\mathcal{F}_7$	skewness
$\mathcal{F}_8$	kurtosis
$\mathcal{F}_9$	entropy
	All statistical features $\mathcal{F}_{St}(S)$ are computed from input sources in $S = \{\mathcal{I}_x, \nabla\mathcal{I}_x, \Delta\mathcal{I}_x, pLBP(\mathcal{I}_x), pLBP(\nabla\mathcal{I}_x), pLBP(\Delta\mathcal{I}_x)\}$ with $x = \{NDVI, G, B, H, S, L\}$ this leads to 324 statical features, i.e. $9\,\mathcal{F}$ (features) $\cdot\, 6\,x$ (channels) $\cdot\, 6\,S$ (input sources)
$\downarrow$	***Shape features*** $\mathcal{F}_{Sh}$ computed on binary image $\mathcal{I}_{\mathcal{V}}$
$\mathcal{F}_{10}$	Convexity
$\mathcal{F}_{11}$	Compactness
$\mathcal{F}_{12}$	Solidity
$\mathcal{F}_{13}$	Rectangularity
$\mathcal{F}_{14}$	Aspect ratio of the minimum enclosing ellipse
$\mathcal{F}_{15}$	Area change under smoothing
$\mathcal{F}_{16}$	Form factor
$\downarrow$	***Other features***
$\mathcal{F}_{17}$	$\mathcal{F}_9(\nabla\mathcal{I}_{NDVI})/\mathcal{F}_9(\Delta\mathcal{I}_{NDVI})$
$\mathcal{F}_{18}$	plant arrangement prior, see Eq. (16)

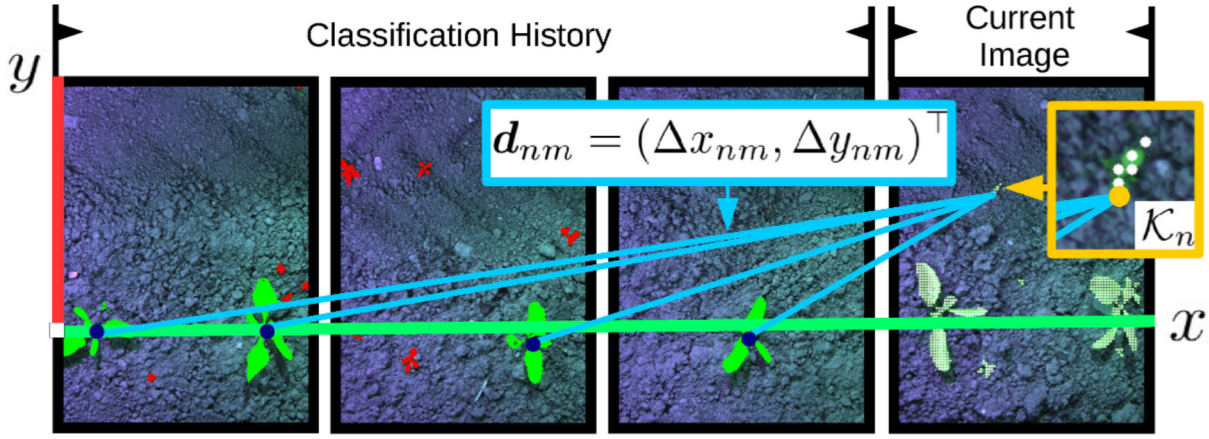

Figure 6: Computed coordinate differences $\boldsymbol{D}_n$ (blue) of query keypoint $\mathcal{K}_n$ (orange *in zoomed area*) to all classified crops in neighboring images. These images are aligned using odometry data from robot's wheel encoders.

3.3.3 Other Features

We use two other features, which turned out to be effective for outF classification problem. First a ratio between the entropy of the first order gradient of the NDVI image and its Laplacian:

$$\mathcal{F}_{17} = \mathcal{F}_9(\nabla \mathcal{I}_{NDVI}) / \mathcal{F}_9(\Delta \mathcal{I}_{NDVI}) \tag{14}$$

Second, a relative plant arrangement prior that describes the expected spatial distribution of the value crop on the field. Due to the fact that crop is often sowed in an automated fashion, prior information about the arrangement of the plants on the field can further improve the classification performance. We incorporate a plant arrangement as an additional feature. We use the probability for a certain keypoint $\mathcal{K}_n$ or object $\mathcal{O}_n$, corresponding to a crop based on it relative location to previously classified crops.

To represent the relative arrangement of crops, we use a coordinate system that is defined according to the crop row, which defines the x-direction. Given this coordinate system, we compute all absolute coordinate differences $\boldsymbol{D}_n$ of the keypoint $\mathcal{K}_n$ / object $\mathcal{O}_n$ in x and y direction between $\mathcal{K}_n$ / $\mathcal{O}_n$ and all positively classified M crops. See Figure 6 for an example illustrated by means of the keypoint-based approach. The coordinate differences are given by

$$\boldsymbol{D}_n = (|\boldsymbol{d}_{n1}|, \ldots, |\boldsymbol{d}_{nM}|)^\top \text{ with } |\boldsymbol{d}_{nm}| = (|\Delta x_{nm}|, |\Delta y_{nm}|)^\top, \quad \text{with} \quad n = 1 \ldots \mathcal{K} \text{ or } \mathcal{O}, \quad m = 1 \ldots M \tag{15}$$

Based on these distances, we compute

$$\mathcal{F}_{18} = p(\omega_c \mid \boldsymbol{D}) = \frac{p(\boldsymbol{D} \mid \omega_c)\, p(\omega_c)}{\sum_\omega p(\boldsymbol{D} \mid \omega)\, p(\omega)}, \tag{16}$$

where ω refers to a class label, with ω_c corresponding to crops. The class conditional probability distribution $p(\boldsymbol{D} \mid \omega_c)$ and the relative frequency of the classes $p(\omega)$ are learned from training data or can be provided if the information is known from sowing. For $p(\boldsymbol{D} \mid \omega_w)$, we consider a uniform distribution, due to the fact that in reality weed can grow anywhere. Note that this feature is only available if at least k neighboring images have already been classified. In our current implementation, we select k=5. Our tests have shown that this choice favors a good trade off between uncertainty of the relative pose estimation, caused by odometry errors, and the probability to find crop in this image sequence which in necessary to compute $\boldsymbol{D}_n$.

3.4 Random Forest Classification

i For the classification, we apply a random forest (Breiman, 2001) because it provides comparably robust classification results. As an ensemble method, random forests reduce the risk of overfitting to some degree and can implicitly estimate confidences for the class labels. The key idea is to construct a large number of decision trees ("a forest") at training time by randomizing the use of features and elements from the training data. During the operational phase, they output a class label or a distribution over the class label based on the output of the individual decision trees.

Random forests are capable of solving multi-class problems. In our approach, we are basically interested in distinguishing two classes, i.e, the "value crop", here sugar beets, referred to as ω_c and the "weed" class ω_w. For the object-based approach, however, we introduce an addition class "mixed" (ω_m) for segmented objects that contain both weeds and sugar beets as the plants overlap. This is only relevant for the object-based approach and not for the keypoint-based one. Based on the outputs of the different trees, we can compute a probability distribution $p(\omega \mid \mathcal{F})$ over the possible class labels.

Random forests run efficiently on large datasets and trained classifiers are fast to evaluate and are also easy to parallelize. In our current implementation, we use all cores of our CPU by running the individual trees of the forest in different threads. An interesting property of the random forest allows for dealing with missing data, for example, if feature is not available. This is relevant for our application in case the relative plant arrangement prior is not known.

3.5 Thorough Smoothing via Markov Random Fields for Keypoint-Based Classification

This section is only relevant for the keypoint-based approach and does not apply for the object-based one. The keypoint-based classification system described so far, computes each label assignment independently of the other nearby labels. In order to improve the classification results and to exploit the topological relationships between keypoints, we apply a Markov random field (MRF). We compute a global classification based on the individually computed class labels $\omega(\mathcal{K})$ of the keypoints by considering their spatial distribution and class confidences $p(\omega(\mathcal{K}) \mid \mathcal{F}(\mathcal{K}))$. We achieve this by minimizing the energy function

$$E(\omega(\mathcal{K})) = \sum_{\mathcal{K}} \left(D(p(\omega(\mathcal{K}) \mid \mathcal{F}(\mathcal{K}))) + \sum_{\mathcal{K}' \in \mathcal{N}_4(\mathcal{K})} V(\omega(\mathcal{K}'), \omega(\mathcal{K})) \right) \tag{17}$$

through belief propagation. Here, $E(\omega(\mathcal{K}))$ describes the quality of a labeling under the key assumption that neighboring labels vary slightly, but also can change fiercely at class borders. Therefore, two energy terms are needed. The first one D considers the confidence of a class label and through this defines the energy which is needed to change the label. The term V describes the energy for smoothing the four-connected neighborhood, i.e., how many neighboring labels agree. We minimize Eq. (17) using efficient belief propagation according to Felzenszwalb and Huttenlocher (2006). The MRF optimizes the classification results as it reduces wrong local estimates by exploiting neighborhood information and considers the confidence of the individual keypoint classifications.

In order to obtain a prediction per pixel instead of per keypoint, we perform a straightforward nearest-neighbor interpolation of the predicted class labels with respect to the vegetation mask between the keypoints. Figure 7 depicts a typical example of a beet, where MRF optimization leads to a better performance. One effect of the MRF smoothing is that wrongly classified stem areas as depicted in Figure 7 (left) are corrected.

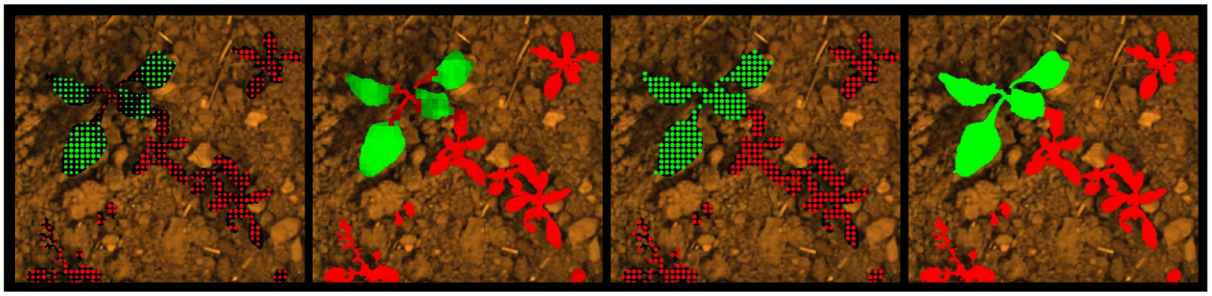

Figure 7: Left to right: Random forest classification, interpolation of the Random Forest results to full resolution, keypoints after spatial smoothing with MRF and classification map after interpolation of the MRF results. The MRF smoothing eliminates the few wrongly classified keypoints at the plant center and outliers.

3.6 Combining the Keypoint-Based and Object-Based Classification

To achieve both, the fast execution time of the object-based approach as well as the ability to deal with overlapping plants of the keypoint-based classifier, we combine both approaches in a cascade. Through our cascaded classification, we initially apply the object-based approach for the whole image. All objects, which are identified as weeds or sugar beet with high certainty, keep their labeling. For objects with uncertain classification results or which are classified as "mixed" objects, i.e. undersegmented objects due to overlap, we apply the keypoint-based approach. As a result of that, the features only need to be computed for a comparatively small number of keypoints and thus we can maintain an overall fast computation time.

In more detail, the keypoint-based approach is only executed for objects for which at least one of the two conditions hold. Either the random forests suggests a mixed object

$$\operatorname*{argmax}_{\omega} p(\omega \mid \mathcal{F}) = \omega_m \tag{18}$$

or the random forest is too uncertain about is results, i.e.

$$\max\, p(\omega \mid \mathcal{F}) < t^{\mathcal{O}}_{min} \tag{19}$$

where $t^{\mathcal{O}}_{min}$ indicating the minimum probability for the class suggested by the random forest. All those objects are passed to the keypoint-based classifier for a further in-depth investigation.

4 Experiments

The evaluation is designed to illustrate the performance of our plant classification system and to support the main claims made in this paper. These claims are: (i) our approach is suitable for classifying sugar beets and weeds under real world conditions, (ii) we show that both, the keypoint-based as well as the object-based approach perform well on our datasets and can be combined to compensate their drawbacks respectively, (iii) we illustrate that our approach also provides good classification results on real fields if the grow stage of the vegetation has changed, and (iv) show that using pose information from sowing can have a significant impact to the classification results.

In Sec. 4.1 and Sec. 4.2, we first introduce the experimental setup including a description of the field robot, the camera system for plant image acquisition and a listing of relevant information about datasets. In Sec. 4.3, we define an evaluation metric to quantify performance and present several experiments to demonstrate the performance of our proposed approach. Furthermore, we evaluate the importance of the used features in Sec. 4.4 and analyze both classification approaches in terms of runtime in Sec. 4.5.

Figure 8: Left/Middle: Bonirob V3 with support structure for mounting the camera as well as the shading and attached halogen spots for illumination of the shaded space under Bonirob. Right: Two JAI AD-130 GE cameras mounted on support structure straight looking downwards.

4.1 Field Robot and Camera System

All experiments have been conducted with different generations of the Bonirob field robot, shown in Figure 1 and Figure 8. The Bonirob is a multi-propose field robot by BOSCH DeepField Robotics and has been developed for agricultural applications such as selective spraying, weed control as well as plant and soil monitoring. The Bonirob provides empty slot to install different tools, called apps, for these specific tasks. In terms of navigation on rough terrain, the robot is equipped with four independently steerable wheels and allows for flexible movements.

For the data acquisition, we mounted the camera under the robot and built a pond foil curtain around it to shade cameras field of view in order to be independent from natural light source. We attached several halogen spots for artificial illumination. See Figure 8 for an illustration of the robot together with the support structure and setup of the halogen spots.

For the image acquisition, we used a 4-channel JAI AD-130 GE camera pointing downwards on the field approximately 70 cm above soil. The RGB+NIR images of the JAI camera were captured with a resolution of 1296×966 pixels using a Fujinon TF15-DA-8 lens with a fixed focal length of $8mm$, which yields a ground resolution of approximatively $3\ \frac{px}{mm}$ and a field of view of 24 cm in driving direction and 31 cm orthogonal to it. See Figure 1 (middle) for an illustration of the camera while robot is operating on the field and Figure 8 (right) for the mounting of the camera. The latter depicts two JAI AD-130 GE cameras equipped with different lenses. For the purposes of this work, we only use images acquired with the Fujinon TF15-DA-8 lens. This camera system allows the acquisition of time synchronized and aligned images through a prism based mapping of the incoming light to the CCD arrays, one for RGB and NIR respectively. The images were captured with a frequency of 1 Hz while the robot was moving over the field with approximately $0,3\ \frac{m}{sec}$. For example images obtained with this sensor see Figure 9. The robot uses ROS and all data was logged in the standard *rosbag* format.

4.2 Datasets

We used the robot on sugar beet fields near Stuttgart, Germany. In the evaluation reported below, we mainly use three datasets, here called A, B and C. The datasets A and B have been collected on the same field with a temporal difference of one week. Dataset C has been collected on a different field. The three datasets vary between 2-leaf and late 4-leaf growth stages, which are the main growth stages for which (manual) weed removal is most effective. If the weeds are been eliminated during these growth stages, the sugar beet plants typically have a sufficient competitive advantage over newly growing weeds.

Figure 9 illustrates an example image of a typical field situation for each dataset regarding to the growth stage of the plants. Related to the growth stage in dataset B, the area of crop is twice as large as in dataset A

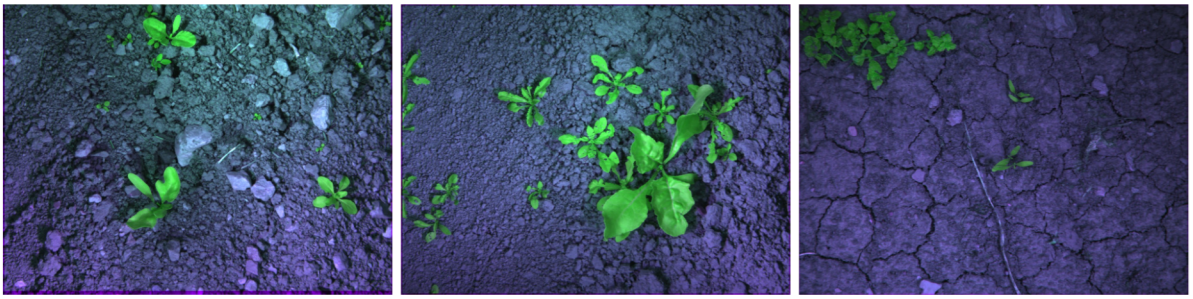

Figure 9: Example images of the captured datasets used for our experiments. For best view, images are shown as false color representation (Red, NIR, Green). From left to right: Dataset A, B and C. A was captured one week before B on the same field and C was captured on a different field.

Table 2: Information about the datasets. Priors are related to the number of keypoints $p_{\mathcal{K}}(\omega)$ or number of objects $p_{\mathcal{O}}(\omega)$ considered from ground truth data. area coverage is given by class wise relation of biomass pixels.

Parameter	**Dataset A**	**Dataset B**	**Dataset C**
#non-overlapping images	1024	694	974
#crops	1315	831	1145
$p_{\mathcal{K}}(\omega_c)$ / $p_{\mathcal{K}}(\omega_w)$	0.74 / 0.26	0.66 / 0.34	0.32 / 0.68
$p_{\mathcal{O}}(\omega_c)$ / $p_{\mathcal{O}}(\omega_w)$ / $p_{\mathcal{O}}(\omega_m)$	0.25 / 0.73 / 0.02	0.21 / 0.74 / 0.05	0.53 / 0.47 / -
area coverage ω_c / ω_w / ω_m	0.71 / 0.17 / 0.02	0.58 / 0.28 / 0.14	0.32 / 0.68 / -
growth stage	4-leaf	late 4-leaf	2-leaf/early 4-leaf

and for the weed is approximately four times as large as in A. Dataset C contains crops of an earlier growth stage and different weed species. See Table 2 for further information about the datasets. To allow for a ground truth evaluation, we manually labeled all sugar beet plants and weeds in the images on a per-pixel basis.

4.3 Evaluation of the Quality of the Classification Results

The experiments presented here are designed to analyze the quality of the classification output. We evaluate respectively the keypoint-based as well as the object-based approach and furthermore the combination of both. Regarding the class labels, we refer for analyzing the classification results to sugar beet plants (ω_c) as *sugar* and weeds (ω_w) as *weed*. For the evaluation of the object-based classification we additionally refer to mixed objects (ω_m) as *mixed*. For our experiments, we define an object as mixed object, if it consists of both, sugar and weed pixels, where either constitutes more the 10% of the total pixels.

We illustrate the overall performance of the classification results by ROC curves and Precision-Recall plots. Furthermore, we illustrate some of the classification results visually to highlight typical results achieved by our system. For the ROC curves and Precision-Recall plots, we varied the threshold $t \in [0, 1]$ for class labeling concerning to the estimated probability distribution $p(\omega \mid \mathcal{F})$ of the random forest. In case of the keypoint-based approach, i.e. binary classification, this leads to the following mapping function of the class label distribution to the class labels

$$\omega = \begin{cases} \omega_c \text{ (sugar beet)}, & \text{if } p(\omega_c \mid \mathcal{F}) \geq t \\ \omega_w \text{ (weed)}, & \text{otherwise} \end{cases} \tag{20}$$

By changing the parameter t from its default value 0.5, we can influence the classifier to minimize false negatives or to minimize false positives.

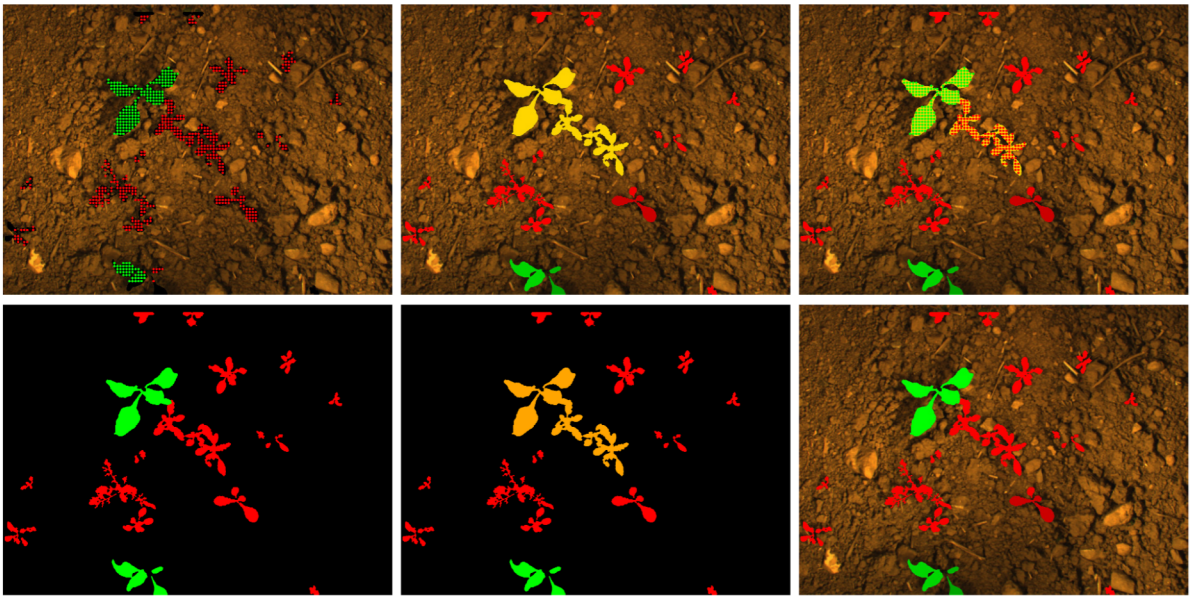

Figure 10: Visual illustration of classification results (i) left column: Keypoint-based classification of the vegetation into the value crop (green) and weed (red) and corresponding ground truth image, (ii) middle column: Object-based classification and labeled ground truth information for objects, including mixed objects (orange) and (iii) right column: Combined classification results according to Sec. 3.6.

Regarding the evaluation of the object-based classification including the "mixed" class, we binarize the multi-class problem in order to illustrate the results also as ROC curves and Precision-Recall plots. To achieve this, we compute the individual performance of a class in one-vs.-all mode. Note that during all experiments, we only use independent image scenes, i.e., no overlapping images have been considered.

All performance measures are based on a comparison between predicted vs. ground truth area in image space. Note that we exclusively evaluate the vegetation parts of the images. Considering all image pixels would lead to even better results, as a large number of pixels cover soil, which is trivially identified as non-vegetation using the NDVI. Thus, this renders the evaluation more challenging.

In ROC and Precision-Recall plots, the term "RF" refers to random forest- only classification and "MRF" to the combination of random forest and MRF. The term "without prior" refers to the approach neglecting the relative plant arrangement prior, "sugar" to sugar beet plants, "weed" to weeds and "mixed" to mixed class objects. For the combined classification approach "min threshold for objects" refers to the applied minimum probability $t^{\mathcal{O}}_{min}$ for the most likely class according to Eq. (19).

4.3.1 Cross Validation on Dataset *B*

The first set of experiments are designed to illustrate the performance of our sugar beet classification system and to highlight the individual capabilities of the keypoint- and object-based approach. We present here the results from dataset B because it is the most difficult dataset due to substantial plant/weed overlaps. We apply a K-fold (with $K = 15$) cross validation according to Alaydin (2004) on dataset B. For each of the K training sets, we learn a random forest with 150 trees, a maximum tree depth of 10 and consider $\text{ceil}(\sqrt{\#\text{features}}) = \text{ceil}(\sqrt{333}) = 19$ randomly selected features to find the best split at each node in a tree.

Figure 10 illustrates an example classification result of the cross validation on B and the corresponding ground truth information. We show these results to highlight the potential of the combined classification.

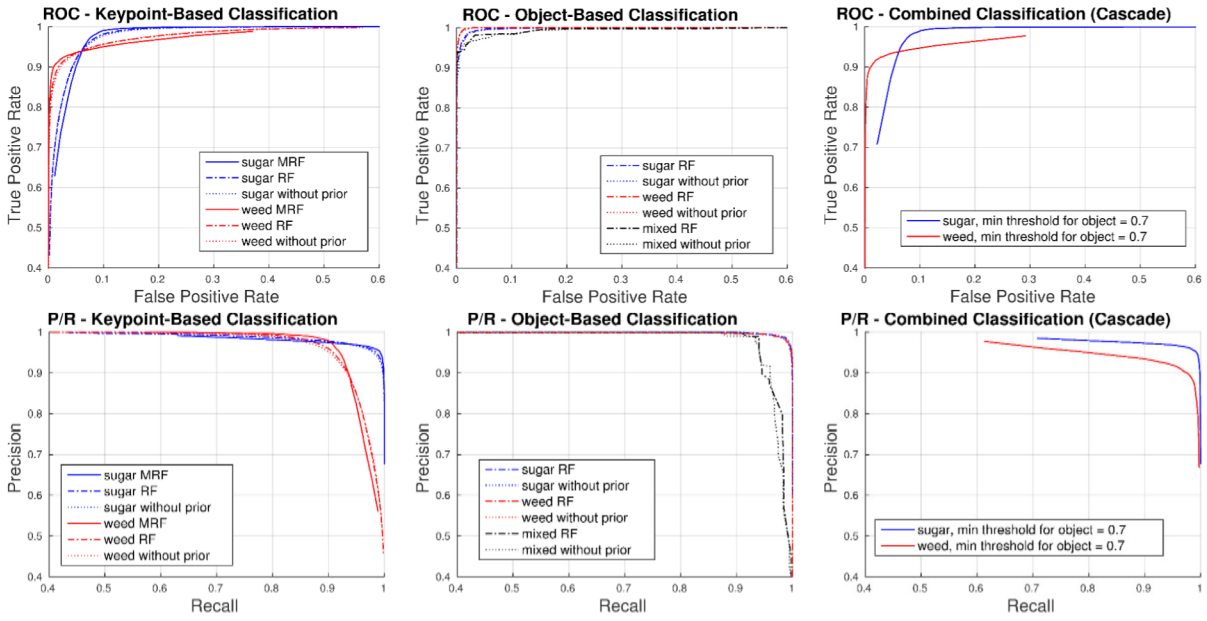

Figure 11: ROC curves (top) and Precision-Recall plots (bottom) from evaluation of our different classification approaches using 15-Fold cross validation on dataset *B*. Left: Performance of the keypoint-based approach. Middle: Performance of the object-based approach. Right: Performance of the combined approach. For a more detailed view on the results, we show only a cropped area of ROC- and Precision-Recall space, i.e. the best 60%.

The keypoint-based classification (Figure 10, left column) is able to separate the sugar beet plant from the adjacently grown weeds, but still fails to correctly predict a few keypoints. In this particular case, the object-based classification performs perfectly. Also the mixed object in center of the shown image is correctly classified. However, as a stand alone classification system, the object-based approach is still not able to separate whole vegetation into crops and weeds (if no mixed labels are allowed). To overcome this issue, we apply both classifiers in a cascade as described in Sec. 3.6 and thus achieve the classification performance shown in the right column. A further advantage of the combined approach is that the execution time for this particular image was twice as fast as for the keypoint-based classification due to the fast object-based classification and the reduced number of keypoints. More details on the runtime analysis are given in the remainder of this evaluation.

The resulting ROC- and Precision-Recall plots are depicted in Figure 11 and illustrate overall performance on dataset *B*. For the keypoint-based classification, we achieve a maximum overall accuracy of 96% at $t = 0.5$, i.e., a labeling with the most likely class, according to Eq. (20). This means that the keypoint-based approach has neither a preference for sugar beet nor for weed. For $t = 0.5$, we achieve a true positive rate (TPR/recall) of 98% for sugar with a precision of 95%. In terms of weeds, we obtain a true negative rate (TNR) of 90% with an precision of 98%. Thus, the system classifies the majority of plants correctly and keeps the false negative rate (FNR), i.e. the percentage of sugar beet pixels, which are classified as weeds, small (FNR = 2%). The corresponding Precision-Recall plot manifests that nearly all vegetation pixels, which are classified as sugar beet are predicted with a high confidence.

In terms of MRF smoothing, we gain on average an improvement of 1% in overall accuracy. The main reason for that increase in precision for weeds is due to smoothing of stem regions as depicted in Figure 7.

The object-based approach can select among the three classes $\{\omega_c, \omega_w, \omega_m\}$ using $\omega = \text{argmax}_\omega\, p(\omega \mid \mathcal{F})$. We classify 99.5% of the area covered by sugar beet plants correctly with a precision of 96%. For weeds, we obtain a positive rate of 98% with a precision of 98%. Thus, the overall error rate for non- overlapping

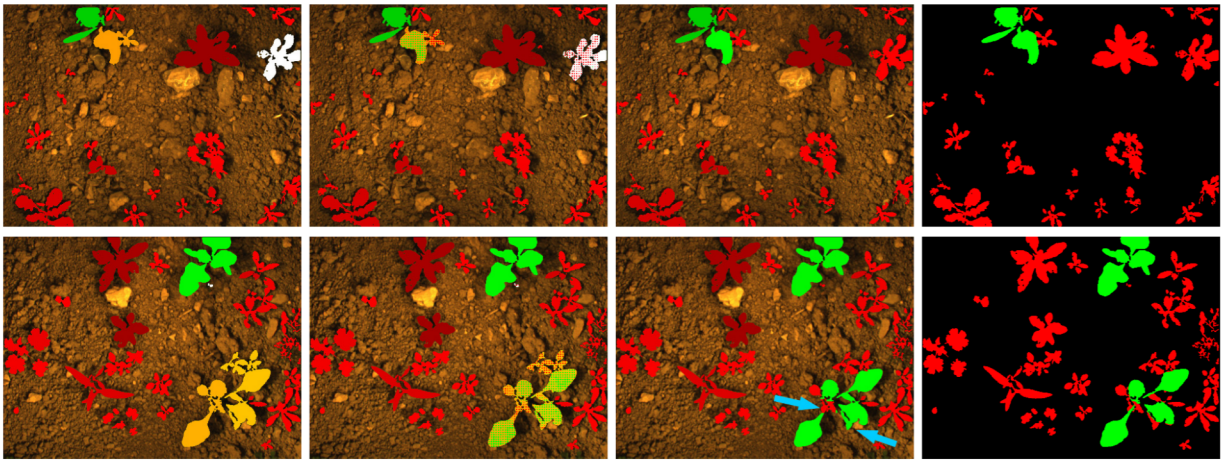

Figure 12: Visual illustration of results (one result per row), achieved by the combined classification approach. From left to right: object-based classification, combined classification including mixed (orange) and uncertain (white) objects, combined classification at full resolution, ground truth. The confidences for crops (green) and weeds (red) are encoded with the radius of the keypoints or in case of objects with the color intensity. Drawn arrows (blue) indicate classification errors.

Table 3: Limiting the false negative rate, i.e. the percentage of sugar beets classified as weeds to the values shown in the first row of the table, yields the given true negative rates (TNR), i.e. percentage of correctly classified weeds that will be eliminated.

postulated FNR	0.1%	0.5%	1%	2.5%	5%
keypoint TNR	72%	85%	89%	92%	93%
combined TNR	82%	88%	91%	92%	93%

objects of is close to zero. Even most of mixed object are classified correctly with a positive rate of 94%, but with a lower precision of 90% related to crops and weeds. This is due the significantly smaller number of training examples for the mixed class. In dataset B, the mixed objects cover around 14% of the vegetation. Because of this, the overall performance of the object-based approach is limited. To deal with these mixed objects both classification systems should be combined.

For the combined approach, we report the results with a minimum probability of $t^{\mathcal{O}}_{min} = 0.7$ in Eq. (19). Regarding the keypoint-based approach, we again achieve an maximum overall accuracy of 96% at $t = 0.5$ and achieve a TPR of 99% for sugar beets with a precision of 95%. In terms of weeds, we obtain a true TNR of 90% with an precision of 98%. Generally, the combined classification offers a similar performance as for the keypoints. The main differences are (i) a better precision for weeds, which arises from high precision for weed of the object-based classification and (ii) a substantially faster execution time. Figure 12 depicts two example classification results of the cross validation on B obtained in difficult situations with substantial weeds growing close to sugar beets.

Table 4: Limiting the false positive rate (FPR), i.e. the percentage of weeds that are classified as sugar beets specified in the first row, yields the true positive rates (TRP), i.e the percentage of the correctly classified sugar beets.

postulated FPR	2%	5%	10%	15%	30%
keypoint TPR	71%	89%	99%	99.5%	99.9%
combined TPR	72%	88%	98%	99%	99.9%

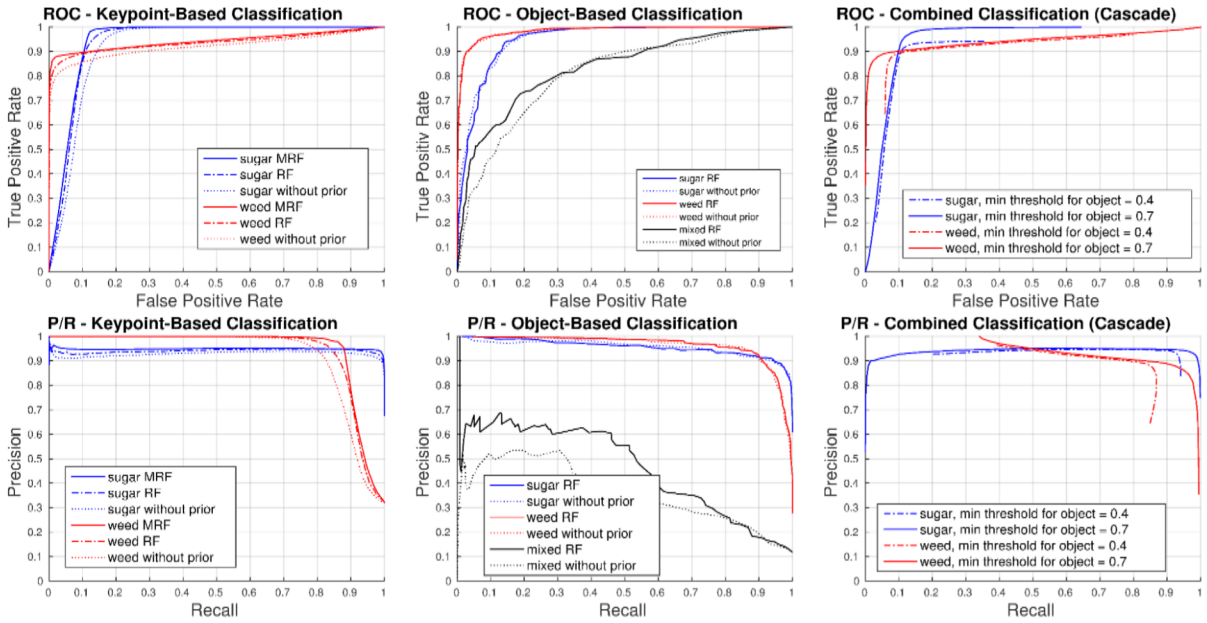

Figure 13: Prediction of dataset B with a random forest classifier learned on dataset A. ROC curves (top) and Precision-Recall plots (bottom) from prediction of our different classification approaches. Left: Performance of the keypoint based approach. Middle: Performance of the object-based approach. Right: Performance of the combined classier.

As precision farming robots are also envisioned to manually eliminate weeds, a wrong classification of a sugar beet plant will lead to its elimination, while a false positive means that a weed will not be treated. In order to avoid eliminating value crops, the plant classification system has to avoid false negatives with a higher priority than false positives. Therefore, we provide in Table 3 the expected true negative rate *for a given false negative rate*. The inverse, which may be relevant for other application in the context of phenotyping, is shown in Table 4.

4.3.2 Dataset B Predicted with Classifier learned on Dataset A

This experiment is designed to analyze the ability of our classifier to generalize over different datasets. Generalizing over different datasets of fields is more challenging as features computed from one single dataset are likely to be correlated due the spectral and physical appearance of the vegetation at the time of the data acquisition.

To measure the generalization performance, we evaluated our classification system in a setup where training is performed on a field with a different growth stage than in the test dataset. We tested on dataset B but trained only on dataset A and the resulting ROC- and Precision-Recall curves are shown in Figure 13. Both datasets differ in terms of the growth stage, early 4-leaf vs. late 4-leaf stage.

In case of the keypoint-based classification, we obtain a maximum overall accuracy of 92% with $t = 0.45$ by using the random forest classification in combination with the MRF. We achieve a TPR of 97% for sugar beets with a precision of 94%, which is comparable to the result for the cross validation on dataset B. In terms of weeds, we obtain a true TNR of 87% with an precision of 94%, which means that performance decreases around 3% for the TNR. This is mainly caused by weeds, which are wrongly classified as sugar beets (FP), which grow close to crops, as shown in Figure 14. Dataset A does not provide enough training samples for the keypoint classifier, where overlap is present. Without MRF smoothing, we obtain 90% overall accuracy, because the results of the random forest are more noisy due to neighboring class labels. Thus, the

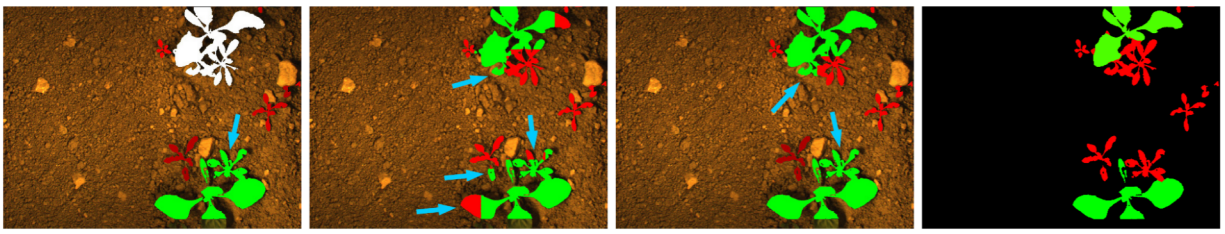

Figure 14: Visual illustration of results of the prediction of dataset B with the classifier learned on dataset A showing typical errors. From left to right: object-based classification also showing uncertain crops (white), keypoint-based classification including MRF smoothing, combined classification and ground truth.

gain of the MRF smoothing is larger than for the first experiment.

In contrast to that, the Precision-Recall plots for the object-based approach indicate, that weeds are classified with a significantly higher precision for a recall $\geq 90\%$. In terms of sugar beets, the results are less accurate. This is mainly caused by the decreased performance for mixed objects. Manual inspection of the data showed that a substantial number of mixed objects are classified as sugar beets, see Figure 14 for an example. Thus, the recall for sugar beets is high while the precision is lower and for weed vice versa. Here, the problem is again the missing training examples for the mixed class in A.

The combined classification system benefits from the better performance for weeds of the object-based classification approach. A comparison between the corresponding Precision-Recall plots illustrates an improved performance as the precision decreases much smoother while recall increases. This means, that the system is more robust regarding small changes in the parameter t and provides a higher probability for detecting weeds correctly (TNR) while keeping the FNR, i.e. crops classified as weeds, small. The performance for sugar beets is comparable to the keypoint-based approach. In sum, the obtained overall accuracy is around 92% for $t^{\mathcal{O}}_{min} = 0.7$ (see Eq. (19)) and $t = 0.45$ (see Eq. (20)). The dashed line in the ROC and Precision-Recall plot shows the effect of setting $t^{\mathcal{O}}_{min} = 0.4$. Here, the performance of the combined approach is mainly based on the object-classification. Most false classifications of the object-based approach directly affect the performance of the combined approach without a further investigation of the keypoint classifier. This includes around 50% off mixed area, i.e. 7% of all vegetation.

The overall accuracy of the keypoint-based and combined classification system decreases from 96% to 92% when generalizing between the datasets. This is caused by the notable altered visual and physical appearance of the plants and the missing overlaps in A used for training. Figure 14 depicts typical classification errors. Basically, we observe two kinds of error sources. The first one is given by overlapping plants. Due to the fact that dataset A does not provide training samples for those regions, the keypoint classifier is not able to separate them compared to the cross validation on B. The second error source is given by mixed objects, which are wrongly classified as crops by the object-based classier. This error source leads to weeds, which the robot is not able to recognize.

4.3.3 Dataset *A* Predicted with Classifier learned on Dataset *B*

Furthermore, we investigate the reverse experiment to show the performance of our classification system when overlap is not an issue. We train the random forest with all data of B and predict the one week earlier captured dataset A. Here, we compare the performance achieved by the keypoint-based including MRF smoothing vs. the object-based classification neglecting the mixed class as no substantial overlap ($< 2\%$), is given in A.

Figure 15 illustrates the resulting ROC- and Precision-recall plots. In terms of overall accuracy, we achieve 92% for the object-based and 89% for the keypoint-based classification. The results indicate a better per-

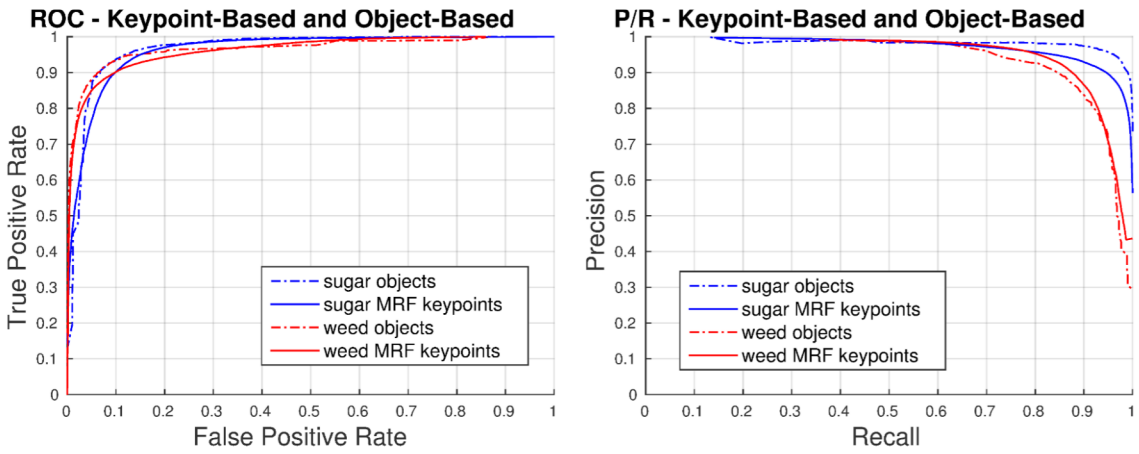

Figure 15: ROC curves (left) and Precision-Recall plot (right) from prediction of dataset A with a random forest classifier learned on dataset B. Comparison of the performance of object-based vs. keypoint-based classification.

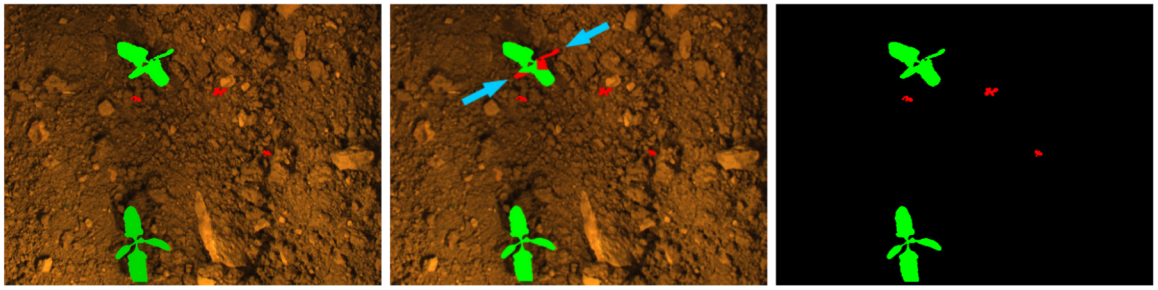

Figure 16: Visual illustration of results of the prediction of dataset A with the classifier learned on dataset B showing typical errors. Left: object-based classification. Middle: Keypoint-based classification including MRF smoothing. Right: Ground truth.

formance for the object-based classification on our data when no substantial overlap is given. Figure 16 depicts some errors of the keypoint-based classification, which are responsible for the lower precision for weeds, compared to the object-based classification.

4.3.4 Dataset C Predicted with Classifier learned on Dataset A

For the next experiment, we increase the difficulty for classification and feed our trained system with images of an unknown field that contains sugar beet of an earlier growth stage in a 2-leaf and partially early 4-leaf stage (dataset C). In most real world applications, one would avoid such situations by providing a classifier trained for the appropriate growth stage but it is worth investigating the loss in performance.

The results are summarized in Figure 17. We obtain a maximum overall accuracy of 80% by a labeling with $t = 0.6$. Obviously, for the keypoint-based classifier the performance further decreases but the achieved TPR of 81% and TNR of 79% indicates that the keypoint-based classification also delivers usable results. Here, the plant arrangement prior and the spatial smoothing through the MRF boosts the performance. As no substantial overlap is given by dataset C, we achieve a TPR of 83% and TNR of 91% by the object-based approach for a labeling with $t = 0.6$. Here, the estimated overall accuracy is given by 87%, which outperforms the keypoint-based classification.

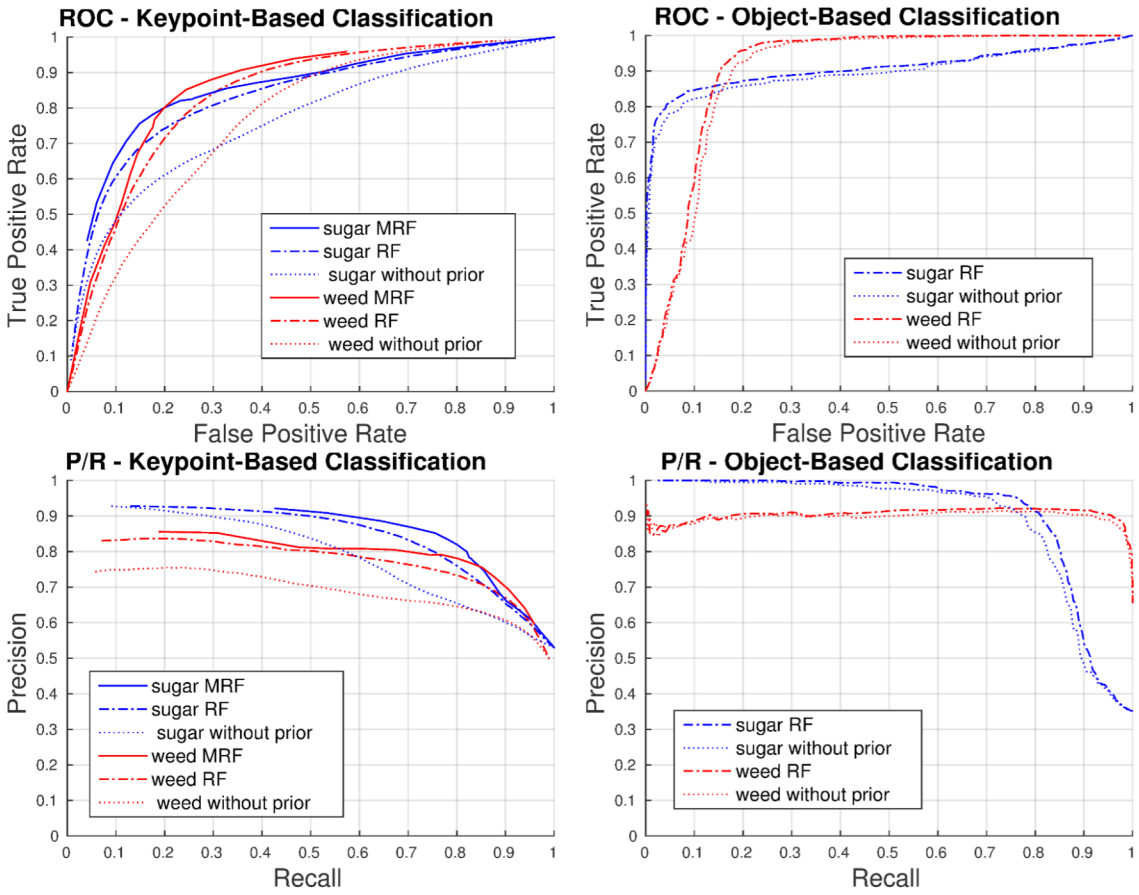

Figure 17: Prediction of dataset C with a random forest classifier learned on dataset A. ROC curves (top) and Precision-Recall plots (bottom). Left: Performance of the keypoint-based approach. Right: Performance of the object-based approach.

Figure 18 illustrates some classification results of both approaches. In terms of sugar beet, the 4-leaf crops are mostly classified correctly and errors are mainly related to the 2-leaf crops. This statement holds for the object-based as well as for the keypoint-based classification. Regarding weeds, the object-based approach performs significantly better, because comparably large areas of weed are classified correctly. In sum, the performance of our classification system is sufficient for selective spraying applications but probably not for mechanical weed removal, as too many crops (especially 2-leaf sugar beets) would be removed by the robot in such challenging settings.

4.4 Feature Importance

This next analysis is designed to investigate which features are important for the classification task. The ten most important features for both, the keypoint-based as well as the object-based approach, are listed in Table 5. As can be seen, the NDVI and Hue information and their gradient as well as texture information are key supporters for the keypoint-based classification task. Regarding the object- based classification, also the most relevant features are related to the NDVI information. Furthermore, the shape features appear highly relevant, probably, as they describe mostly complete plants in this settings. For both approaches, the best 30 features (9% of all used features), hold approximately 40% of the overall feature importance.

In all experiments, we further evaluated the effect of using the relative plant arrangement prior from sowing

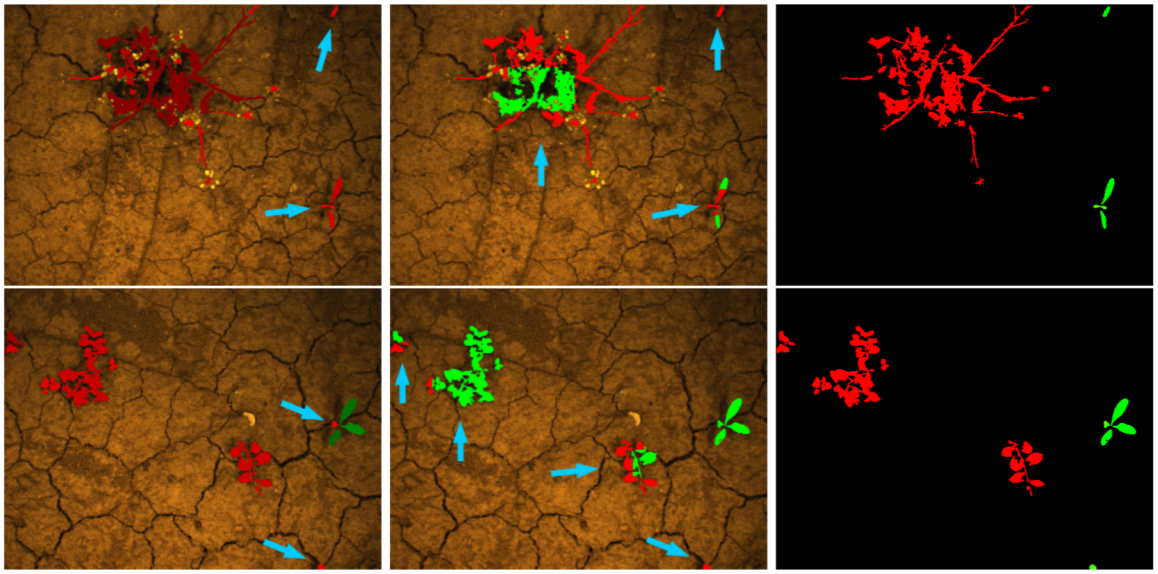

Figure 18: Visual illustration of results of the prediction of dataset C with the classifier learned on dataset A. From left to right: object-based classification, keypoint-based classification and ground truth. Drawn arrows (blue) indicate errors.

or learned from training data. The dotted lines in all plots show the result of the classification system but ignoring the relative plant arrangement information. In most cases, the prior helps to obtain better classification results. In case of the keypoint-based classification, we obtain a gain in average classification accuracy of 1% for the first experiment (Sec. 4.3.1) and 3% for the second (Sec. 4.3.2) experiment. In these experiments, the object-based classification doesn't profit significantly from it. In case of the classification of plants with different growth stage and different species, the impact of the prior information becomes even larger, i.e. an improvement of 9% for the keypoint-based and 3% for the object-based classification in terms of overall accuracy on dataset C. In sum, the results illustrate that using spatial prior information about the crop arrangement is a valuable information for classification systems, especially for the keypoint classifier. This is also documented by Table 5, as the arrangement prior is the third important feature for keypoints.

4.5 Runtime Analysis

For on-field applications in agricultural robotics such as selective spraying or mechanical plant treatment, the runtime performance of the perception system is an important factor. We evaluated the individual runtime of our whole classification pipeline on a standard computer with an Intel i7 CPU and a GeForce GTX 980 GPU. Table 6 illustrates an overview of the execution time of different parts of our classification system. The evaluation is performed on dataset B as it contains the biggest vegetative coverage of our captured data and therefore requires the the largest computational effort, i.e., is the most challenging dataset.

Compared to our conference paper Lottes et al. (2016), we exploit the GPU for the preprocessing and feature extraction part. We use the CUDA framework and through the GPU usage, we achieve a speed-up of 20-100 times for the feature extraction. The average execution time to extract all features is around 0.4 ms for a keypoint and 9.75 ms for an object. Furthermore, we use the GPU for our preprocessing step to obtain normalized intensities for each input channel. Through this, the execution time for this task is around 5 times faster as compared to the CPU implementation, see Table 6. The most time consuming part is currently the MRF smoothing of the keypoint- based approach, which runs on the CPU.

Table 5: The 10 most expressive features over all datasets. Left column: Features for keypoint-based classification. Right column: Features for object based classification. See Table 1 for a description of the features.

Rank	Keypoint Feature	Object Feature
1	$\mathcal{F}_9(\nabla\mathcal{I}_H)$	$\mathcal{F}_3(\Delta\mathcal{I}_{NDVI})$
2	$\mathcal{F}_4(\Delta\mathcal{I}_H)$	$\mathcal{F}_8(\nabla\mathcal{I}_{NDVI})$
3	$\mathcal{F}_{18}$ plant arrangement prior	$\mathcal{F}_{16}$ Formfactor
4	$\mathcal{F}_5(\nabla\mathcal{I}_L)$	$\mathcal{F}_4(\nabla\mathcal{I}_{NDVI})$
5	$\mathcal{F}_3(pLBP(\Delta\mathcal{I}_{NDVI}))$	$\mathcal{F}_7(\nabla\mathcal{I}_L)$
6	$\mathcal{F}_7(pLBP(\Delta\mathcal{I}_{NDVI}))$	$\mathcal{F}_8(\Delta\mathcal{I}_{NDVI})$
7	$\mathcal{F}_{17} \rightarrow \mathcal{F}_9\nabla\mathcal{I}_{NDVI}/\mathcal{F}_9\Delta\mathcal{I}_{NDVI}$	$\mathcal{F}_6(\nabla\mathcal{I}_L)$
8	$\mathcal{F}_3(pLBP(\Delta\mathcal{I}_{NDVI}))$	$\mathcal{F}_3(pLBP(\Delta\mathcal{I}_{NDVI}))$
9	$\mathcal{F}_3(\mathcal{I}_{GREEN})$	$\mathcal{F}_{14}$ aspect ratio
10	$\mathcal{F}_9(pLBP(\mathcal{I}_{NDVI}))$	$\mathcal{F}_9(\nabla\mathcal{I}_H)$

Table 6: Runtime of the individual steps of our classification pipeline in milliseconds.

Function	Mean [ms]	Std [ms]	Max [ms]
general			
Preprocessing (GPU)	21	4	58
Vegetation Detection (CPU)	74	21	190
object-based			
Feature Extraction (GPU)	105	3	239
Classification (CPU)	3	1	13
Overall	**203**		**500**
keypoint-based			
Feature Extraction (GPU)	190	39	401
Classification (CPU)	189	45	301
MRF Smoothing (CPU)	322	150	951
Overall	**796**		**1653**

Given our datasets, we need much more keypoints than objects to cover the vegetation of an image. On average, we can currently provide high quality classification results at 1-2 Hz with our object- based approach and, if we neglect the MRF smoothing, also with our keypoint-based approach. In the worst case of the keypoint-based classification, we are not able to classify an image within 1 second. Regarding the combined classification, the execution time lies in between the object-based and keypoint-based classification timing. For our experiments, the execution time of the combined approach is around 70% of the keypoint- based approach.

4.6 Discussion of the Experimental Results

We provided a large number of real world experiments on sugar beet fields in Germany to evaluate our approach and to support our key claims made in this paper. In summary, the experiments show that our proposed classification system provides high quality results for identifying sugar beets and weeds on the field. We presented two different approaches for the feature extraction and in addition a combined approach, which combines the benefits of the individual approaches with respect to classification performance and runtime.

Given there is no substantial overlap of the plants, the object-based classification shows a slightly better performance in terms of accuracy and is 4 times faster to execute on the same data than the keypoint-based classification. The drawback of the object-based classification is that it is not capable of separating undersegmented objects due to overlap of value crop and weeds. We showed in our first and second experiment that our combined approach is able to identify mixed objects and exploit the keypoint-based classification only for few objects. Thus, the combined approach is faster than the keypoint-based classification as it profits from the fact that the majority of the vegetation is classified well on the object level. The overall execution time shows that our proposed classification system is able to provide usable results within 0.2 seconds for the object-based approach and 0.8 seconds for the keypoint-based approach on average.

Our experiments with classifying changed growth stages of the crop and different weed types illustrate that our classification system provides feasible results when the appearance does not change dramatically. For both approaches, the best 30 features, which corresponds to 9% of all used features, hold 40% of the overall feature importance. This fact offers further potential to limit the extraction to the most relevant features and in this way save further computational effort. To further reduce execution time of the keypoint-based classification, we could (i) apply an adaptive grid size for keypoints to only compute as much as needed, (i) perform the classification (RF) and smoothing (MRF) task also on GPU, (iii) reduce the number of extracted features and (vi) perform MRF smoothing only on local parts of the image.

5 Conclusion

Robots for precision farming must be able to distinguish the crops on the field from weeds. We addressed the problem of detecting sugar beet plants and weeds using a camera installed on a mobile robot operating on a real field. We developed two systems that perform vegetation detection, feature extraction and classification. One performs the classification on object-level, the second one on keypoint-level and performs additional smoothing. Our system uses statistical and shape features computed on the different channels of the image and has the ability to exploit a spatial arrangement prior from sowing. To decide which area of an image corresponds to sugar beets and weeds, we combine random forest classification and in addition exploit the neighboring information through a Markov random field. In order to exploit the advantages of both classifiers, we combined both in cascaded fashion and in this way achieve a high quality classification at 1-2 Hz. We implemented our approach as ROS modules and thoroughly evaluated it on a real farm robot on different sugar beet fields and illustrate that our approach allows for accurately identifying the weed on the field.

Acknowledgments

This work has partly been supported by the European Commission under the grant number H2020-ICT-644227-FLOURISH.

References

Aitkenhead, M., Dalgetty, I., Mullins, C., McDonald, A., and Strachan, N. (2003). Weed and crop discrimination using image analysis and artificial intelligence methods. *Computers and electronics in Agriculture*, 39(3):157–171.

Alaydin, E. (2004). *Introduction to Machine Learning*. MIT Press.

Borregaard, T., Nielsen, H., Norgaard, L., and Have, H. (2000). Crop-weed discrimination by line imaging spectroscopy. *J. Agric. Eng. Res.*, 72(4):389–400.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Burks, T., Shearer, S., Gates, R., and Donohue, K. (2000). Backpropagation neural network design and evaluation for classifying weed species using color image texture. *Transactions of the American Society of Agricultural Engineers*, 43(4):1029–1037.

Cerutti, G., Tougne, L., Mille, J., Vacavant, A., and Coquin, D. (2013). A model-based approach for compound leaves understanding and identification. In *IEEE Intl. Conf. on Image Processing*, pages 1471–1475.

Elhariri, E., El-Bendary, N., and Hassanien, A. E. (2014). Plant classification system based on leaf features. In *Computer Engineering Systems (ICCES), 2014 9th International Conference on*, pages 271–276.

Felzenszwalb, P. F. and Huttenlocher, D. P. (October 2006). Efficient belief propagation for early vision. *Int. Journal on Computer Vision*, 70.

Feyaerts, F. and van Gool, L. (2001). Multi-spectral vision system for weed detection. *Pattern Recognit. Lett.*, 22:667–674.

Hall, D., McCool, C., Dayoub, F., Sunderhauf, N., and Upcroft, B. (2015). Evaluation of features for leaf classification in challenging conditions. In *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, pages 797–804.

Haug, S., Michaels, A., Biber, P., and Ostermann, J. (2014). Plant classification system for crop / weed discrimination without segmentation. In *Proc. of the IEEE Winter Conf. on Applications of Computer Vision (WACV)*, pages 1142–1149.

Hemming, J. and Rath, T. (2001). Computer-vision-based weed identification under field conditions using controlled lighting. *Journal of Agricultural Engineering Research*, 78(3):233 – 243.

Kumar, N., Belhumeur, P., Biswas, A., Jacobs, D., Kress, W., Lopez, I., and Soares, J. (2012). Leafsnap: A computer vision system for automatic plant species identification. In *Proc. of the European Conference on Computer Vision (ECCV)*.

Latte, M. V., Anami, B. S., and Kuligod, V. B. (2015). A combined color and texture features based methodology for recognition of crop field image. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8(2):287–302.

Lottes, P., Hoeferlin, M., Sander, S., Mter, M., Lammers, P. S., and Stachniss, C. (2016). An effective classification system for separating sugar beets and weeds for precision farming applications. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*.

Müter, M., Lammers, P. S., and Damerow, L. (2013). Development of an intra-row weeding system using electric servo drives and machine vision for plant detection. In *Proc. of the Agricultural Engineering Conference*.

Nieuwenhuizen, A. (2009). *Automated detection and control of volunteer potato plants*. PhD thesis, Wageningen University.

Ojala, T. and Pietikinen, M. (1999). Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, (32):477–486.

Ranganathan, A. and Dellaert, F. (2007). Semantic modeling of places using objects. In *Proc. of Robotics: Science and Systems (RSS)*.

Rouse, Jr., J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring Vegetation Systems in the Great Plains with Erts. *NASA Special Publication*, 351:309.

Shearer, S. and Holmes, R. (1990). Plant identification using color co-occurrence matrices. *Transactions of the American Society of Agricultural Engineers*, 33(6):2037–2044.

Stachniss, C., Martínez-Mozos, O., Rottmann, A., and Burgard, W. (2005). Semantic labeling of places. In *Proc. of the Int. Symposium of Robotics Research (ISRR)*, San Francisco, CA, USA.

Tellaeche, A., Burgos-Artizzu, X., Pajares, G., and Ribeiro, A. (2008). A vision-based method for weeds identification through the bayesian decision theory. *Pattern Recognition*, 41(2):521–530.

Wang, X.-F., Huang, D., Du, J., Xu, H., and Heutte, L. (2008). Classification of plant leaf images with complicated background. *Applied Mathematics and Computation*, 205:916–926.