UAV-Based Crop and Weed Classification for Smart Farming

Philipp Lottes Raghav Khanna Johannes Pfeifer Roland Siegwart Cyrill Stachniss

***Abstract*— Unmanned aerial vehicles (UAVs) and other robots in smart farming applications offer the potential to monitor farm land on a per-plant basis, which in turn can reduce the amount of herbicides and pesticides that must be applied. A central information for the farmer as well as for autonomous agriculture robots is the knowledge about the type and distribution of the weeds in the field. In this regard, UAVs offer excellent survey capabilities at low cost. In this paper, we address the problem of detecting value crops such as sugar beets as well as typical weeds using a camera installed on a light-weight UAV. We propose a system that performs vegetation detection, plant-tailored feature extraction, and classification to obtain an estimate of the distribution of crops and weeds in the field. We implemented and evaluated our system using UAVs on two farms, one in Germany and one in Switzerland and demonstrate that our approach allows for analyzing the field and classifying individual plants.**

I. Introduction

Herbicides can have several side-effects on the biotic and abiotic environment and bear a risk to harm human health [7]. Therefore, a reduction of the amounts of herbicides used in modern agriculture is a relevant step towards sustainable agriculture. In conventional weed control, the whole field is typically treated uniformly with a single herbicide dose, spraying the soil, crops and weed in the same way. This practice is simple for the user as neither the knowledge about the spatial distribution nor about the type of the weeds is required and thus is commonly used in conventional farming. The availability of such knowledge, however, offers the potential to reduce the amount of agro-chemical applied to the fields. One way to achieve this is by selectively spraying different weed species as they show variable sensitivities to different herbicides. Another approach consists in the physical control of weeds from a moving platform by destroying them mechanically or with lasers.

Thus, modern approaches as agriculture system management and smart farming typically require detailed knowledge about the current status on the fields. One popular way to monitor farm land is the use of aerial vehicles such as UAVs. Compared to ground vehicles, UAVs can cover large areas in a comparably short amount of time and do not impact the fields through soil compaction as ground vehicles do. For successful on-field intervention, it is important to know the type and spatial distribution of the weeds on the field *at an early state*. The earlier mechanical or chemical weeding actions are executed, the higher the chances for obtaining a high yield. A prerequisite to trigger weeding and intervention task is a detailed knowledge about the spread of weeds. Therefore, it is important to have an automatic perception pipeline that monitors and analyzes the crops automatically and provides report about the status on the field.

Fig. 1: Low-cost UAV used for field monitoring (left) as well as an example image analyzed by our approach (right).

In this paper, we address the problem of analyzing UAV imagery to inspect the status of a field in terms of weed types and spatial crop and weed distribution. We focus on a detection on a per-plant basis to estimate the amount of crops as well as various weed species to provide this information to the farmers. Thus, the main contribution of this paper is a vision-based classification system for identifying crops and weeds in both, RGB-only as well as RGB combined with near infra-red (NIR) imagery of agricultural fields captured by an UAV. Our perception pipeline is capable of detecting plants in such images only based on their appearance. Our approach is able to exploit the geometry of the plant arrangement without requiring a known spatial distribution to be specified explicitly. However, optionally it can also utilize prior information regarding the crop arrangement. In addition to that it can deal with vegetation, which is located within the intra-row space.

We implemented and tested our approach using different types of cameras and UAVs, see Figure 1 for an example, on real sugar beet fields. Our experiments suggest that our proposed system is able to perform a classification of sugar beets and different weed types in RGB images captured by a commercial low cost UAV system. We show that the classification system achieves good performance in terms of overall accuracy even in images where neither a row detection of the crops nor a regular pattern of the plantation can be exploited for geometry-based detection. Furthermore, we evaluate our system using RGB+NIR images captured with a comparable expensive camera system to those obtained using the standard equipped RGB camera mounted on a consumer quad-copter (DJI Phantom 4 and camera for around $1,500).

Philipp Lottes and Cyrill Stachniss are with the University of Bonn, Germany, while Raghav Khanna, Johannes Pfeifer and Roland Siegwart are with the Swiss Federal Institute of Technology, Zurich, Switzerland. This work has partly been supported by the EC under contract number H2020-ICT-644227-FLOURISH.

II. Related Work

UAVs equipped with different sensors serve as an excellent platform to obtain fast and detailed information of arable field environments. Monitoring crop height, canopy cover, leaf area, nitrogen levels or different vegetation indices over time can help to automate data interpretation and thus to improve crop management, see for example [9], [20], [22], [3]. Geipel et al. [3] as well as Khanna et al. [9] focus in their work on the estimation of crop height using UAV imagery. Both these works apply a bundle adjustment procedure to compute a terrain model and perform a vegetation segmentation in order to estimate the crop height based on the obtained 3D information. Tokekar and Hook [22] introduce a concept for a collaboration of an unmanned ground vehicle (UGV) and a UAV in order to measure nitrogen levels of the soil across a farm. The basic idea is to use the UAV for the measurements and the UGV for the transport of the UAV due to its limited energy budget.

Several works have been conducted in the context of vegetation detection by using RGB as well as multispectral imagery of agricultural fields [5], [6], [23]. Hamuda et al. [6] present a comprehensive study about plant segmentation in field images by using threshold based methods and learning based approaches. Torres Sanchez et al. [23] investigate an automatic thresholding method based on the Normalized Difference Vegetation Index (NDVI) and the Excess Green Index (ExG) in order to separate the vegetation from the background. They achieve an accuracy of $90-100\,\%$ for the vegetation detection based on their approach. In contrast, Guo et al. [5] apply a learning approach based on decision trees for vegetation detection. They use spectral features for the classification exploiting different color spaces based on RGB images. We use a threshold based approach based on the ExG and NDVI in order to separate the vegetation from the background, i.e. mostly soil.

The next level of data interpretation is the classification of the detected vegetation by separating it into the classes crop and weed. Several approaches have been proposed in this context. Peña et al. [16] introduced a method for the computation of weed maps in maize fields based on multispectral imagery. They extract super-pixels based on spatial and spectral characteristics, perform a segmentation of the vegetation and detect crop rows in the images. Finally, they use the information about the detected crop rows to distinguish crops and weeds. In a following work by Peña et al. [17], they evaluated a similar approach according to [16] for different flight altitudes and achieve the best performance, i.e. around 90% overall accuracy for crop/weed classification, using images captured at an altitude around $40\,\mathrm{m}$ with a spatial resolution of $15\frac{mm}{px}$. Furthermore, they conclude that using additional near infra-red (NIR) information leads to better results for vegetation detection.

Also machine learning techniques have been applied to classify crops and weeds, in UAV imagery of plantation [8], [4], [18], [19]. Perez-Ortiz et al. [18] propose a weed detection system based on the classification of image patches into the values crop, weed and soil. They use pixel intensities of multispectral images and geometric information about crop rows in order to build features for the classification. They evaluate different machine learning algorithms and achieve overall accuracies of $75-87\%$ for the classification. Perez-Ortiz et al. [19] used a support vector machine classifier for crop/weed detection in RGB images of sunflower and maize fields. They present a method for both inter-row and intra-row weed detection by exploiting statistics of pixel intensities, textures, shape and geometrical information as features. Guerrero et al. [4] propose a method for weed detection in images of maize field, which allows to identify the weeds after its visual appearance changed in image space due to rainfall, dry spell or herbicide treatment. Garcia et al. [8] conducted a study for separating sugar beets and thistle based on multispectral images with a comparably large number of narrow bands. They applied a partial least squares discriminant analysis for the classification and achieved a recall of 84% for beet and 93% for thistle by using four narrow bands at 521, 570, 610 and $658\,\mathrm{nm}$ for the feature extraction. Another noteworthy approach is the one by Mortensen et al. [15]. They apply a deep convolutional neural network for classifying different types of crops to estimate individual biomass amounts. The use RGB images of field plots captured at $3\,\mathrm{m}$ above soil and report an overall accuracy of 80% evaluated on a per-pixel basis.

Our approach extracts visual features as well as geometrical features of the detected vegetation and uses a Random Forest [2] to further classify the vegetation into the value crop and weed. Optionally, we perform a crop row detection and extract an additional feature to integrate this information into the classification system. In the past, we have proposed a classification pipeline [10], [11] for ground vehicles operating in farms. In this paper, we extend our pipeline in terms of (i) geometric features for the use on UAVs, (ii) the ability to deal without sunlight- shields and 100% artificial lighting as used in both previous papers, and (iii) a way to handle RGB images without the requiring NIR information.

III. Plant/Weed Classification for UAVs

The primary objective of the proposed plant classification is to identify sugar beet crops and weeds in UAV imagery in order to provide a tool for an accurate monitoring of the plantation on real fields. The target is to determine a detailed map of the crops and weeds on a per-plant basis in each image, also including weeds located in the intra-row space. We furthermore target the detection of common weed species on sugar beet fields in Northern Europe, which is an important problem and a challenging task for precision farming systems in Germany. The input of the system are either 4-channel RGB+NIR images or regular RGB images, depending on the sensor setup of the UAV. The optional NIR information that is available on more advanced systems, supports the vegetation analysis but is not essential.

The addressed task is strongly related to UGV-based plant classification such as our previous approaches [10], [11]. Us-

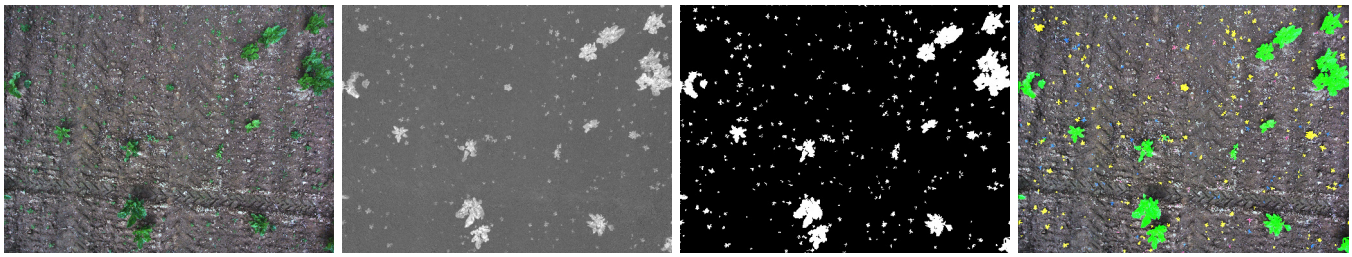

Fig. 2: Example image captured by a *DJI Phantom 4* at 3 m altitude at different levels of progress within the classification pipeline. From left to right: normalized RGB image, computed Excess Green Index (ExG) according to Eq. (2), vegetation mask $\mathcal{V}$ and multi-class classification output of our plant/weed classification system

ing UAV data instead of data recorded with an UGV, is more challenging as the imagery is naturally exposed to varying lightning conditions and different scales. Furthermore, UGV-based systems can exploit more assumptions about the data and are capable for controlled illumination as the sun-light is typically shielded and artificial light-sources are applied. The approach presented in this paper builds upon [10], [11] but extends the classification system, adds more relevant features, can work with RGB-only imagery, and is tailored to UAVs.

The key steps of our systems are the following: First, we apply a pre- processing in order to normalize intensity values on a global scale and detect the vegetation in each image. Second, we extract features only for regions, which correspond to vegetation, exploiting a combination of an object- based [11] and a keypoint-based [10] approach. Third, we apply a multi-class Random Forest classification and obtain a probability distribution for the predicted class labels. We adapt our previous perception pipeline for the crop/weed classification from the ground vehicle to the UAV and introduce the following innovations:

- multi-class detection for discrimination of different weed species,
- classification on RGB+NIR or RGB only imagery, and
- geometric features for exploitation of plant arrangement and for exploiting crop rows if available.

A. Vegetation Detection

The goal of vegetation detection is to remove the background, i.e. the soil or other objects given an image $\mathcal{I}$ by compute the vegetation mask

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

for each pixel location (i,j). Depending on the input data (RGB+NIR or RGB images), we apply different vegetation indices to separate the vegetation parts from the background. In case NIR information is available, we exploit spectral characteristics of plants in the NIR and RED channel through the normalized difference vegetation index (NDVI) according to [21]. In case we only have RGB data available, we rely on the Excess Green Index (ExG) given by

$$\mathcal{I}_{ExG} = 2\,\mathcal{I}_{GREEN} - \mathcal{I}_{RED} - \mathcal{I}_{BLUE} \qquad (2)$$

and compute the masking based on a threshold.

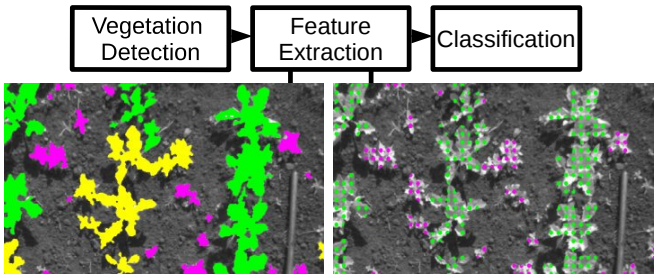

Fig. 3: Pipeline of our Plant/Weed classification system and concept for object-based (left) and keypoint-based (right) approach for the feature extraction. Green refers to sugar beet keypoints and objects, red refers to weed keypoints and objects and yellow refers to mixed objects, which consists of both sugar beet and weed pixels

See Figure 2 for an illustration of the ExG and the obtained vegetation mask by thresholding. Compared to the NDVI, the ExG provides an appropriate index distribution for the identification of the vegetation in RGB images, confirming the results of [12], [14], [9]. However, this approach can lead to wrong segmentations in case of green objects in the image that do not correspond to vegetation. On agricultural fields, however, there are usually only few green objects except plants so that the ExG index is a good choice for our application if no near infra-red channel is available.

B. Objects-based vs. Keypoint-based Feature Extraction

Given the vegetation mask $\mathcal{V}$, we extract features as the input for our classification system. Here, we consider two approaches. The first one computes a single feature vector *per segmented object* from the vegetation mask $\mathcal{V}$, whereas the second approach computes the feature vector on a dense grid of *keypoints* within areas of vegetation, similar to our previous approach described in [11].

Initially, the object-based approach searches for objects, i.e. connected vegetation pixels, in $\mathcal{V}$. Then, we compute a feature vector for each object $\mathcal{O}$ using all pixels that belong to the segment. In contrast, the keypoint-based approach makes no topological assumptions to the vegetation pixels. A keypoint $\mathcal{K}$ is given by its position (i,j) and a certain neighborhood $\mathcal{N}(\mathcal{K})$. We define the positions of the keypoints $\mathcal{K}$ on a grid with a constant lattice distance over the whole vegetation pixels in the image and extract a feature vector for $\mathcal{K}$ taking its neighborhood $\mathcal{N}(\mathcal{K})$ into account. In our current implementation, we use a lattice distance of 10 pixel by 10 pixel for the keypoint and chose 20 pixel by 20 pixel

for the neighborhood $\mathcal{N}(\mathcal{K})$. Figure 3 depicts the concept of both approaches visually. The object-based approach has the advantage that features mostly describe whole plants and comparably less objects are needed to represent the vegetation. The keypoint-based approach has the advantage to deal with under-segmented and overlapping plants. In [11], we showed that combining both approaches in a cascade benefits from the individual advantages respectively. In this work, we use the cascade approach.

For the feature extraction, we start with the same set of statistical and shape features as those described in [11] and extract a set of features $\mathcal{F}$ for each object $\mathcal{O}$ as well as each keypoint $\mathcal{K}$. This set contains statistics regarding the (i) intensity values of the captured images, (ii) their gradient representations, (iii) different color space representations and (iv) texture information.

C. Geometric Features

In addition to the features described in [11], we consider additional geometric features, to exploit the field geometry for the image analysis. Usually, UAV images of plantation capture larger areas compared to UGVs. Thus, they observe a sufficient amount of crops within an image to perform a row detection and to measure spatial relationships among multiple individual plants. We investigate additional geometric features to exploit the fact that crops mostly have a regular spatial distribution without explicitly specifying it. Note that weeds may also appear spatially in a systematic fashion, e.g. in spots or frequently in border regions of the field. First, we perform a line set detection to find parallel crop rows and use distances from potential rows to $\mathcal{O}$ and $\mathcal{K}$ as a feature for the Random Forest classifier. Second, we compute distributions based distances and angles in the local neighborhood around objects and keypoints and extract statistics from it to use them as additional features.

In most agricultural field environments, the plants are arranged in rows, which share a constant inter-row space r, i.e. the distance between two neighboring crop rows. The main goal the line feature

$$f_l = \frac{d}{r} \tag{3}$$

is to exploit the distance d of an object $\mathcal{O}$ or keypoint $\mathcal{K}$ to a crop row. We normalize d by the inter-row space r and use f_l as an additional feature for the classifier. The values d and r are measured in pixels and can be directly obtained in image space. From a mathematical point of view, crop rows can be represented as a finite set of parallel lines

$$\mathcal{L}(\theta, \boldsymbol{\rho}) = \{l_1(\theta, \rho_1), \ldots, l_I(\theta, \rho_I)\}_{i=1}^{I}. \tag{4}$$

where, θ refers to the orientation of the line set and $\boldsymbol{\rho}$ are the distances from each line l_i to the origin of the image frame.

1) Line Feature for Crop Rows: Figure 4 depicts an exemplary result of a detected line set and illustrates the concept of our line-based feature. We introduce the constraint that ρ_i are equidistant to exploit the fact that the inter-row space of crop rows is constant. Note that we do not make any assumptions about the size of r, i.e., the inter-row space. To detect the set $\mathcal{L}(\theta, \boldsymbol{\rho})$ of parallel lines, we employ the Hough transform on the vegetation mask $\mathcal{V}$. This Hough space accumulates the number of votes $v_{\rho,\theta}$, i.e. the number of corresponding vegetation pixels for a given line with the parameters θ and ρ. To compute $\mathcal{L}(\theta, \boldsymbol{\rho})$, we analyze the Hough space and perform the following three steps.

a) Step 1: Estimating the main direction of the crop rows: We compute the main direction $\theta_{\mathcal{L}}$ of the vegetation in an image in order to estimate the direction of the crop rows. This direction can be estimated by considering the votes for parallel lines in Hough space. Here, we follow an approach similar to the one proposed by Midtiby and Rasmussen [13]. To obtain $\theta_{\mathcal{L}}$, they compute the response

$$E(\theta) = \sum_{\rho} v_{\rho,\theta}^2 \tag{5}$$

for each direction and select the maximum $E(\theta)$. The term $v_{\rho,\theta}$ refers to the number of votes for a certain line with the parameters θ, ρ. In contrast to [13], we do not only select the maximum of $E(\theta)$ but consider the N best values for $E(\theta)$ to evaluate the N best voted directions in Hough space given the vegetation in the subsequent steps. In our implementation, we use $N = 15$. We consider the 15 best main directions supported by the vegetation in order to handle scenarios with large amounts of weed. Tests under high weed pressure show that the maximum response is not always the correct choice as many weed plants may lead to more votes for a false detection of the rows.

b) Step 2: Estimating the crop rows as sets of parallel lines: Given the N best voted orientations of possible line sets from Step 1, we want estimate in which direction we find the best set of parallel lines with equidistant spacing.

We search for an unknown but constant spacing r between neighboring lines as well as the offset s of the first potential crop row in image space, see Figure 4 for an illustration. Thus, we search for the maximum response of

$$E(\theta, r, s, L_r) = P + \sum_{r=1}^{R} \sum_{s=0}^{r} \sum_{l=0}^{L_r - 1} v_{(s + l\,r),\theta}, \tag{6}$$

with the penalty term

$$P = -L_r\, \bar{v}_\theta, \tag{7}$$

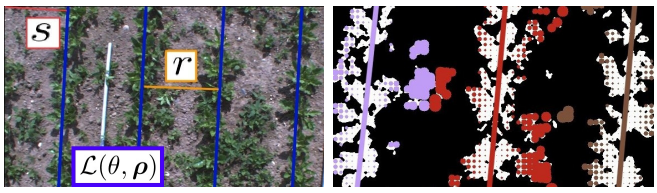

Fig. 4: Left: Result of the line set detection. s (red) refers to the distance of the first line in the set $\mathcal{L}(\theta, \boldsymbol{\rho})$ to the origin of the image frame. r (orange) refers to the inter-row space. Right: visual illustration of the line model feature f_l for the keypoint-based approach. The different colors refer to different lines of the detected line set. The value of f_l is encoded by the radius of a keypoint. Weeds located within the inter-row space get a higher value in f_l.

by varying the size of r and s. The term L_r refers to the number of lines that intersect with in the image for a given r. The penalty term P is an additional cost term that is introduced for each line of the set in order to penalize an increasing number of lines. Here, $\bar{v}_\theta$ is the mean response over the column corresponding to $\theta_\mathcal{L}$ in the Hough space. This leads to the effect that E, according to Eq. (6), increases for lines, which have a better response $v_{s+l\,r,\theta} > \bar{v}_\theta$ and vice versa decreases if the response is lower. The maximum response according to Eq. (6) provides the best voted line set, which has a constant inter-row space.

c) Step 3: Refitting the best line set to the data: Crops are commonly sown out in fixed assemblages of a certain number of rows, called plots. It can happen that the inter-row space between plots slightly differs due to the limited position accuracy of the sowing machine. To overcome this, we finally fit each line of the best set obtained from step 2 to the data by using a robust estimator based on a Huber kernel and obtain a robust estimate $\mathcal{L}(\theta, \boldsymbol{\rho})$ for the crop rows.

2) Spatial Relationship Features: In order to describe spatial relationships among individual plants, we first compute the distances and azimuths from a query object $\mathcal{O}_q$ or keypoint $\mathcal{K}_q$ to all other nearby objects or keypoints in world coordinates (which requires to know the flying altitude of the UAV). We compute the differences of the measured distances between the query object and its neighbors to obtain a distribution in form of a histogram. Similarly, we obtain the distribution over angles from the observed azimuths. From these distributions, we compute common statistical qualities such as *min, max, range, mean, standard deviation, median, skewness, kurtosis* and *entropy* and use them as features for the classifier. In addition to that, we count the number of vegetation objects $\mathcal{O}$ or keypoints $\mathcal{K}$ in their neighborhood in object space.

Both the spatial relation features as well as the line features allow for encoding additional geometric properties and in this way to improve the random forest classifier used to make the actual decision.

D. Random Forest Classification

We apply a rather standard random forest [2], which is a comparably robust ensemble method, that is capable of solving multi-class problems. The central idea is to construct a large number of decision trees at training time by randomizing the use of features and elements from the training data. It also allows for implicitly estimating confidences for the class labels by considering the outputs of the individual decision trees.

IV. Experiments

This evaluation is designed to illustrate the performance of our UAV-based plant/weed classification system and to support the main claims made in this paper. These claims are: (i) Our classification system identifies sugar beets as well as common weed species on sugar beet farms in RGB imagery captured by an UAV, even if the crop is not sowed in rows and the line detection cannot support the classification, (ii) we demonstrate that our plant classification system is able to exploit an arrangement prior about crop rows using our proposed line model and in addition to that benefits from geometric features, which capture spatial relationships in a local neighborhood, and (iii) our method provides good results under challenging conditions such as overlapping plants and is capable for weed detection in intra-row space. Finally, we analyze the effect of exploiting additional NIR-channel in terms of vegetation detection and classification performance.

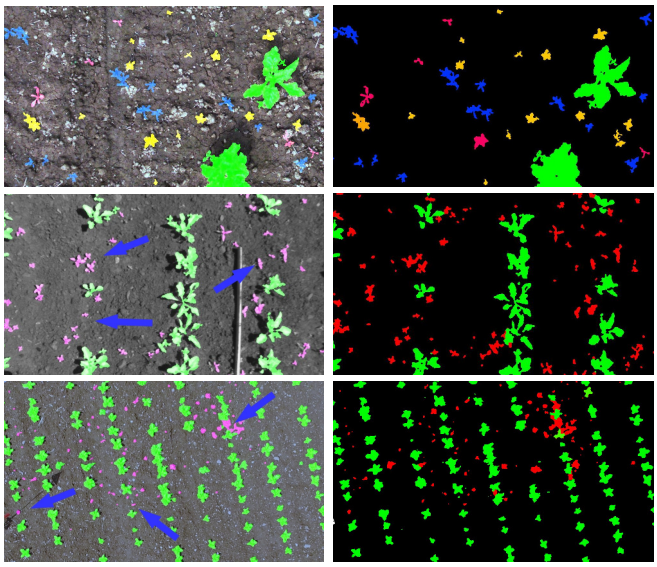

Fig. 5: Zoomed view of images analyzed by our plant/weed classification system (left column) and corresponding ground truth image (right column). Top row: Multi-class results for the *PHANTOM-dataset*. Center row: Result for the *JAI-dataset*. Bottom row: Result for the *MATRICE-dataset*. Arrows point to weeds detected in intra-row space. Sugar beet (green), chamomile (blue), saltbush (yellow) and other weeds (red)

A. Evaluation Metric

We illustrate the performance of the classification results by ROC curves and precision-recall (PR) plots. For these plots, we varied the threshold for the class labeling concerning the estimated confidences of the random forest. The parameters influencing the performance of the crop/weed classification are exclusively evaluated on the detected vegetation parts of the images in order to analyze the quality of the classification output. We labeled all the plants manually in the images and perform the evaluation on object-level, i.e. predicted object vs. ground truth object, in order to obtain a performance as close as possible to a per-plant basis. For the keypoint-based classification we compute the class-wise ratios of predicted keypoints concerning the ground truth object to keep the evaluation metric on the object-level.

B. Datasets

We collected three different datasets, with different characteristics and challenges, to thoroughly evaluate the per-

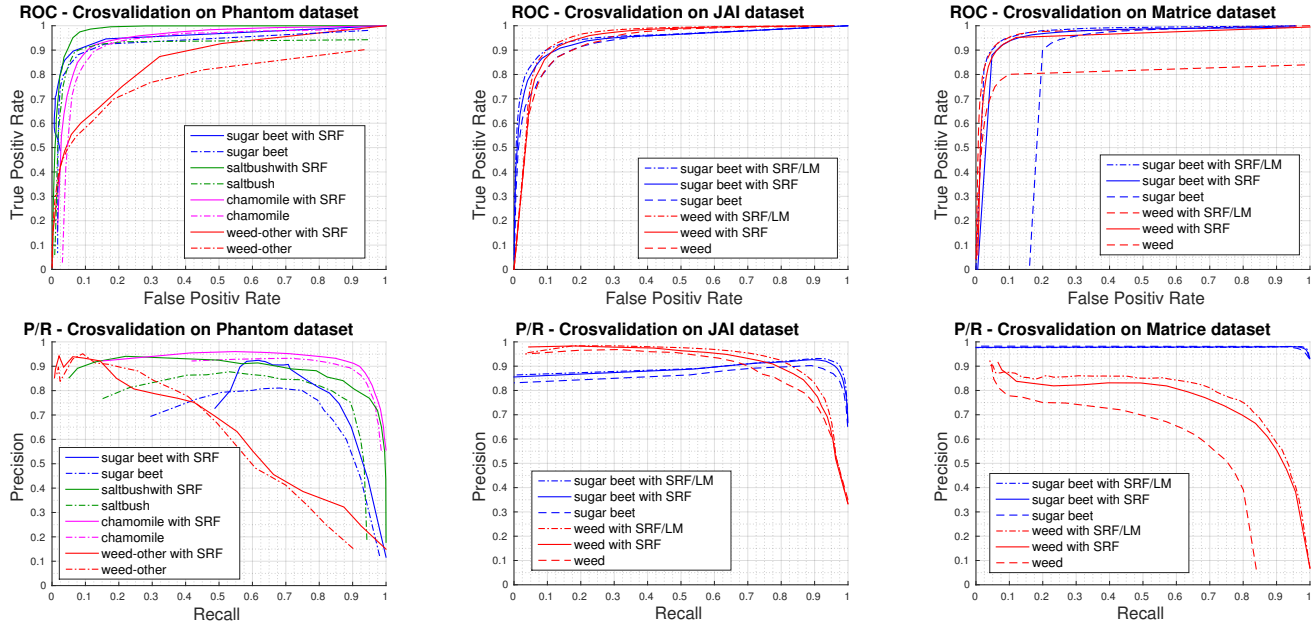

Fig. 6: ROC curves (top) and PR plots (bottom) for each dataset based on a leave-one-out cross-validation scheme according to [1]. Left: Multi-class performance evaluated on the *PHANTOM-dataset*. Middle: Performance of our approach on the *JAI-dataset*. Right: Performance of our approach on the *MATRICE-dataset*. The term label "SRF" refers to the additional use of the spatial relationship features and "LM" to the line model feature.

TABLE I: Information about the datasets.

Parameter	JAI	MATRICE	PHANTOM
# images	97	31	15
flight mode	manual	way-point-mission	manual
weather conditions	sunny	sunny/cloudy	overcast
$\approx$ flight altitude	2-3 m	15 m	3 m
$\approx$ ground resolution	2 $\frac{mm}{px}$	5 $\frac{mm}{px}$	0.2 $\frac{mm}{px}$

formance of our plant/weed classification system. All the recorded datasets represent real field conditions and contain sugar beet crops and weeds.

The *JAI-dataset* is captured with a 4-channel JAI AD-130 GE camera on a sugar beet farm in Eschikon, Switzerland. The dataset contains 97 RGB+NIR images of sugar beets arranged in crop rows and weeds captured at a weekly basis over 1.5 months in May 2016. The growth stage of the sugar beet is from early 4-leaf up to late 6-leaf stadium. The camera was pointing downwards to the field at a hight of approximately $2-3$ m above the soil. The image resolution of 1296×966 pixels in combination with Fujinon TF4DA 8 lens with 4 mm focal length yields a ground resolution of roughly 2 $\frac{mm}{px}$. The *MATRICE-dataset* is captured by the *Zenmuse X3* camera of the *DJI Matrice 100* UAV. The data has been recorded on two days with one week temporal difference in May 2016 on the same field as the *JAI-dataset*. The dataset contains 31 RGB images where the crop is in early 6-leaf growth stage. We flight with the UAV roughly 15 m above the soil. The achieved ground resolution is around 5 $\frac{mm}{px}$. The *JAI-dataset* and the *MATRICE-dataset* provide the basis to evaluate the line feature and to analyze the performance for the crop/weed classification. The dataset contains weeds that are located within both inter and intra-row space. In these datasets the amount of weeds is much smaller compared to the crop. The *PHANTOM-dataset* is captured on a field in Bonn, Germany, where the sugar beets are not sowed in crop rows. The dataset provides images obtained by an unmodified consumer *DJI Phantom 4* UAV. The obtained ground resolution of 0.2 $\frac{mm}{px}$ is comparably high as the images were captured with a resolution of 4000×3000 pixels at a flight altitude of 3 m. Due to the resolution, we can visually identify typical weeds on common sugar beet fields, i.e. *saltbush* (Atriplex) as a common problem weed in terms of mass, *chamomile* (Matricaria chamomilla), and *other weeds*.

C. Classification of Crops and Different Weed Types

This first experiment in our evaluation is designed to demonstrate that our system is capable for classifying sugar beets and weed species, which are common on sugar beet farms. Therefore, we analyze the performance on the *PHANTOM-dataset* for the classification of sugar beets, *saltbush*, *chamomile* and *other weeds*. Figure 6 (left column) depicts the ROC and PR plots computed in a one-vs-all mode to illustrate the obtained performance.

The ROC curve shows that for all explicitly specified classes a recall around 90%, for *saltbush* even 95%, can be obtained depending on the selected threshold. The class labeling based on the predicted maximum confidences of the random forest leads to the following results. The system achieves a recall of 95% for *saltbush* and 87% for *chamomile*. Both species are predicted with a precision of 85%. For sugar beet a recall of 78% with a precision of 90% is obtained. Generally, the precision suffers from the obtained recall for *other-weeds* of 45%. This result is affected by (i)

having a small number of examples within the datasets and (ii) probably by a higher intra-class variance since all other weeds, which occur in this dataset, are represented by this class. More datasets with different weed types are needed to clarify that. In terms of overall accuracy, 86% of the predicted objects and 93% of the area are classified correctly.

If we only focus on crop vs. weed classification, the overall accuracy increases by 11% to 96% for the detection on object-level. For sugar beets, the performance do not change but the fusion of the weed species leads to a recall of 99% with an obtained precision of 97%.

D. Impact of Geometric Features

We designed the second experiment in order to illustrate the impact in performance when using geometric features. Therefore, we evaluated all datasets with the usage of all geometric features, i.e. the line feature and the spatial relationship features, and compare the results with the performance neglecting them. Figure 6 illustrates the results for each dataset by using all geometric features ("with SRF/LM") using only the spatial relationship features ("with SRF") without exploiting the line feature.

Our evaluation shows that the use of geometric features supports the classification based on visual features as it improves the overall accuracy and the precision, especially for weeds, in all tested datasets. Even for the *PHANTOM-dataset*, which not involves any regular pattern of the plants, the performance benefits from the use of the spatial relationship features. In numbers, the gain for *saltbush* is about 6% for both precision and recall. For all other classes the recall raises of around 3% on average for a given precision

For the *JAI-dataset* and *MATRICE-dataset*, we measured the effect of the spatial relationship features and all geometric features, including the line model feature, by an individual cross-validation on the whole dataset respectively. The biggest gain in performance can be observed for the *MATRICE-dataset*. Here, the detection based only on visual appearance, i.e. features ignoring geometry, suffers from the comparably low ground resolution. Thus, geometric features become great supporters for the detection as they are rather invariant to the image resolution. On average, the spatial relationship features are responsible for an increase in performance of around 10% and in combination with the line feature of 13% in terms of overall accuracy. The corresponding PR plot indicates that this amount is mainly caused by better detection. This statement also holds for the *JAI-dataset*. The PR plot illustrates a significant gain in recall of around 5% for weeds even when the precision is greater than 85%. In sum, the effect of the geometric features is smaller as for the *MATRICE-dataset* and the *PHANTOM-dataset*, i.e. 5%. That is because the classification based on pure visual appearance performs better and is more stable with respect to a varying threshold for the class labeling. We conclude that using geometric features for the classification task is an appropriate way to exploit spatial characteristics of plantation in agricultural field environments.

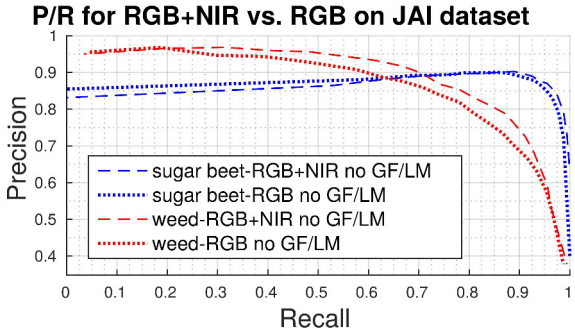

Fig. 7: PR plot obtained by a leave-one-out cross validation on the *JAI-dataset* by using only RGB and RGB+NIR information.

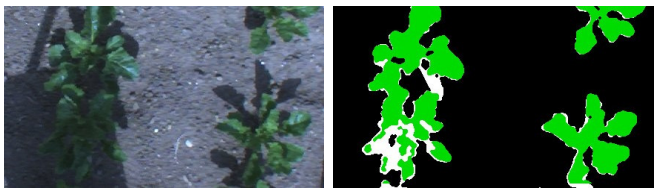

Fig. 8: Left: Detailed view on RGB image containing sugar beets. Right: Masking based on NDVI (white) and ExG (green). All vegetation pixels detected by ExG based approach are also detected by the NDVI based approach

E. Intra-Row Weed Detection

Most existing approaches that exploit the prior knowledge about the crop rows perform the following two steps sequentiality: (i) detect the rows and (ii) use geometry to force vegetation corresponding to the rows to be weeds. In contrast to that, our approach is suited for both exploiting the geometry and identifying intra-row-weeds. The usage of f_l as feature in the random forest leads to the fact that vegetation located within intra-row space still has the chance to get detected as weed based on its visual features.

Figure 5 visually illustrates typical examples of the performance of our approach. A visual comparison of example images from the *JAI-dataset* (middle row) and the *MATRICE-dataset* (bottom row) against the corresponding ground truth image shows that weed are correctly detected in intra-row space. This results originates from the classification using also the geometric features.

F. Effect of the Availability of NIR Information

In our last experiment, we evaluate the effect of using additional NIR information for the vegetation detection as well as for the crop/weed classification. We separate this experiment into two parts. First, we evaluate the performance for the vegetation detection given RGB+NIR and RGB only images and compare it with the ground truth. Second, we analyze the impact of additional NIR information for the classification. For this part, we extract features based on RGB and RGB+NIR images receptively by using the mask obtained by the ExG in order avoid effects of the masking to the considered shape features, see [11]. Finally, we neglect the use of geometric features for this experiment in order to rely only on visual information.

TABLE II: Recall for objects consisting of at least 50 pixels.

index	NDVI	ExG	ExGR	NGRDI	CIVE
recall	92%	86%	83%	83%	81%

We tested several commonly applied vegetation indices as basis for a threshold based vegetation segmentation, see [6], [9], [23], on the *JAI-dataset*, where both RGB and NIR images are available. To evaluate the segmentation performance on object-level, we compare the obtained vegetation masks $\mathcal{V}_{index}$ with the labeled data. We define an object as detected, if 75% of its pixels are detected as vegetation pixel. Table II gives an overview of the detection rate for vegetation objects by using NDVI, ExG, Normalized GreenRed Difference Index (NGRDI), Excess Green minus Excess Red Index (ExGR) and Color Index of Vegetation Extraction (CIVE)..

Using the NIR information to exploit the NDVI for the detection outperforms all RGB-based methods. Figure 8 depicts a typical result of $\mathcal{V}$ when some parts in the image are shaded. Visually, the NDVI- based masking (white) gives better results compared to the ExG-based (green) vegetation detection. This is because the NIR-channel observes a comparably high reflectance for vegetation pixels, even in shaded areas. Given only RGB data, we use the ExG as it is the best choice given our data.

Figure 7 depicts the precision recall plots for the performance of the classification using the additional NIR information (RGB+NIR) as well as only relying on RGB. Using the NIR-channel for the features increases the performance around 3% in terms of overall accuracy. To conclude, using the additional NIR leads to better classification results and provides more robustness in terms of vegetation segmentation in shaded image regions but our approach is also operational if the NIR information is missing.

V. Conclusion

UAVs used in precision farming applications must be able to distinguish crops from weeds on the field to estimate the type and distribution of weeds. In this work, we focus on sugar beet plants and typical weeds observed on fields in Germany and Switzerland using a regular RGB camera as well as a RGB+NIR camera. We described a system for vegetation detection, feature extraction, and classification for aerial images relying on object-features and keypoints in combination with a random forest. Our system can exploit the spatial arrangement of crops through geometric features. We implemented our approach and thoroughly evaluated it on UAV derived data captured on two farms and illustrate that our approach allows for identifying the crops and weeds in the field, which is an important information for several precision farming applications.

References

[1] E. Alaydin. *Introduction to Machine Learning*. MIT Press, 2004.

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

[3] J. Geipel, J. Link, and W. Claupein. Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. *Remote Sensing*, 6(11):10335, 2014.

[4] J. M. Guerrero, G. Pajares, M. Montalvo, J. Romeo, and M. Guijarro. Support vector machines for crop/weeds identification in maize fields. *Expert Systems with Applications*, 39(12):11149 – 11155, 2012.

[5] W. Guo, U. K. Rage, and S. Ninomiya. Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model. *Computers and Electronics in Agriculture*, 96:58–66, 2013.

[6] E. Hamuda, M. Glavin, and E. Jones. A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, 125:184–199, 2016.

[7] L. Horrigan, R. S. Lawrence, and P. Walker. How sustainable agriculture can address the environmental and human health harms of industrial agriculture. *Environ Health Perspect*, 110:445–56, 2002.

[8] G. R. F. Jose, D. Wulfsohn, and J. Rasmussen. Sugar beet (beta vulgaris l.) and thistle (cirsium arvensis l.) discrimination based on field spectral data. *Biosystems Engineering*, 139:1–15, 2015.

[9] R. Khanna, M. Möller, J. Pfeifer, F. Liebisch, A. Walter, and R. Siegwart. Beyond point clouds - 3d mapping and field parameter measurements using uavs. In *Proc. of the IEEE Int. Conf. on Emerging Technologies Factory Automation (ETFA)*, pages 1–4, 2015.

[10] P. Lottes, M. Höferlin, S. Sander, M. Müter, P. Schulze-Lammers, and C. Stachniss. 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)*, 2016.

[11] P. Lottes, M. Höferlin, S. Sander, and C. Stachniss. Effective vision-based classification for separating sugar beets and weeds for precision farming. *Journal of Field Robotics*, 2016.

[12] G. E. Meyer and J. C Neto. Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, 63(2):282 – 293, 2008.

[13] H. S. Midtiby and J. Rasmussen. Automatic location of crop rows in uav images. In *NJF Seminar 477. Future arable farming and agricultural engineering*, pages 22–25, 2014.

[14] M. Montalvo, G. Pajares, J. M. Guerrero, J. Romeo, M. Guijarro, A. Ribeiro, J. J. Ruz, and J. M. Cruz. Automatic detection of crop rows in maize fields with high weeds pressure. *Expert System Applications*, 39(15):11889–11897, 2012.

[15] A. K. Mortensen, M. Dyrmann, H. Karstoft, R. N. Jörgensen, and R. Gislum. Semantic segmentation of mixed crops using deep convolutional neural network. In *Proc. of the International Conf. of Agricultural Engineering (CIGR)*, 2016.

[16] J. M. Peña, J. Torres-Sánchez, A. I. de Castro, M. Kelly, and F. López-Granados. Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle uav images. *PLoS ONE*, 8, 2013.

[17] J. M. Peña, J. Torres-Sánchez, A. Serrano-Perez, A. I. de Castro, and F. López-Granados. Quantifying efficacy and limits of unmanned aerial vehicle uav technology for weed seedling detection as affected by sensor resolution. *Sensors*, 15(3), 2015.

[18] M. Perez-Ortiz, J. M. Peña, P. A. Gutierrez, J. Torres-Sánchez, C. Hervás-Martínez, and F. López-Granados. A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Applied Soft Computing*, 37:533 – 544, 2015.

[19] M. Perez-Ortiz, J. M. Peña, P. A. Gutierrez, J. Torres-Sánchez, C. Hervás-Martínez, and F. López-Granados. Selecting patterns and features for between- and within- crop-row weed mapping using uav-imagery. *Expert Systems with Applications*, 47:85 – 94, 2016.

[20] J. Pfeifer, R. Khanna, C. Dragos, M. Popovic, E. Galceran, N. Kirchgessner, A. Walter, R. Siegwart, and F. Liebisch. Towards automatic UAV data interpretation for precision farming. *Proc. of the International Conf. of Agricultural Engineering (CIGR)*, 2016.

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

[22] P. Tokekar, J. V. Hook, D. Mulla, and V. Isler. *Sensor planning for a symbiotic UAV and UGV system for precision agriculture*, pages 5321–5326. 2013.

[23] J. Torres-Sanchez, F. López-Granados, and J.M. Peña. An automatic object-based method for optimal thresholding in uav images: Application for vegetation detection in herbaceous crops. *Computers and Electronics in Agriculture*, 114:43 – 52, 2015.