Semi-Supervised Online Visual Crop and Weed Classification in Precision Farming Exploiting Plant Arrangement

Philipp Lottes Cyrill Stachniss

***Abstract*— Precision farming robots offer a great potential for reducing the amount of agro-chemicals that is required in the fields through a targeted, per-plant intervention. To achieve this, robots must be able to reliably distinguish crops from weeds on different fields and across growth stages. In this paper, we tackle the problem of separating crops from weeds reliably while requiring only a minimal amount of training data through a semi-supervised approach. We exploit the fact that most crops are planted in rows with a similar spacing along the row, which in turn can be used to initialize a vision-based classifier requiring only minimal user efforts to adapt it to a new field. We implemented our approach using C++ and ROS and thoroughly tested it on real farm robots operating in different countries. The experiments presented in this paper show that with around 1 min of labeling time, we can achieve classification results with an accuracy of more than 95% in real sugar beet fields in Germany and Switzerland.**

I. Introduction

Moving towards a sustainable agriculture is one of the 17 sustainable development goals of the United Nations. A step towards this goal is to reduce the reliance on herbicides and pesticides that must be applied in our fields. Precision farming robots have the potential to contribute to this goal through a targeted, plant specific intervention, for example through selective spraying, mechanical or laser-based weed control applications. In order to treat plants on an individual level, it is essential that a robot is able to distinguish weeds from the value crop planted on the field.

In the past, different methods have been proposed in the context of vision-based crop and weed detection [4], [5], [8], [9], [10], [11], [15]. Most approaches employ supervised learning algorithms to address the detection problem and rely on feature distributions, which encode the visual appearance of the plants. Typically, they report classification performances within a range of 80-95% for the overall accuracy. However, most systems have not been thoroughly evaluated under substantial changes in the appearance of the plants. Especially if during the learning and the operational phase, the robots operates on different fields and/or growth stages of the plants, the performance often drops substantially. Slaughter *et al.* [13] conclude in their review about robotic weed control systems that the lack of *robust* crop/weed detection technology is the main limitation to the commercial development of such a system. A key reason for the lack of robustness is the change of the underlying feature distribution induced by a change in the appearance of the plants between training time and operational phase of the detection system.

All authors are with the University of Bonn, Institute of Geodesy and Geoinformation, Bonn, Germany. This work has partly been supported by the EC under the grant number H2020-ICT-644227-Flourish.

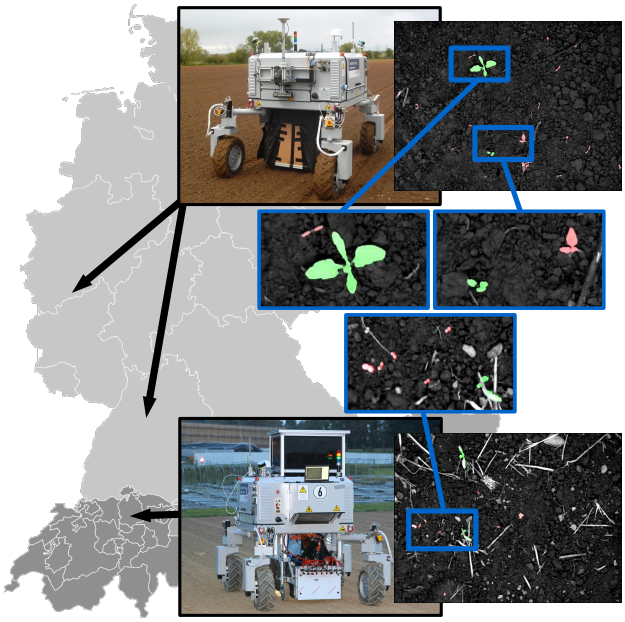

Fig. 1. We aim at keeping the classification performance of the crop/weed detection system high, even if the training and the operational phase of the classifier are executed on different fields and in different countries as depicted in the image above. We achieve this while requiring only minimal re-training efforts through a semi-supervised approach. The example images are analyzed by our approach containing sugar beets (green) and weeds (red) detected with a size of 0.5-2 cm^2

In this work, we propose a method to bridge the performance gap in visual crop and weed detection if the distribution of the features at training differs from the one observed during operation. The targeted use-case is sketched in Fig. 1: A robot has been trained on one sugar beet field (here near Bonn, Germany) but is used somewhere else (here near Zurich, Switzerland). To achieve this, we exploit the fact that a large number of crops are seeded in rows. Sugar beet plants, for example, are arranged in rows and often share a similar lattice distance along these rows. Such geometric information is typically similar within a field and in addition independent of the visual appearance of the plants. Thus, we exploit it to support and retrain the vision-based classifier in a semi-supervised way.

The main contribution of this paper is a semi-supervised online approach for the vision-based classification of crops and weeds by exploiting additional arrangement information of the crops in order to adapt the visual classifier. We realize this by combining the visual and geometric classifiers that complement each other through independent predictions. We furthermore adapt the visual classifier in an online manner

by using the predictions of the geometric classifier. We use a probabilistic model representing the arrangement of the plants and employ a Bayesian approach to perform the crop and weed classification based on that model. The plant arrangement model can be learned from a comparably small amount of data and initializes the visual classification system, which is then updated as new data is processed. We can even handle plants at a small growth stage (size of 0.5-2 cm^2), which is attractive for targeted weed control as an early intervention is typically desired.

We make three key claims in this paper: (i) Our approach is able to identify sugar beets and weeds with an accuracy of above 95% by only using a very small amount of labeled data to initialize the system, which can be achieved within one minute of human effort. (ii) We show that our approach is able to adapt a previously trained visual classifier to the current situation on the field, where the distribution of the visual features has notably changed. (iii) Our approach provides suitable results for online operation of a robotic weed control system with a processing time of 5 Hz. These three claims are explicitly backed up our experimental evaluation.

II. Related Work

Several approaches have been developed in the context of supervised crop and weed classification based on visual features such as [4], [5], [6], [7], [9], [10], [15]. A recent work by Mortensen *et al.* [9] employ a CNN on images of field plots containing crops and different weeds. They aim at estimating individual biomass amounts based on a pixel-wise classification and report an overall accuracy around 80%. Nieuwenhuizen [10] presents an approach on the automated detection and control of volunteer potato plants in sugar beet fields based on a Bayesian classifier. Haug *et al.* [4] proposes a method to distinguish carrot plants and weeds in RGB+NIR images. They obtain an average accuracy of 94% on an evaluation dataset containing 70 images. Hemming and Rath [5] propose a vision-based approach to classify carrots, cabbage, and weeds based on a fuzzy logic classifier. They evaluate their approach using field experiments and achieve an overall performance between 72% and 88%. In our previous work [6], [7], we presented a vision-based classification system based on RGB+NIR images for sugar beets and weeds. We combine appearance and geometric features using a random forest classifier and obtained classification accuracies of up to 96% on pixel level. The work also shows that the performance decreases to an unsuitable level for weed control applications when the appearance of the plants changes substantially. This submission builds on top of our previous work [7] by presenting an approach to initialize and re-train the classification system effectively and to adapt the classifier online.

Hall *et al.* [3] also address the issue of changing feature distributions. They evaluate different features for leaf classification by simulating real-world conditions with respect to translation, scaling, rotation, shading, and occlusion. They compare the obtained performance by selectively using different handcrafted and ConvNet features and conclude that ConvNet features can support the robustness and generality of a classifier. Strothmann *et al.* [14] use a multi-wavelength line scanner for crop and weed classification using a Bayesian approach. To adapt their classifier they label a small amount of data covering the actual feature distribution and retrain their classifier.

A further challenge with the supervised classification approaches is the necessary amount of labeled data, which is typically obtained at high cost. Wendel and Underwood [16] address this issue by proposing a method for training data generation in order to feed a classifier for crop and weed detection with it. They use a multi-spectral line scanner mounted on a field robot and perform a vegetation segmentation followed by a crop-row detection. Subsequently, they assign the label crop for the pixels corresponding to the crop-row and the remaining ones as weed. Rainville *et al.* [12] propose a vision-based method to learn a probability distributions of morphological features based on a previously computed crop-row. Similar to [16], they also exploit crop-row structure as a prior for labeling.

Di Cicco *et al.* [2] try to reduce the human effort for labeling by constructing synthetic datasets using a physical model of a sugar beet leaf. Their results indicate that such artificially generated datasets can support traditional approaches by providing additional training data. Potena *et al.* [11] reduce the required amount for labeling through an unsupervised dataset summarization before the actual labeling process takes place. The key idea is to select a subset of a fixed size, which gives the most informative description the whole dataset.

To the best of our knowledge, there is no similar approach for crop/weed detection that uses a geometry-based and vision-based classification system to support and initialize each other, that requires minimal human training efforts in the order of 1 min and that provides classification rates of more than 95%.

III. Vision-Based Crop and Weed Classification

This Sec. III briefly summarized our recently published approach [7] for the vision-based online classification of crops and weeds by using RGB+NIR data for field robots. As an illustration, Fig. 2 depicts the RGB+NIR input images and different processing stages of the visual classifier. We use a three step pipeline for providing a per-pixel labeling of the input images: (i) threshold-based vegetation detection exploiting spectral characteristics of the plants, (ii) extraction of visual features for the detected vegetation using the object-based approach described in [7], and (iii) random forest classification. Note that we do not claim any novelty for this section in this submission but provide it to allow the reader to understand the paper as a whole.

A. Vegetation Segmentation

The first step is to separate the vegetation from the background, which is mostly soil. We achieve this by computing the popular Normalized-Difference-Vegetation-Index (NDVI) and the Excess-Green-Index (ExG) from the

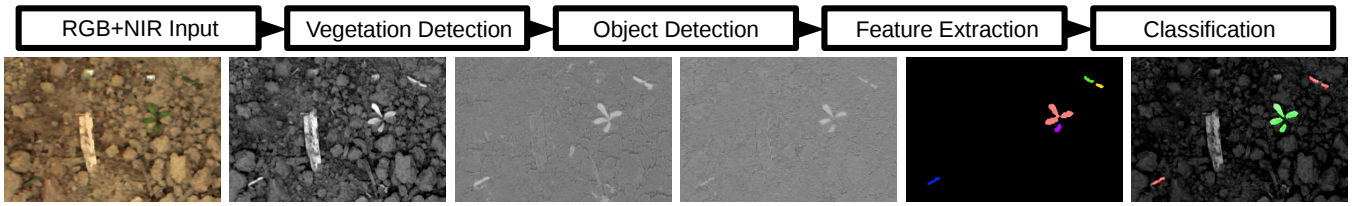

Fig. 2. Pipeline for the vision-based classification. From left to right: the Input (1) RGB and (2) NIR image, computed (3) NDVI and (4) ExG image. NDVI and ExG are shown for illustration reasons to have values between 0 and 255, (5) shows the obtained vegetation mask containing the segmented vegetation objects Ψ, where different colors refer to individual objects, and (6) the classification output containing sugar beets (green) and weeds (red).

different channels, see also [8]. A simple threshold operation to the NDVI distribution combined with $ExG > 0$ for a pixel to be considered as vegetation is sufficient to obtain a high quality segmentation. Fig. 2 depicts the computed NDVI and ExG images as well as the obtained vegetation mask.

B. Object-Based Feature Extraction

We compute the 100 most expressive features of [7] for each segmented object (called "object-based approach" in [7]). An object Ψ is identified through connected vegetation pixels and we perform the feature extraction, i.e. one feature vector per object, based on all pixels corresponding to Ψ at ones. The visual feature contains statistics regarding the intensity values of the captured images, their gradients, different color spaces, and texture information.

C. Random Forest Classification

We employ a random forest [1] for the vision-based classification of the detected Ψ into crops or weeds. Random forests are an ensemble method combining a large number of decision trees. The key idea is to construct a large number of *weak* learner $\phi(\boldsymbol{f}_s, \Theta_s)$ by training each tree on a randomized subset s of the features $\boldsymbol{f}$. This leads to a diverse optimization of the individual model parameters Θ_s of each tree. The random forest is given by

$$\phi(\boldsymbol{f}, \Theta) = \{\phi_1(\boldsymbol{f}_{s_t}, \Theta_{s_t}), \ldots, \phi_t(\boldsymbol{f}_{s_t}, \Theta_{s_t})\}_{t=1}^{T}. \quad (1)$$

A pseudo-probability for a predicted class label ω can be estimated by considering the outputs of the T individual decision trees:

$$p(\omega \mid \phi(\boldsymbol{f}, \Theta)) = \frac{1}{T}\sum_{1}^{T} p_t(\omega \mid \phi_t(\boldsymbol{f}_{s_t}, \Theta_{s_t})). \quad (2)$$

These three steps lead to labeled images as shown in Fig. 2 and provides results that are sufficient for weed control given a solid amount of representative training data.

IV. Geometry-Based Classification

The goal of the geometric classifier is to assign a class label $\omega = \{c, w\}$ (c = crop, w = weed) to each detected object Ψ based *only* on spatial information by exploiting the relative arrangement of the plants in the field. The key idea is (i) to model the arrangement for the sugar beets as well as for weeds as two probability distributions of coordinate differences observed between plants and (ii) to employ a Bayesian approach to obtain a probabilistic output for the prediction. Specifically, we exploit the specific pattern of the sugar beet plants that is given by its row structure and a similar spacing between the sugar beets along the row. In contrast, weeds grow rather randomly on the field and can be assumed to follow a uniform spatial distribution.

A. Probabilistic Plant Arrangement Model

We define our relative arrangement model through conditional probability distributions

$$p(\boldsymbol{d} \mid \omega) \quad \text{with} \quad \omega = \{c, w\}, \quad (3)$$

of intra-class coordinate differences observed in a coordinate system, for which the x-axis is aligned to the actual crop-row direction, where

$$\boldsymbol{d} = \{\Delta\boldsymbol{x}_1^{row}, \ldots, \Delta\boldsymbol{x}_N^{row}\}_{n=1}^{N} \quad (4)$$

is a set of size N consisting of 2D coordinate differences. The intra-class coordinate differences are given by

$$\Delta\boldsymbol{x}^{row} = [|\Delta x^{row}|, |\Delta y^{row}|]^{\top} \quad (5)$$

and measured between the 2D positions of plants. Considering the crops, $|\Delta x^{row}|$ is the distance between sugar beet plants along the crop-row reflecting the similar spacing between them and $|\Delta y^{row}|$ is the distance between two sugar beets across the crop-row, which tends to take mainly small values around 0, see Fig. 3. For the computation of $\Delta\boldsymbol{x}^{row}$, we use the positions $\boldsymbol{x}_{\Psi}^{row}$ of the center of mass for each object Ψ as the reference point.

B. Learning the Crop Arrangement from Data

As new data arrives, we perform three steps to learn the crop arrangement model: (i) We represent the detected vegetation in a local map of a fixed size, (ii) estimate the actual crop-row considering the already classified crops in the local map, and (iii) compute $\boldsymbol{d}$ according to Eq. (4) and use it to update the plant arrangement model $p(\boldsymbol{d} \mid \omega)$.

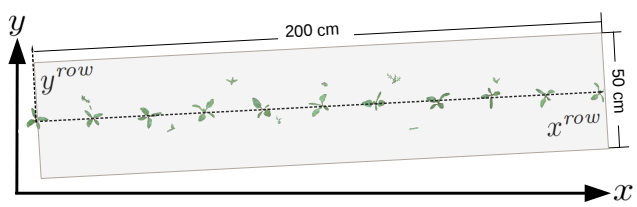

Fig. 3. Local map of the segmented objects Ψ. The dashed line depicts the estimated crop-row defining the coordinate system to compute $\boldsymbol{d}$.

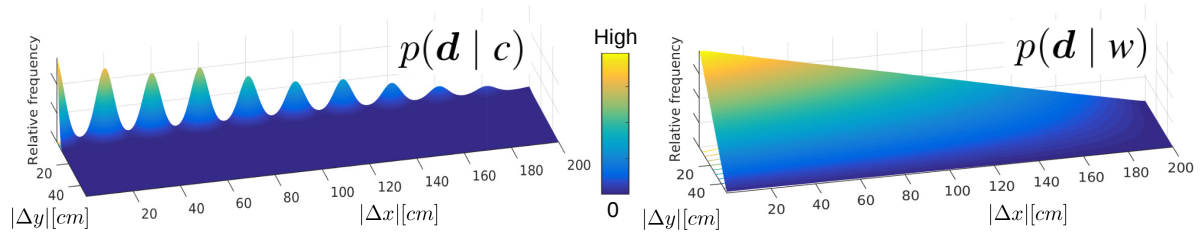

Fig. 4. Plant arrangement model according to Eq. (3) represented by the probability distributions of intra-class coordinates differences $\boldsymbol{d}$ measured between plants in a coordinate system, which is aligned to the actual crop-row. Left: Probability distribution $p(\boldsymbol{d} \mid c)$ for sugar beets learned from data (real distribution learned after approx. 20 m of traveling on a field in Bonn). Right: Probability distribution $p(\boldsymbol{d} \mid w)$ for weeds obtained under the assumption that weeds spatially follow a uniform distribution in object space. The shape of $p(\boldsymbol{d} \mid w)$ is caused by the computation of $\boldsymbol{d}$ in a finite space leading to smaller probabilities for the observation of large coordinate differences.

TABLE I
ACTIONS FOR SEMI-SUPERVISED APPROACH.

No.	Vision	Geometry	Prediction	Action
1	confident	confident	agree	nothing
2	uncertain	uncertain	(any)	nothing
3	uncertain	confident	(any)	add label to vision
4	confident	confident	contradict	check crop-row
5	confident	uncertain	(any)	check crop-row

1) Vegetation Mapping: We build a map of the segmented objects Ψ. We use the wheel odometry measurements to determine the motion of the camera and apply a pinhole camera model to project Ψ to the surface of the field, which we assume as a plane. For estimating $\boldsymbol{d}$, we consider only objects Ψ that lie within an area of 2 m along the crop-row and 0.25 m across the crop-row with respect to the current position of the camera. The reason for that is to minimize the effect of drift to the mapping induced by the integration of the wheel odometry measurements over time. Fig. 3 shows a sketch of a obtained map.

2) Crop-Row Detection: We perform the crop-row detection using a Hough transform, searching for the topmost voted line given the actual crops (sugar beets) in the local map. Finally, we optimize the result using a least-square estimator for line estimation by considering the supporters of the detected line by the Hough transform.

3) Update of the Plant Arrangement: We compute $\boldsymbol{d}$ between sugar beet objects Ψ_c and update our model $p(\boldsymbol{d} \mid c)$ for the crops by accumulating a 2-dimensional histogram of $\boldsymbol{d}$. This accumulation can be done as the used *coordinate differences* are not tight to an external coordinate system. As we have only a limited amount of data, we smooth $p(\boldsymbol{d} \mid c)$ using a Gaussian kernel. For the weed class, we obtain the distribution $p(\boldsymbol{d} \mid w)$ by *assuming* an uniform spatial distribution of weed objects within the local map. Fig. 4 depicts the learned arrangement model $p(\boldsymbol{d} \mid c)$ for crops as well as the assumed $p(\boldsymbol{d} \mid w)$ for weeds.

C. Prediction

In the prediction step, we compute the coordinate differences $\boldsymbol{d}$ from a new object Ψ to already classified crops in the local map and employ Bayes rule to obtain the probability

$$p(c \mid \boldsymbol{d}) = \frac{p(\boldsymbol{d} \mid c)\, p(c)}{\sum_{\omega} p(\boldsymbol{d} \mid \omega)\, p(\omega)} \tag{6}$$

for an object Ψ belonging to the crop class.

V. Combining Visual and Geometric Classifiers

This section describes how we combine the visual and the geometric classifier (i) to compute a joint classification of the crops and weeds and (ii) to achieve an online adaption of the visual classifier to match better with the actual distribution of the visual features.

A. Joint Classification

It is safe to assume that the features used by the visual and geometric classifier are independent of each other. As the training labels used by both classifiers are partially the same, the resulting classifiers may not necessarily independent. Nevertheless, we compute the class label ω^* for an object Ψ by maximizing the product of both distributions:

$$\omega^* = \underset{\omega}{\operatorname{argmax}}\, p(\omega \mid \phi(\boldsymbol{f}, \Theta))\, p(\omega \mid \boldsymbol{d}). \tag{7}$$

In case one of the classifiers has not been initialized yet, we consider a uniform distribution and thus ω^* turns into the response of the other classifier.

B. Semi-Supervised Learning by Crop Arrangement

The goal of the semi-supervised approach is to (i) exploit the predictions of the geometric classifier that are not affected by the potential change in visual appearance of the plantation in order to generate new training data for the visual classifier and to (ii) use the predictions of the visual classifier to identify errors in the crop-row estimation. Considering the predictions provided by both classifiers, we perform different actions as listed in Tab. I.

In cases *No. 1* and *2*, we perform no action. In case *No. 3*, we add the label provided by the geometric classifier to the visual training data to adapt the system to the field-specific visual feature distribution in the next retraining step. For the cases *No. 4* and *No. 5*, we see the following reasons to check the crop-row: (i) the appearance of the plants changed substantially so that the visual classifier fails while assuming to provide confident results. From our experience, this is the most common failure case. (ii) The crop-row detection is wrong. In practice, the crop-row detection can fail if crops are not present for a longer period, e.g. due to errors during sowing or because the robot moved outside the crop-row so that no plants are visible (in combination with a failure of the odometry system). To check which classifier to trust

in such situations, we estimate two new plant arrangement models, one based on the current visual classifier and one only from the geometric one. In both cases we use the plant information from the last 2 m of travel. This yields the models $p(\boldsymbol{d} \mid c_{vis})$ for the visual and $p(\boldsymbol{d} \mid c_{geo})$ for the geometric classifier as well as two independent estimates of the location of the crop-row. Then, we compare the distributions $p(\boldsymbol{d} \mid c_{vis})$ and $p(\boldsymbol{d} \mid c_{geo})$ to our currently used relative plant arrangement model $p(\boldsymbol{d} \mid c)$. The comparison is done using the Kullback-Leibler divergence, a general measure for the similarity of distributions. We trust the model which has the smaller distance under the Kullback-Leibler divergence. If the geometric classifier is assumed to be correct, we follow case *No. 3*. Otherwise, we use crop-row estimated by the visual classifier and proceed with the existing plant arrangement model.

C. Online Adaptation of the Visual Classifier

The main goal of the online adaption of the visual classifier is optimizing its model to obtain high quality classification outputs for the vegetation objects currently being observed on the field. The random forest framework offers two options: Either we retrain individual trees or we gradually replace individual trees in the random forest as new training data arrives.

Here, we explicitly utilize the newly gathered training data generated during the operation in the actual field environment to achieve the adaption of the visual classifier model to the current distribution of features, see Tab. I. We construct a new tree after having obtained a given amount of training samples generated by the geometric classifier and combine them with randomly chosen training samples obtained during the whole operation in the actual field. This leads to a field-specific adaptation of the random forest and thus generally leads to a better classification for this environment. In this way, for each field a new adaptation is possible, starting from an existing classifier (typically learned over multiple fields).

VI. Experimental Evaluation

The experiments are designed to show the capabilities of our method and to support the key three claims we made in the introduction section: Our approach is able to (i) identify sugar beets and weeds at an accuracy of over 95% requiring a small amount of training data, (ii) adapt a previously trained visual classifier to the current situation on the field, where the distribution of the visual features has notably changed, and (iii) is suitable for online operation at a rate of 5 Hz.

TABLE II
Dataset Information

	Bonn	**Zurich**	**Stuttgart**
# images	867	667	1175
# sugar beet objects Ψ_c	2393	1031	4568
approx. sugar beet size	2-9 cm^2	2-4 cm^2	4-12 cm^2
# weed objects Ψ_w	5647	3836	3195
approx. weed size	0.5-3 cm^2	0.5-5 cm^2	3-30 cm^2
intra-row spacing quality	high	low	high

Fig. 6. Labeling of the sugar beets with markers placed next to the plant.

We illustrate the performance of the classification results by ROC curves and Precision-Recall plots obtained by varying the threshold $t \in [0, 1]$ for the assignment of the class labels:

$$\omega = \begin{cases} c & \text{(sugar beet)}, \quad \text{if } p(\omega \mid \phi(\boldsymbol{f}, \Theta), \boldsymbol{d}) \geq t \\ w & \text{(weed)}, \quad \text{otherwise} \end{cases} \tag{8}$$

We evaluate performance on object-level, i.e. by comparing the predictions for each segmented object Ψ with the corresponding ground truth information. To obtain the ground truth, we manually labeled all images used for this evaluation. For the whole labeling process we spent more than one week of human effort.

All experiments are conducted on different sugar beet fields located near Bonn or Stuttgart in Germany as well as near Zurich in Switzerland, see Tab. II. The data has been recorded with different variants of the BOSCH DeepField Robotics BoniRob platform. All robots use the 4-channel RGB+NIR camera JAI AD-130 GE mounted in nadir view.

A. Classification with Minimum Labeling Effort

The first set of experiments is designed to show the performance of our approach and to support the claim that it is well-suited for identifying sugar beets and weeds by only requiring a small amount of training data. We target a labeling effort of approx. 1 min for a human and do not consider any pre-trained classifier. We achieve this 1 min labeling effort by placing printed markers next to a set of sugar beet plants in the beginning of the row. We can place around 10-15 markers within a minute, which corresponds to approx. 2-3 m of sugar beets along a row, see Fig. 6. The placement of the markers is the only labeling effort that we use. Based on this information, we can initialize the plant arrangement model and start training the visual classifier.

We conducted this experiment for the datasets recorded in Bonn and Zurich. The solid lines in Fig. 5 depict the achieved performance. For the Bonn dataset, we achieve an overall classification accuracy of around 97%. For sugar beets, we obtain a recall of 97% at a precision of 96% and for weeds a recall of 94% at a precision of 98% when assigning the class labels according to Eq. (7).

The most challenging dataset is the one in Zurich. Here, the size of the plants is around 0.5-5 cm^2 for weeds and 2-4 cm^2 for sugar beet. We achieve a recall of 90% for sugar beets and 97% for weeds at an overall accuracy of 95% according to Eq. (7). The primary objective in a robotic weed control scenario is to keep the number of false negatives small, i.e. sugar beet plants that are considered as weeds, to

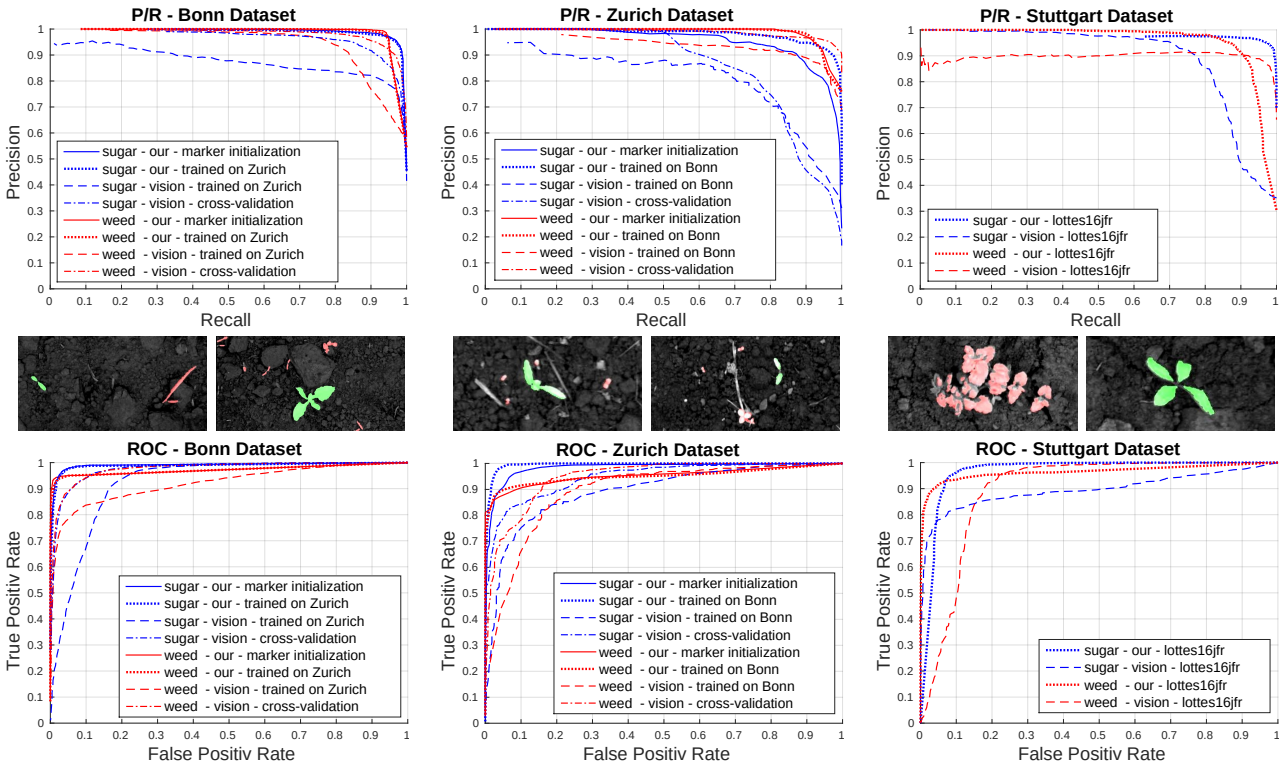

Fig. 5. Precision-Recall plots (top) and ROC curves (bottom) from or experimental evaluation. Left: classification performance on the Bonn dataset. Middle: classification performance on the Zurich dataset. Right: performance on the Stuttgart dataset. The term "our" refers our *proposed approach*, whereas "vision" refers to the results obtained by the "pure" *vision-based* classification.

Fig. 7. Plant arrangement model $p(\boldsymbol{d} \mid c)$ according to Eq. (3) learned only using ground truth data to analyze the quality of the intra-row spacing.

prevent the crops to be eliminated by the robot. Nevertheless, due to the comparably high precision for the weeds we can adjust the threshold for the class assignment according to Eq. (8) and detect around 85% of the weeds correctly at recall of 99.9% for sugar beets, which means that only 1 of 1000 crops is wrongly considered as weed by robot.

Compared to the Bonn dataset, the reason for the lower performance on the Zurich dataset is due to the higher variance for the spacing of the sugar beets along the crop-row. This can be seen in Fig. 7, which shows the arrangement models $p(\boldsymbol{d} \mid c)$ learned by using only labeled ground truth data. Here, the lower quality of the intra-row spacing in Zurich leads to a smaller support by the geometric classifier.

In addition to that, we compare our results with the performance of the visual classifier exploiting the full amount of training data. We apply a 10-fold cross validation on these datasets using the visual classifier. The results are depicted by the dotted-dashed lines and show that our proposed approach provides a better performance for both, the Bonn and Zurich dataset exploiting only a small amount of the training data for the initialization of the classification system.

B. Adaption of the Visual Classifier

The second set of experiments is designed to support the claim that our approach is suitable for adapting a visual classifier to the specific situation on the field. Here, we do not learn the visual classifier from scratch as in the first set of experiments, but adapt an already learned random forest as described in Sec. V. We conducted this experiment for the Bonn and Zurich dataset to show the performance and to compare the results with the first experiment. We do the same for the Stuttgart dataset and compare our proposed approach with the performance achieved in our previous work [7], which uses the same data.

We learn the visual classifier using the training data of the Zurich dataset to perform the classification on the Bonn dataset and also perform reverse experiment. The dotted lines in Fig. 5 depict the achieved performance for this experiment. For the Zurich dataset, we achieve a recall of 94% at a precision of 95% for the sugar beets, which represents a gain of 4% for both measures respectively compared to the first experiment. For weeds, the gain in performance is around 1% for both recall and precision. Thus, the overall accuracy increases by 2% to 97%. In the reverse experiment on the Bonn dataset, we do not obtain a considerable gain in performance. In terms of overall accuracy we keep 97% as obtained in the first experiment. This indicates that the pre-trained visual classifier becomes crucial as the quality of the intra-row spacing of the sugar beets deteriorates. See Fig. 7 for an illustration of intra-row spacing quality.

TABLE III

RUNTIME OF THE CLASSIFICATION PIPELINE IN MILLISECONDS.

Function	Mean [ms]	Std [ms]	Max [ms]
Preprocessing	21	1	22
Vegetation Detection	19	1	20
Feature Extraction	32	3	46
Vis. Classification	4	2	21
Vis. Classifier Update	7	4	12
Geom. Classification	9	1	16
Geom. Classifier Update	3	2	5
Whole pipeline (incl. rest)	**124**	**14**	**191**

In sum, these results suggests that a pre-trained model can support the performance, especially for the sugar beets. The smaller support for the weeds is probably due to the different weed types growing in Bonn compared the ones occurring in Zurich. This causes larger variations in their visual feature distributions compared to the features of the sugar beets.

The dashed lines in Fig. 5 illustrate the performance achieved by the *transfered* visual classifiers without the support of the geometric classifier. Either way, the results demonstrate that our proposed approach, using a pre-trained base classifier or not, outperforms the "pure" vision-based classification. In sum, we gain 14% for the Bonn and 20% for the Zurich dataset in terms of overall accuracy. Compared to our recently published results in [7] for the Stuttgart dataset, we obtain an improvement of round 10%. Especially, the comparably lower performance for sugar beets in [7] is dramatically improved by our new approach.

Thus, our proposed semi-supervised approach makes a big step towards bridging the performance gap in visual crop and weed detection if the distribution of the features at training differs from the one observed during operation and has the ability to adapt a pre-trained classification model to the current feature distribution on the field.

C. Runtime Analysis

The last experiment in designed to support the claim that our approach runs fast enough to provide suitable results for online operation of a robotic weed control system. We evaluated the runtime on the computer that is installed on the field robot (Intel i7 CPU, GeForce GTX-1080 GPU). Tab. III lists the execution time of the whole pipeline and for specific processing steps per image. The evaluation is performed using all datasets used in this paper. On average, we can provide high quality classification results at 8 Hz. In the worst case, we can operate at 5 Hz.

VII. CONCLUSION

In this paper, we presented a practical approach for obtaining high performance crop and weed classifiers for agriculture robotics. Our approach exploits an image-based as well as a geometric classifier and combines both in a probabilistic way that only minimal labeling efforts are needed. It exploits the fact that several crops such as sugar beets are planted in rows with a similar intra-row spacing. This fact can be used to update the vision-based system to better handle the variations in the visual appearance of the plants. We implemented and evaluated our approach on different field robots and on different fields in different cities and provided a detailed experimental analysis that supports all claims made in this paper. The experiments suggest that we achieve a high classification performance requiring only 1 min of labeling effort, we can easily adapt the classifiers, and we can execute the system online on the field robot.

ACKNOWLEDGMENTS

We thank R. Pude and his team from the Campus Klein Altendorf for their great support as well as the team from ETH Zurich around A. Walter, F. Liebisch and J. Pfeifer. We furthermore thank the people from BOSCH Corporate Research and DeepField Robotics for their support during joint experiments. Further thanks to N. Chebrolu, J. Behley and S. Wenzel for fruitful discussions.

REFERENCES

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

[2] M. Di Cicco, C. Potena, G. Grisetti, and A. Pretto. Automatic model based dataset generation for fast and accurate crop and weeds detection. *arXiv preprint arXiv:1612.03019v1*, 2016.

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

[4] S. Haug, A. Michaels, P. Biber, and J. Ostermann. 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, 2014.

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

[6] 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.

[7] 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.

[8] P. Lottes, R. Khanna, J. Pfeifer, R. Siegwart, and C. Stachniss. Uav-based crop and weed classification for smart farming. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2017.

[9] 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.

[10] A.T. Nieuwenhuizen. *Automated detection and control of volunteer potato plants*. PhD thesis, Wageningen University, 2009.

[11] C. Potena, D. Nardi, and A. Pretto. Fast and accurate crop and weed identification with summarized train sets for precision agriculture. In *Proc. of Int. Conf. on Intelligent Autonomous Systems (IAS)*, 2016.

[12] F. M. De Rainville, A. Durand, F. A. Fortin, K. Tanguy, X. Maldague, B. Panneton, and M. J. Simard. Bayesian classification and unsupervised learning for isolating weeds in row crops. *Pattern Analysis and Applications*, 17(2):401–414, 2014.

[13] D.C. Slaughter, D.K. Giles, and D. Downey. Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture*, 61(1):63 – 78, 2008.

[14] W. Strothmann, A. Ruckelshausen, J. Hertzberg, C. Scholz, and F. Langsenkamp. Plant classification with in-field-labeling for crop/weed discrimination using spectral features and 3d surface features from a multi-wavelength laser line profile system. *Computers and Electronics in Agriculture*, 134:79 – 93, 2017.

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

[16] A. Wendel and J. Underwood. Self-supervised weed detection in vegetable crops using ground based hyperspectral imaging. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 5128–5135, 2016.