Evaluation of intra variability between annotators of weed species in color images

Mads Dyrmann [a,*], Henrik Skov Midtiby[a], Rasmus Nyholm Jørgensen [b]

[a] The Maersk McKinney Moller Institute, University of Southern Denmark, Odense, Denmark

[b] Department of Engineering, Aarhus University, Aarhus, Denmark

* Corresponding author. Email: mady@mmmi.sdu.dk

Abstract

In order to develop automated systems for plant recognition in images, it is necessary to possess large amounts of training data. In the project, RoboWeedSupport, plant production consultants annotate images, distinguishing between different weeds, and thereby producing the training data for our automated system. Usually we assume that the annotations made by the consultants are correct, but with this study, we will clarify whether this assumption is correct: Some species may be difficult to recognize, which may cause errors in the training data. Consequently, the automated weed discriminating model will be less precise.

In this study, we will compare how 8 consultants recognize weeds in images from a Danish maize field. These weeds are at a growth stage where some of the plants have developed the first true leaves, but the majority are at the dicotyledonous stage. The results show that 29 different weed species has been registered in the images and that the consultants in average agree on 88% of the weeds.

Keywords: Annotations, plant species recognition, precision agriculture, herbicide optimization

1. Introduction

In recent times, there has been an increased focus on reducing the use of pesticides in agriculture. This is by the increased number of rules and documentation requirements for the farmer. Among these rules is *EU Directive 2009/128/EC*, which indicates that the farmer must follow some general requirements about the use of Integrated Pest Management.

These requirements include that the farmer must first spray his after determine the needs of his field and then target the weed control for the specific needs. Inspecting fields and recognizing weeds can, however, be close to impossible for the individual farmer as he might neither have the knowledge about the individual weed species nor have the time to do the field inspection. Determine the weeds in the field automatically from images, like in the RoboWeedSupport project, would be a way to solve this task.

When you train a classifier to recognize images, such as a neural network or support vector machine, you need images of known weeds. These images can be from weeds that have been sown, whereby the species of the weeds are known. Alternatively images from the field can been collected and experts can annotate the plants in these images. In case of the latter, there will be uncertainties associated with the annotations, due to the limited information available to the annotators (e.g. they are not able to look at the plant from multiple viewpoints and thus inspect certain characteristics). Still, annotations from such images are often used as ground truth for training.

In this study, eight consultants have annotated images of weeds from a Danish maize field. The ground truth for these images is not known, but each image has been annotated by two or three consultants.

The results of this study will provide training data for an automated weed distinguishing algorithm, potentially leading to recognition abilities beyond those of the average plant production consultant.

2. Data Materials

The data material that is used in this study consists of 47 images of weeds and crops collected on a single day in June from a Danish maize field. All weeds are at an early growth stage and cover primarily the growth stage from BBCH 12 to 14 (Meier, 2001).

The images have been photographed vertically towards the ground by a Samsung NX1000. The ground resolution is about 9pixels/mm, which provides details of even small plants. An image sample is shown in Figure 1.

Figure 1. Sample of one of the 47 images from which the weeds has been cropped out

Figure 2. Plant samples

The 47 pictures are divided among the consultants so that the plants from each image is annotated by a minimum of three consultants and distributed so that all consultants annotate against each of the other consultants at least 6 times.

For each of the 47 images, plants are found by examining connected components in the green chromaticity. Afterwards, plants objects are manually marked and the boundaries between plants are marked, so that each plant can be cut out. Lastly the individual plants can be analysed by the consultants.

From these 47 images about 3350 weed objects have been segmented.

Samples from the images are shown in Figure 2.

3. Methods and Results

This study will show how much uncertainty there is related to these manual annotations of plant images. Since we do not have any ground truth annotations, we can only analyse how often a the consultants disagree on the right species of the plants, but not whether one, two or all three consultants are correct or not

3.1. Number of annotations

A total of 4336 annotations have been made. Of these annotations 2929 annotations are from unique plants. The consultants could choose the species from a list of 114 Danish weed species. Furthermore, if the consultants were unable to tell the exact species of a plant, but only the family, they could choose to only annotate a plant as belonging to a specific family. For instance by annotation a plant as cereal rather than barley. Likewise the consultants could annotate grass as either broad-leaved or narrow-leaved grasses rather than the specific species, which can be hard to discriminate at early growth stages. In order to be able to compare as many annotations as possible, the comparison is only made at the "family level" for the species for which at least one sample has been annotated only as a specific family.

The 4,336 annotated plants cover 29 different plant species, which are distributed as shown in Figure 3. As can be seen particularly Field Pansy and Knotweed account for most of the annotated plants in the images, while 14 species are detected less than 10 times.

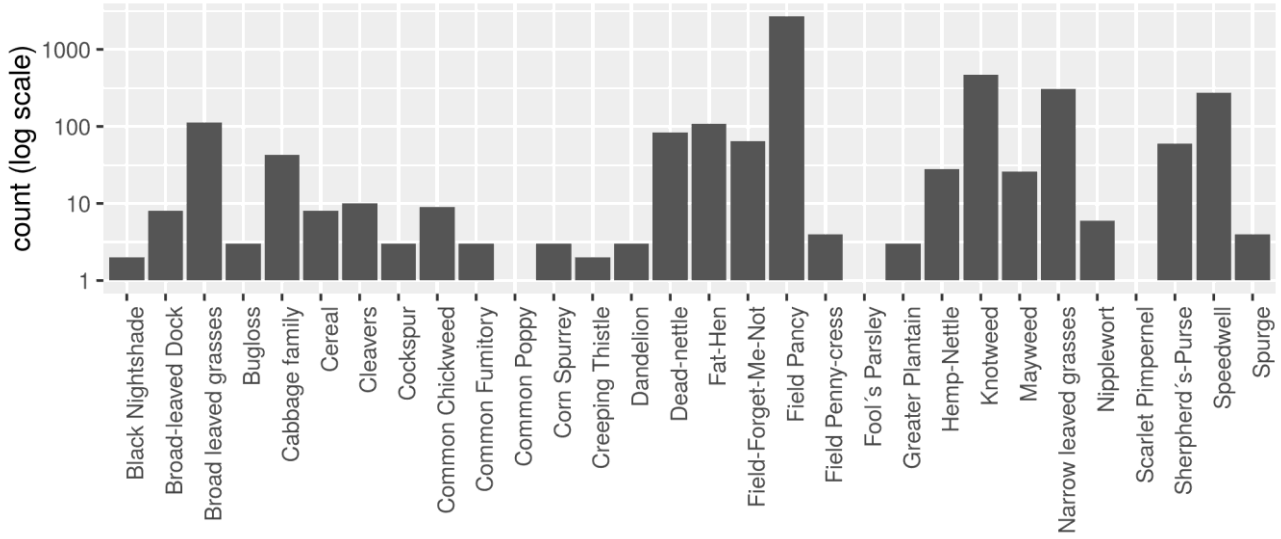

Figure 3. Distribution of annotation labels (log scale)

Table 1. Number of annotated plants for each consultant

Consultant	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
# of annotations	396	278	828	1457	372	310	306	389

These 4,336 annotations are distributed between the consultants, as shown in Table 1, which ranges from 278 to 1,457 plants per consultant.

Table 1 shows that some consultants have annotated more plants than other, which means that there is not three annotations for all plants. Of the 2929 unique plants, 2617 plants have been annotated at least two times, while 311 have been annotated three times. In the following study, we will only consider the plants with at least three annotations.

3.2. Congruence between three annotations

Some plants might be easier to recognize than others. Therefore, we will now look at how the congruence and incongruence is between the annotations for each of the detected plant species. Of these 311 plants, all of which have been annotated three times, there are 276 plants, where all three consultants agree and 35 plants, for which at least one consultant disagree.

Figure 4a shows the number of plants where all consultants agree, while Figure 4b shows the number of plants for each species, where at least one of the consultants disagree. An interesting figure here is speedwell, which primarily is represented in the group of plants for which the consultants disagree on the species.

However, as there is different numbers of annotations for each species, this must be taken into account, by normalizing the number of occurrences, where the consultants disagree by the total number of occasions for a given annotation. Figure 5 shows the number of times a plant of a given species is involved in an annotation for which all three consultants agree compared to the number of times a species is involved in an annotation, for which at least one of the consultants disagree. Here we see, that the consultants agree in less than 10% the annotations of what at least one consultant says is speedwell. The overall fraction of plants, where three consultants agree is 88.7%.

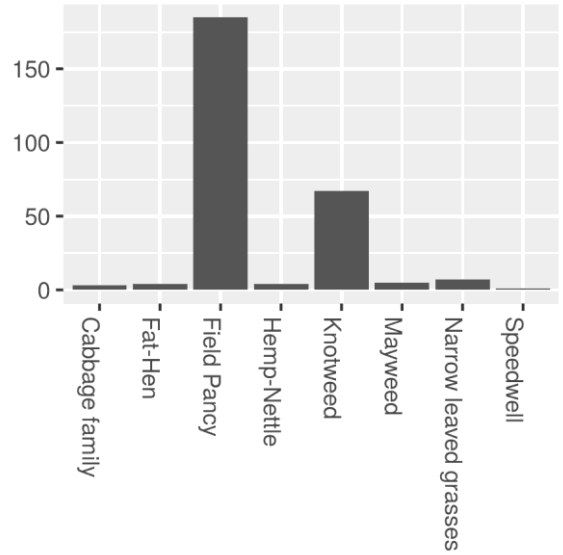

Figure 4a. Number of samples for the plants for which all consultants agree

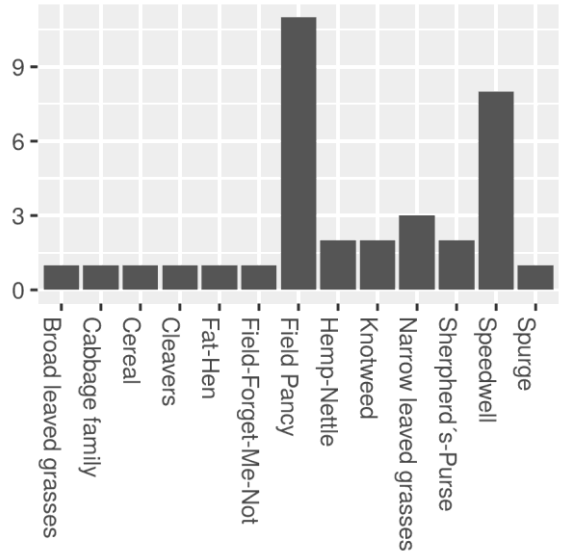

Figure 4b. Number of samples for the plants for which at least one consultants disagree

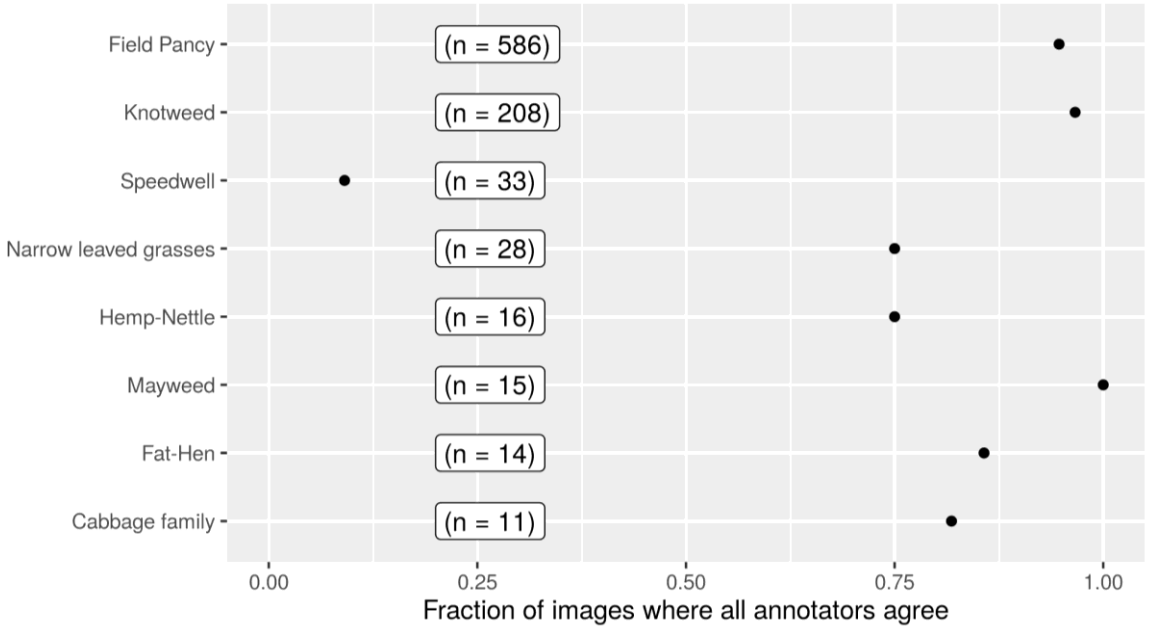

Figure 5. The fractions of which all three consultants agree on the given species. The bubbles on the axis shows the number of samples, n, that has been annotated as that specific species

3.3. Congruence for all pairs of annotations

We will now consider all plants with more than one annotation and look at how the confusions between the annotations of the plants are distributed. Because we have no ground truth annotation, all annotations for each plant takes turn in being used as reference for the given plant and then compared to the remaining annotations for that plant.

A way to see this is as a game between two consultants, where one consultant says which species he believes a given plant is, which is the 'reference species'. Another consultant then tells which species he believes it is, which then is the 'annotated species'. The result is plotted in the confusion matrix in Figure 6.

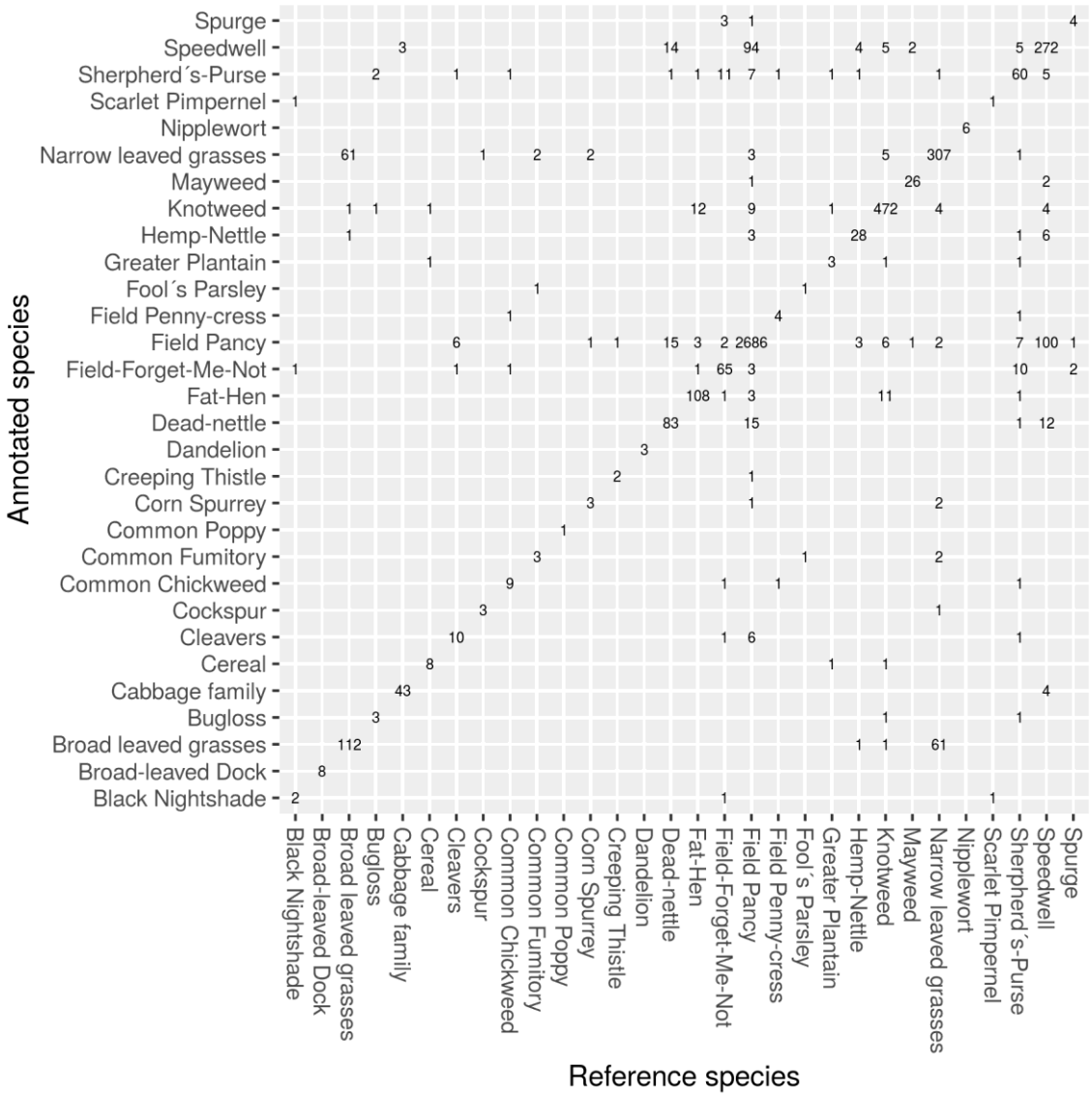

Figure 6 Distribution of pairwise annotations of plants. The annotations take turn at being reference and the remaining annotations for that plant are marked as 'annotated species'.

From this confusion matrix we can see that 88% of the annotation pairs are at the diagonal, which means that when one consultant puts a given label one a plant, another consultant will agree in 88% of the cases. However, when one consultant says that a plant is speedwell the remaining consultants only agree in 67% of the cases.

Likewise, there are large incongruence between the annotations of plants labelled as narrow-leaved grasses and broad leaved grasses.

3.4. Distribution of inconsistent annotations for the consultants

We will now see how often each of the eight consultants is involved in annotations with inconsistence. Table 2 shows each annotator and how often the given consultant is involved in an annotation where at least one consultant disagree. Here it is clear that consultant number 6 has an inconsistence rate, which is much higher than the remaining seven consultants. Field Pansy is the plant for which consultant number 6 most often disagree with the other consultants. Here number 6 has made 133 out the total number of 280 annotations.

The low inconsistence rate for the remaining consultants might also be because they more often choose not to annotate plants, for which they are in doubt of which species it is.

Table 2. Number of times a consultant is involved in annotations for which at least one of the other consultants disagree on the annotation

Annotator	# of Inconsistent images	Total # of images	Inconsistence rate
1	57	396	0.16
2	31	278	0.11
3	131	828	0.16
4	106	1457	0.07
5	27	372	0.07
6	135	310	0.44
7	41	306	0.13
8	59	389	0.15

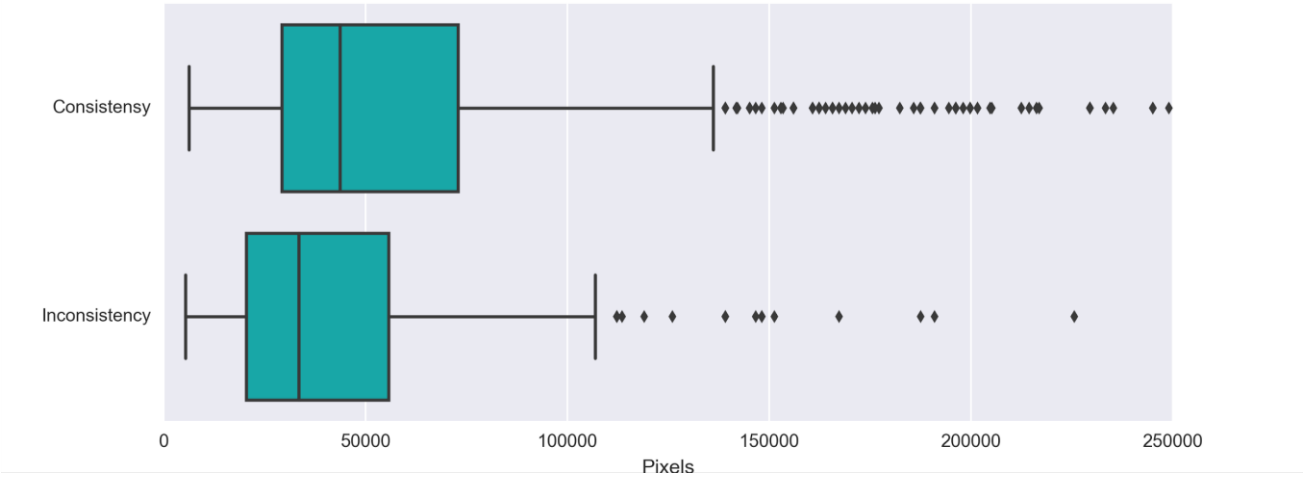

Figure 7. Distribution of sizes for plants for which the consultants agree and disagree on the species

We will now see if there is a relationship between the size of plants and the number of times, they are confused. Figure 7 shows the sizes of bounding boxes in pixels for all the plants where there are consistence and inconsistence between the annotations. From this figure, it can be seen, that the sizes of the plants, for which the consultants do not agree, tend to be smaller, than for the plants, for which they agree. As the images are sharp and have a high resolution, these numbers indicate that plants at earlier growth stages are the hardest ones to recognize.

4. Conclusions

In the present study, we have shown how annotations are distributed when eight plant consultants are set to annotate weeds in images. The study has shown that for the majority of the plants, the consultants agree on plant species. However, for about 12% of the plants, the consultants disagree on the species.

The study has also given knowledge about which species the consultants have difficulties in recognizing. When an automatic system is to be trained to recognize these difficult species, it will be an advantage to have more consultants to annotate the same image and then use majority voting to determine the species. Similarly, it will be sufficient to have a single consultant to annotated species for which there rarely is doubt about the species.

Acknowledgements

The work was founded by Green Development and Demonstration Programme (GUDP) under the Danish Ministry for Food, Agriculture and Fisheries. The agro consultants are from SEGES, Gefion, Centrovice, LMO, BLF, LandboNord, AgriNord, and IPM Consult.

References

Meier, U. (2001). Growth stages of mono-and dicotyledonous plants - BBCH monograph. Federal Biological Research Centre for Agriculture and Forestry. Retrieved from http://www.bba.de/veroeff/bbch/bbcheng.pdf