Master Thesis Proposal
A Deep Learning Approach for Stem-Base Detection of Row Crop Plants

Author:
Stefan Schweng, BSc

Supervisors:
Ao. Univ. Prof. Dipl.-Ing. Dr.techn. Markus Vincze
Dipl.-Ing. Dr.techn. Peter Riegler-Nurscher

October 2022

1 Motivation & Problem Statement

In order to increase agricultural profit, farmers aim for low weed density on their fields. A cheap and convenient way of achieving this is chemical weed control, i.e. the use of herbicides. However, several studies in the past two decades revealed a significant negative impact of herbicides and other types of pesticides on climate resilience, biodiversity [14] and especially the abundance of pollinators [13] in ecosystems around the world. Further, the use of herbicides in high quantities in agriculture for food production has a negative impact on human health [3].

In 2020 the European Commission approved a set of policy initiatives, referred to as European Green Deal [16], with the aim to make Europe a net-zero emitter of greenhouse gases by 2050. One of these initiatives is the reduction of use and risk of chemical pesticides by 50% by 2030.

Due to the reasons mentioned above and the growing herbicide-resistance of weed [5], the demand for organic food production is increasing, which makes it necessary to find and use alternatives to chemical weed control. One alternative approach for row crop is a process called hoeing, where metal tools are pulled through farmland in order to remove weed mechanically. One of the big challenges for applications like these is to make sure that weed gets removed efficiently while the crop stays unharmed.

Figure 1 shows a hoeing machine with camera control on a field. The metal tools, which get pulled through the field, are mounted on a frame with a fixed spacing in between them. The camera, which is mounted on top of the device, is part of a computer vision system which processes farmland images in real-time in order to gather information on the crop row alignment. Further, this information can be used to automatically adjust the tool-frame’s linear offset normal to the driving direction of the tractor.

Up to a certain level of weed coverage and variance in crop height, state-of-the-art systems, as shown in Figure 1, achieve significantly higher efficiency while enabling higher tractor speeds compared to hoeing approaches with manual tool-frame offset adjustment. Stem-base detection enables not only precise row guidance but also in-row weeding. With in-row weeding, tools are mechanically moved between individual plants, allowing for even less weed coverage.
However, under sub-optimal conditions there is still room for improvement for systems like these when it comes to detecting crop rows.

The goal of this master thesis is to establish a deep learning approach for stem-base detection of row crop plants as well as a simulation environment for pre-training neural networks for applications as described above.

2 Aim of the Work

As mentioned above, state-of-the-art computer vision systems have problems detecting crop rows correctly under sub-optimal conditions. Examples for sub-optimal conditions in this case are the following:

a) high weed-pressure,

b) high variance in crop height,

c) similar height of crop and weed,

d) varying light conditions,

e) curved crop rows or

f) bad image quality (e.g. dusty conditions during image acquisition).

The aim of this thesis is to investigate deep learning methods to detect stem-bases of row crop plants with the goal to enable higher accuracy in crop row detection and to tackle the limitations mentioned above. In the course of these investigations three sub-goals shall be defined as follows:
1. Establish a deep learning approach for stem-base detection and a crop row detection algorithm, which surpasses state-of-the-art systems in at least one of the sub-optimal conditions mentioned above.

2. Create a simulation environment, which can be used to pre-train neural networks for row crop stem-base detection.

3. Validate the established stem-base and crop row detection approach in a real world scenario and in real-time on farmland.

3 Methods

The following subsections describe the methodological approach which shall be followed in the course of this thesis.

3.1 Literature Review

There are several different approaches to the problem of crop row detection, which include the use of different sensing technologies. Due to the high availability and the possibility for cheap acquisition of 2D-RGB image data, the focus for this thesis shall be on this type of sensing. Considering this, the literature review for this thesis will be focused on stem-base and crop row detection approaches, based on 2D-RGB data. The outcomes of this thesis shall be compared to the state-of-the-art approaches found in the literature review.

3.2 Evaluation of different Deep Learning Methods

The main focus of the practical thesis part shall be on evaluating different deep learning methods for stem-base detection. For purposes like these, the two most common deep learning strategies for applications like these are called Semantic Segmentation and Object Detection. At least one implementation of each type shall be evaluated with adequate measures. PytorchAutoDrive [15], a framework for self-driving perception, will be considered as a starting point for the implementations.

The algorithm for crop row detection shall build on top of the stem-base detection approaches. There shall be less focus on this practical part. However, the outcomes of this part shall be used to compare the overall detection results to the other crop row detection methods found in the literature review.

3.3 Simulation Environment for Neural Network Pre-training

Another goal of this thesis is the implementation of a simulation environment for pre-training neural networks for stem-base detection or similar tasks. This environment shall simulate the perspective of a camera as it would be mounted on a device as shown in Figure 1. There shall be the possibility to specify camera parameters like the 3D-position, angles of view and similar variables.

The visualized content of the simulation shall look as similar to real farmland as possible within a feasible amount of work time. There shall be one type of row crop and weed plant randomized in a way so that it is possible to generate adequate data for pre-training of neural networks. For the implementation, the suitability of environments like Gazebo [9] or the computer game Farming Simulator [17] will be evaluated. There is also a simulator called AgROS [12], which is based on Gazebo and meant for agricultural robotics. AgROS could be considered as a starting point for the this simulation environment as well.
3.4 Real-time Experiments on Farmland

In order to compare the performance of the most promising implementations of this work to approaches described in other scientific articles, there shall be experiments in real-time on real farmland. The implementations for stem-base and crop row detection shall be evaluated by acquiring farmland images from a moving vehicle on a field and processing them in real-time.

4 State-of-the-art

This section serves as an overview for relevant literature on crop row detection and stem-base or plant position estimation respectively. In the initial literature review a significantly higher number of scientific articles on crop row detection than on stem-base detection was found. This is the reason why there shall be an implementation for crop row detection in the course of this thesis, since it enables better comparison to existing approaches. Further, the lack of material regarding stem-base detection additionally yields the need for research in this area.

Zhang et al. [11] introduce and apply a new vegetation index to farmland RGB images, followed by a two-staged thresholding process in order to segment maize, weed and background. From this segmentation a binary image is obtained with all the pixels classified as maize. After applying a dilation process to fill gaps between maize pixels in the binary image, feature points (i.e. maize plant positions) are extracted and the number of crop rows is determined via two separate processes, each based on the vertical projection method [4]. The last step of the proposed method, the detection of crop rows, aims to find feature points belonging to the same crop row by clustering them based on distance and angle constraints between the feature points. In order to refine the found point sets the authors use the Floyd algorithm [1] to find the shortest path from the bottom to the top of the image for each of the sets. Finally, the least squares method is applied on the final point sets to end up with straight-line equations representing the detected crop rows. The proposed method shows outstanding performance under high weed-pressure conditions and superior performance in direct comparison with the Hough Transform method [18]. The angle error between detected and ground truth crop rows was less than 0.5°.

In [10] Bah et al. use farmland images made from an unmanned aerial vehicle (UAV) to detect crop rows in order to detect inter-row weeds. The images of detected weeds are used to train convolutional neural networks (CNNs), which can then be used to detect weeds on other fields. Like in [11] the authors of this method apply a vegetation index on the RGB data as a pre-processing step in order to segment plants and background (i.e. soil, shadows and stones). The pixels classified as plants are transformed to binary images. The binary images are then skeletonized before applying the Hough Transform method in order to detect crop rows.

Haug et al. [8] apply sliding window feature extraction followed by non-maximum suppression on multi-spectral image data in order to obtain a probability map for plant stems in the processed image. Their implementation aims to detect both, use plant and weed stems. The described sliding window approach calculates feature vectors, consisting of statistical and geometrical features, from image patches which show biomass in their center. The extracted feature vectors are passed to a Random Forest classifier to gather stem probabilities for each patch. Their implementation was tested and evaluated with images from a carrot farm, which resulted in an stem detection rate of 80.4% with a mean position error of 1.88mm.

In [6] Kiani et al. implement a crop detection and positioning method based on an artificial neural network (ANN) for a mechanical weeding application. The authors apply a vegetation index on digital camera images to segment plants and background, like in [11] and [10]. The dataset for training consists of 180 images from which seven different shape features of plants
have been extracted. With the help of discriminant analysis the number of features needed for crop and weed classification is reduced from seven to four. For the classification itself an ANN is trained to distinguish between crop and weed. The classified crop pixels are then used to calculate corresponding centroids, which are finally used as an estimation of the crop’s stem position. The system’s classification accuracy reached a detection rate of 98.9%. The proposed method achieved crop position errors of less than 1.5 cm.

Fu et al. [7] introduce a method to detect stems of potted tomato plants based on data acquired with the stereo vision system Kinect. The depth information is used to segment the potted plant and its background. The background pixels are removed from the corresponding 2D-RGB image and the remaining pixels are skeletonized after transforming the RGB to a binary image. Finally, the Progressive Probabilistic Hough Transform is applied to detect plant stems in the image. The outcomes of this algorithm are used to extract textures of the plant’s stems from the original 2D-RGB images. These textures are then used to generate virtual 3D models of tomato plants using OpenGL. The plant models are generated based on the L-system [2] algorithm.

5 Relevance to the Curriculum of Computer Engineering

The topic of this master thesis overlaps with several courses in the curriculum of Computer Engineering. The probably most relevant course is 376.054 Machine Vision and Cognitive Robotics since it covers various state-of-the-art methods used in computer vision. Also the use and development of simulation environments is taught in several courses. The following list is a set of courses I attended, which are relevant to the topic of this thesis.

- 182.763 Stochastic Foundations of Cyber-Physical Systems
- 183.660 Mobile Robotics
- 191.119 Autonomous Racing Cars
- 376.054 Machine Vision and Cognitive Robotics

5
References