



DISSERTATION

Mobile vision system for estimation of soil and plant properties

conducted in partial fulfillment of the requirements for the degree of a
Doktor der technischen Wissenschaften (Dr. techn.)

supervised by

Ao.Univ.-Prof. Dr. techn. Markus Vincze
E376 Automation and Control Institute

submitted at the

TU Wien

Faculty of Electrical Engineering and Information Technology

by

Dipl.-Ing. Peter Riegler-Nurscher, BSc

DOB 21.11.1988

Matr. Nr.: 00930832

Vienna, April 2021

Peter Riegler-Nurscher

Abstract

Computer vision serves as the foundation for the automation of many processes and tasks on agricultural fields. Camera-based applications can be used as assistance systems and for the more efficient use of agricultural machines. From an environmental perspective, erosion and nitrate pollution are major challenges in agriculture. Vision systems on agricultural implements can guide tillage and fertilization in a more sustainable way and are the foundation for full automation. This thesis presents soil cover and soil roughness estimation methods from short distances by cameras, as well as, methods for grassland yield and composition estimation. Soil cover decreases soil erosion by reducing the kinetic energy of rain drops impacting the soil surface whereas a certain level of soil roughness reduces erosion as soil wash-off is reduced. Both parameters must be measurable in order to automatically manage them in agronomic processes like tillage. On the other hand, yield estimation in grassland forms the basis for site-specific fertilization, which in turn reduces nitrate pollution in the groundwater.

The proposed methods for soil cover estimation are based on so-called entangled random forests as well as on a convolutional neural network (CNN). Both methods are able to distinguish between soil, living organic matter, dead organic matter and stones by semantic segmentation of RGB images. A stereo camera setup enables soil roughness estimation in real time on a power harrow. The methods for soil cover estimation are integrated for handling residues during the roughness estimation. The final RGB stereo camera setup, equipped with AI inference hardware, is also able to estimate grassland yields by plant height measurement and plant composition estimation using CNNs on a mower. The camera mounted on a rake detects and measures grassland swaths. The system opens various applications in grassland harvesting from machine steering and yield estimation to real-time machine control and logistics optimization. The methods for soil cover estimation exceed state of the art methods and are the basis for reliable soil roughness measurement. The method for soil roughness measurement was compared to the sieve analysis and was judged to be sufficiently accurate for the task of seeding with a power harrow as secondary tillage. In this thesis, the first method is presented to control a tillage machine in real time using soil roughness measurements. It has been shown that the method for controlling a power harrow improves homogeneity of soil roughness on the field during seeding. In the grassland, new methods are proposed for measuring yield and composition directly on agricultural implements like the mower and the rake. Additional sensors (radar, laser and ultrasound) were used during mowing to investigate their suitability for yield estimation. The methods utilized in the grassland are compared to weighed dry matter yield samples over three seasons. All methods were developed in the context of extensive field trials to explore environmental influences. The field trials were conducted across fields and across seasons to investigate and reduce seasonal and regional biases.

Kurzzusammenfassung

Computer Vision ist die Basis für Automatisierung vieler Prozesse und Aufgaben in der Landwirtschaft. Auf dem Weg zur Vollautomatisierung können Bildverarbeitungsanwendungen in einem ersten Schritt als Assistenzsysteme und zur effizienteren Nutzung von Landmaschinen eingesetzt werden. Aus Umweltgesichtspunkten sind Erosion und Nitratverschmutzung große Herausforderungen in der Landwirtschaft. Computer Vision Systeme auf landwirtschaftlichen Maschinen können helfen, Bodenbearbeitung und Düngung auf nachhaltigere Weise durchzuführen und den Weg zur Autonomisierung öffnen. Ziel dieser Arbeit ist es, an Methoden zu forschen, mit denen Boden- und Pflanzenparameter mit Kamerasystemen auf landwirtschaftlichen Maschinen erfasst werden können. In dieser Arbeit werden Methoden zur Schätzung der Bodenbedeckung und der Bodenrauheit aus kurzen Entfernungen mithilfe von Kameras, sowie Methoden zur Schätzung des Ertrags und der Zusammensetzung von Grünland vorgestellt. Bodenbedeckung verringert die Bodenerosion, indem sie die kinetische Energie der auf der Bodenoberfläche auftreffenden Regentropfen reduziert, während eine gewisse Bodenrauheit die Erosion durch reduzierte Bodenabschwemmung verringert. Beide Parameter müssen messbar sein, um sie im Zuge agronomischer Prozesse, wie der Bodenbearbeitung, zu optimieren. Andererseits bildet die Ertragsabschätzung im Grünland die Grundlage für eine teilflächenspezifische Düngung, welche wiederum das Potential zur Reduktion der Nitratbelastung im Grundwassers hat.

Die beiden vorgeschlagenen Methoden zur Schätzung der Bodenbedeckung basieren auf einem sogenannten Entangled Random Forest sowie auf einem Convolutional Neural Network (CNN). Mithilfe von semantischer Segmentierung von RGB-Bildern ist es möglich, zwischen Boden, lebendem organischem Material, totem organischem Material und Steinen zu unterscheiden. Ein Stereokamera-Setup ermöglicht die Schätzung der Bodenrauheit in Echtzeit auf einer Kreiselegge. Die Methode zur Schätzung der Bodenbedeckung ermöglicht es Reststoffe wie Stroh während der Rauheitsschätzung zu erkennen und zu filtern. Das RGB-Stereokamera-Setup, das mit AI-Inference-Hardware ausgestattet ist, ist in der Lage, Grünlanderträge durch Messung der Pflanzenhöhe und Schätzung der Pflanzenzusammensetzung durch CNNs auf einem Mähwerk zu schätzen. Wird die Kamera auf einem Schwader montiert, werden Grünlandschwaden erkannt und vermessen. Das System eröffnet dadurch verschiedene neue Anwendungen in der Grünlandernte, von der Maschinensteuerung über die Ertragsschätzung bis hin zur Echtzeit-Maschinensteuerung und Logistik-Optimierung.

Die Methoden zur Schätzung der Bodenbedeckung übertreffen den Stand der Technik und sind die Grundlage für eine zuverlässige Bodenrauheitsmessung. Die Methode zur Bodenrauheitsmessung wurde mit der Siebanalyse verglichen und für die Anwendung bei der Aussat als ausreichend genau beurteilt. Darauf aufbauend wird die erste Methode zur Steuerung einer Bodenbearbeitungsmaschine in Echtzeit durch Bodenrauheitsmessung

vorgelegt. Es wurde gezeigt, dass die Methode zur Steuerung einer Kreiselegge die Homogenität der Bodenrauheit innerhalb eines Feldes verbessert. Im Grünland werden neue Methoden zur Messung von Ertrag und Zusammensetzung direkt auf Grünlandgeräten wie Mäher und Schwader vorgeschlagen. Zusätzliche Sensoren (Radar, Laser und Ultraschall) wurden während des Mähens eingesetzt, um ihre Eignung für die Ertragsabschätzung zu untersuchen. Die Methoden im Grünland wurden mit Trockenmasse-Ertragsproben über drei Saisonen validiert.

Alle vorgestellten Methoden wurden im Rahmen von umfangreichen Feldversuchen entwickelt und ermöglichen es dadurch, Umwelteinflüsse zu erforschen. Die Feldversuche wurden feld- und saisonübergreifend durchgeführt, um saisonale und regionale Verzerrungen zu untersuchen und zu reduzieren.

Acknowledgements

I could have never completed my PhD alone and many people have supported me along the way who I would like to thank.

First, my advisor Prof. Markus Vincze gave me the opportunity to do a PhD as an external student at his institute. I always appreciated Johann Prankl's support, experience, and cooperation. He has passed on so much technical and scientific knowledge to me over the years at Josephinum Research and was always available for advice. Jürgen Karner always believed in me and motivated me to try new things like starting my PhD. Without the support of Heinrich Prankl, I would not have been able to do my PhD while also working at Josephinum Research. He also believed in me and his inspiration and vision were instrumental in my research. I am grateful to all my colleagues at Josephinum Research and the BLT who supported me at work, especially during the field trials. They stood by me even when it was 30°C or when evening was fast approaching and we had to collect the last bits of data to finish the trial. It is truly a pleasure to work with them. I would like to thank Pöttinger Landtechnik GmbH and its employees for collaborating with me on my research.

I also must thank my family, especially my mother, for its love and unconditional support in all my endeavors, and my friends for keeping me sane through the years. Finally, I want to thank Karin for her love, care and support.

The research leading to this work received funding from the Lower Austrian government and the Austrian Research Promotion Agency. In particular the work was funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) under the program "Bridge 1" in the project SmartSeeding between November 2015 and December 2017 and in the project Mayday between March 2018 and February 2021. The work received funding from the Lower Austrian government as part of the project SoilCover from July 2015 until November 2017.

Contents













1	Introduction	1
1.1	Problem Statement	1
1.2	Goals	3
1.3	Contributions	4
1.3.1	Publications	6
1.4	Results Preview and Outline	7
2	Related Work	9
2.1	Vision Systems for Plant and Soil Parameters on Agricultural Machines	9
2.2	Soil Cover Measurement	9
2.3	Soil Roughness Measurement	11
2.4	Measuring Grassland Yield	12
2.5	Semantic Segmentation	14
2.6	Summary	15
3	Soil Cover Measurement	16
3.1	Hardware Setup	16
3.2	Dataset	18
3.2.1	Data labelling	18
3.2.2	Public datasets	19
3.3	Entangled Random Forest	19
3.3.1	Random Forest	19
3.3.2	Image features	21
3.3.3	Image pre-processing and parameter settings	22
3.4	Convolutional neural network	23
3.4.1	Convolutional neural networks for semantic segmentation	24
3.4.2	ERFNet	24
3.5	Evaluation	26
3.5.1	Validation with the Grid Method	27
3.5.2	Validation including Human Error	29
3.5.3	Inference Speeds	29
3.6	Discussion	30
4	Soil Roughness Measurement	31
4.1	Hardware Setup	31
4.1.1	Hardware Requirements	32
4.1.2	Test Setup	32

4.2	Roughness Estimation	33
4.2.1	Camera Calibration	33
4.2.2	Roughness Estimation Pipeline	34
4.2.3	Roughness Indices	34
4.2.4	Handling of External Influences	35
4.3	Online Machine Control	37
4.4	Field Trials	39
4.4.1	Soil sampling and sieve analysis	39
4.4.2	Experimental Design	39
4.5	Evaluation	42
4.5.1	Aggregate Size Distribution in the Seedbed	42
4.5.2	Relations between Mean Clod Diameter and Stereo Vision Roughness	44
4.5.3	Working Intensity and Controller Effect	46
4.5.4	Effects of External Influences	49
4.5.5	Soil Roughness and Plant Emergence on a Field	49
4.6	Discussion	50
5	Grassland Yield Measurement	53
5.1	Grassland harvesting	54
5.2	Hardware Setup	55
5.2.1	Hardware Requirements	55
5.2.2	Test Setup	56
5.2.3	Additional Sensors	56
5.3	Volume Estimation during Mowing	59
5.4	Family and Genus Classification	62
5.4.1	Dataset	63
5.5	Yield Modelling from Sensors on the Mower	64
5.6	Swath Measurement	65
5.6.1	Swath Segmentation based on Depth Information	67
5.6.2	Swath Segmentation based on RGB Image	67
5.6.3	Swath Volume Estimation based on Laser Scanner Data	68
5.7	Field Trials	69
5.7.1	Outdoor Laboratory Test for Swath Measurement	71
5.7.2	Comparison of different Raking Concepts	72
5.7.3	Targeted Trial for Grassland Family and Genus Classification	72
5.8	Evaluation	73
5.8.1	Volume Estimation and other Sensor Measurements during Mowing	73
5.8.2	Grassland Family and Genus Classification	76
5.8.3	A yield estimation model	78
5.8.4	Swath Segmentation and Volume Estimation	83
5.8.5	Pre-training for Semantic Segmentation	86
5.9	Discussion	87
5.9.1	Volume Estimation during Mowing and Yield Modelling	87
5.9.2	Family and Genus Classification	89

5.9.3	Swath Segmentation and Volume Estimation	89
5.9.4	Pre-training for the Semantic Segmentation	90
6	Conclusion and Outlook	91
6.1	Summary	91
6.2	Outlook	91
6.2.1	Soil Cover Measurement	91
6.2.2	Soil Roughness Measurement	92
6.2.3	Grassland Yield Measurement	92

List of Figures

1.1	Soil erosion on sunflower field (photo credit: Elmar Schmaltz, Bundesamt für Wasserwirtschaft, Institut für Kulturtechnik und Bodenwasserhaushalt)	2
1.2	Nitrogen fertilization on field (photo credit: Innovation Farm Wieselburg, Josephinum Research)	3
1.3	Left: Person capturing an image of the soil surface for soil cover estimation; Middle: image of the soil surface; Right: classification of the soil cover components in the image by a CNN	5
1.4	Left: Tractor with seed drill combination during seeding; Middle: camera view of the soil surface; Right: 3D stereo reconstruction of the soil surface for roughness estimation	5
1.5	Left: Tractor and mower with cameras and sensors mounted; Middle: Camera view on the cut edge; Right: top: classification of legumes and grasses in the image, bottom: 3D stereo reconstruction of the scene for plant height measurement.	6
2.1	Example soil cover segmentation of the method proposed in [16]: Stepwise application of eCognition A) original RGB image B) segmented image C)performed classification with different classes (brown - soil, green - living vegetation, pink - residues, yellow- areas of uncertainty).	11
2.2	Figure of the proposed work of [77]: Microsoft Kinect™ 3-D camera sensors mounted on a tractor with a cultivator for soil roughness measurements	12
2.3	Images of the method proposed in [94]. Left: on-the-go application as a combined real time measurement of ultrasonic sward height and high precision GPS in movement; Middle: experimental layout; Right: point cloud of the sward height measurements with an ultrasonic sensor	13
3.1	Android App for capturing images on smartphones. The applications communicates via a web-service to an inference server.	17
3.2	Hardware for capturing RGB images on a mobile machine. The system includes a DS RGB UI-3251LE-C-HQ camera and a NVIDIA Jetson Nano for inference of CNN models.	17
3.3	Semi-automatic segmentation for ground-truth annotation. A simple interface enables the adjustment of super-pixel size and a color threshold. Finally, the user needs to set labels to pre-segmented areas. From left to right different settings of the pre-segmentation are shown. The right image depicts a semi-transparent overlay of a partially annotated image.	18
3.4	Workflow within the training stage of the random forest.	20

3.5	Random Forest, consisting of a set of decision trees which provide a leaf node distribution for the learned classes, e.g. soil, organic and inorganic material. The node index array holds the current classification for each pixel and is updated every layer. The initial training samples and the splits within the nodes are selected randomly (symbolized by the dices).	20
3.6	Pairwise color comparison (pairwise feature) for a pixel X.	21
3.7	Effect of tree depth on F1 score; The curve PF (dashdotted line) is the forest variant with pairwise features, PF + EF (double line) denotes the variant with additional entangled features and PF + EF-col (dashed line) is a forest variant with a color constraint for the entangled features. The variant PF + EF-col + texture (solid line) includes linearness/pointness features. The curve upper bound (dotted line) shows the F1-Score for validation where the test-set equals the training-set.	23
3.8	Non-bottleneck-1D layer as implemented in the bonnet framework [116].	25
3.9	Examples of different soils surface covers (top) and the classification results (bottom). living plant  , residue  , soil  , stone 	26
3.10	Examples of different soils surface covers (top) and the classification results (bottom). living plant  , residue  , soil  , stone 	27
3.11	Comparison between reference annotations and entangled random forest results using the grid method () for (a) soil, (b) living organic matter and (c) dead organic matter ($N = 99$). The regression equation and the coefficient of determination are shown at the bottom of each graph in orange.	28
3.12	Comparison between reference annotations, entangled random forest results (), ERFNet () and the method proposed in [16] (), for (a) soil, (b) living organic matter and (c) dead organic matter. The regression equations and the coefficients of determination are shown in color for all three methods.	28
4.1	LI-OV580-STEREO (left), LI-USB30-V024STEREO (right)	32
4.2	Depth resolution Δz depending on z for LI-USB30-V024STEREO and LI-OV580-STEREO cameras with full resolution of 2688 pixel and reduced resolution of 640 pixel (assuming no sub-pixel accuracy).	33
4.3	Estimation process from the camera images to the final roughness value.	34
4.4	Height measurements for calculation of the RC indices, b) angles of the surface normals for calculation of the average angle of normals (AoN). .	35
4.5	Maximum exposure time $t_{s,max}$ for different driving speeds $v_{machine}$ and overlapping factors η	36
4.6	Concept of the tillage machine control.	37
4.7	Closed loop controller.	38
4.8	Camera mounted on the power harrow (left). Detailed view of stereo camera with Odroid XU4 and headlights (right)	38
4.9	Sample images with corresponding RC and AoN values in ascending order.	44
4.10	(a) Estimation of the mean weighted diameter of the soil aggregates. (b) Estimation of the soil surface roughness.	45

4.11	Measured results of MWD (x-axis) and RC roughness from the stereo camera (y-axis) for experiment 1 (top) and experiment 2 (bottom). . . .	45
4.12	Roughness RC results at different working intensities (tine loops per m). Each point equals the mean value of one track. $N = 24$	47
4.13	Controller effect: The diagram shows three tracks side by side, with the controlled track in the middle (blue line), and two tracks with constant PTO speeds (gray line). The PTO speed of the controlled track is shown by a black line. The target value for the controller was $RC = 14$	48
4.14	Effects of external influences on the image and the corresponding q values. a) shows a regular example with good stereo detection quality during motion. b) is an example on the same field where dust was created deliberately by stopping the tractor turning the tines on the stand. No drafts remove the dust atmosphere. c) shows a scenario with residues in which peaks are removed from the disparity map to reduce their influence. d) is an image taken during night without lighting. This scenario is also detected by image parameter mean brightness.	49
4.15	Soil roughness (left) and soil cover (right) estimated on a field in Wieselburg (AT). Soil cover is estimated with the entangled random forest method presented in Chapter 3.	50
5.1	Tractor and mower with cameras and sensors mounted, and people taking yield samples in the background.	53
5.2	Typical tasks during harvesting grassland. Grass is mown (a), followed by raking after drying (b) and removal by a baler (c). Grass sward ■, mown grass and swath ■ and grass turf ■	55
5.3	Cameras setup with two RGB cameras, a NIR camera and a NVIDIA Jetson Nano for inference of CNN models.	56
5.4	Pendulum-meter on the mower for measuring the deflection angle of the mower cloth.	58
5.5	Sensors on the mower: one camera setup is facing forwards towards the grass sward and one backwards on the cut edge. The sensors are mounted on the front facing downwards into the grass sward.	59
5.6	Cameras on the mower. One camera is facing forwards towards the grass sward and one camera is facing backwards on the edge between the grass sward ■, the grass turf ■ and the mown grass ■	60
5.7	Front camera view on the mower. Left RGB Image, NIR image (850 nm) and Right RGB image.	61
5.8	Rearview camera on the mower with grass sward (uncut plants), the grass turf, the mower cloth and the mown grass visible. Left RGB image, NIR image (850 nm) and right RGB image.	61
5.9	Results of the rearview mower segmentation: four examples of images captured on the mower and their corresponding segmentation results below. grass sward (uncut plants) ■, grass turf ■, mower cloth ■, mown grass ■.	62

5.10 Example images of the sugar Beet dataset [118] (left) and the grass clover image dataset [136] (right). 63

5.11 Two challenging examples from our grass legumes dataset with motion blurring (top) and high dynamic range by direct sunlight (bottom). Grass ■ and legumes ■ 64

5.12 Cameras on the rake facing towards the swath. 65

5.13 Cameras and sensors mounted on the rake facing backwards on the swath. 66

5.14 Captured images of a swath. Left RGB image, NIR image (850 nm) and right RGB image. 66

5.15 Swath segmentation based on depth data 67

5.16 The swath cross-sectional area estimation from the laser scanner data. Points from the LMS111 are shown in gray. The approximated ground plane is shown in green and all points above the plane after the first intersection are marked in red. 68

5.17 Manual weighing at random sample points during mowing and raking with the suspension scale and a tarpaulin. 69

5.18 Outdoor laboratory trial to validate the swath measurement system under controlled conditions. 72

5.19 Data from the forwards facing camera. The individual field trials are color coded. 74

5.20 Data from the backwards facing camera. The individual field trials are color coded. 75

5.21 Example image (left), prediction (middle) and ground truth labels (right). Grass ■ and legumes ■ 76

5.22 The Pearson correlation between the legumes-grass ratios at seeding and at the weekly measurements. 77

5.23 The Pearson correlation between the legumes-grass ratios at harvest and at the preceding weekly measurements. 77

5.24 An example plot of the weighed and the predicted legumes portions for 2020-07-22. 78

5.25 The predicted dry matter yields from the features estimated by the front facing camera. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis. 79

5.26 The predicted dry matter yields from the features estimated by the front facing camera without the NDVI. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis. 80

5.27 The predicted dry matter yields based solely on the height measurements of the front facing camera. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis. 80

5.28 The predicted dry matter yields from the features estimated by the rear facing camera. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis. 81

5.29	The predicted dry matter yields from the features estimated by the rear facing camera after shifting an entire field trials from the training set to the test set. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis.	82
5.30	The predicted dry matter yields only based on height measurements estimated by the rear facing camera after shifting entire field trials from the training set to the test set. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis.	83
5.31	The swath segmentation results: the sample images followed by the manual labels, the segmentation results based on the depth data and the segmentation results based on the CNN	84
5.32	Results of the outdoor laboratory test for the swath volume measurement. The estimation with a third of the original image resolution (a) and the estimation with half of the original image resolution (b).	86
5.33	The relationship between the estimated swath volume and the weighed swath mass for the different rake concepts	86

List of Tables

3.1	Layers of implementation of ERFNet [112] in bonnet framework [116]. The number of output features and resolution are for input images of 512x384.	25
3.2	Standard deviation between soil cover estimates of 10 different people. .	29
3.3	Inference speeds of ERFNet on different devices with an image resolution of 512x384 px.	29
4.1	Field Experiments design	41
4.2	Characteristics of the soils on the trial sites.	41
4.3	Characteristics of the soils from the additional experiment to observe plant emergence.	42
4.4	Soil aggregate fractions (%) and mean-weight diameter in the seedbed horizon, different tillage operations (n=36), experiment 1.	43
4.5	Soil aggregate fractions (%) and mean-weight diameter in the seedbed horizon, different primary tillage operations (plough, cultivator and disc harrow) and PTO speed of the power harrow (300, 540, 1000 1 min^{-1}), n=3, experiment 2.	43
5.1	Field trials mower	70
5.2	Field trials rake	71
5.3	Clover-grass mixtures seeded for the targeted family and genus classifi- cation trial.	73
5.4	Correlation between sensor readings and the weight of the dry matter yield (DMY) for the sensors on the mower.	76
5.5	The correlation between the estimated volume by the camera and the weight yield mass for the rake field trials.	85
5.6	Model accuracy and IoU, with and without pre-training on the test datasets	87

1 Introduction

With a growing world population, the demand for food and the pressure to produce in economically efficient and sustainable ways increases. The intensification of agricultural production combined with wrong agronomic decisions may lead to nitrogen pollution in groundwater or damage to the soil structure and consequently lead to soil erosion. In addition, climate change intensifies heavy rain which is accompanied by strong soil erosion [1].

Soil erosion can be influenced by agricultural techniques, such as direct drilling, mulching or reduced tillage [2]. The goal of these techniques is to preserve soil structure and to maintain a certain amount of soil cover. For conventional tillage and seeding these goals can be achieved by reducing the tillage intensity and maintaining the soil roughness.

Nitrogen pollution can be reduced by the targeted application of fertilizer based on the potential yield. Being able to estimate and predict yield site-specific forms the foundation for targeted application [3], as well as, optimized logistics and farm resource panning.

Decisions in agriculture are traditionally based on gut feeling and the experience of farmers. To base these decisions on measurable parameters and to be able to automate processes, new methods and tools are needed. Cameras with state-of-the-art vision algorithms are promising for measuring soil and plant parameters since farmers have always based most of their decisions on visual information.

Due to automation and the trend towards autonomous robots and sensors systems, cameras will also become ubiquitous in agriculture. With automation, not only the driving process but also the agronomic process like implement control must become automated. Currently, farmers use intuition and experience to carry out soil cultivation, plant cultivation or harvesting in the best possible way. All these intuitive processes including assessing the of work quality must be studied and mechanically replicated. Cameras with modern algorithms offer a perfect solution to master these challenges.

This thesis investigates methods for measuring soil and plant parameters based on state-of-the-art computer vision techniques, to solve different problems with one camera system and similar technical methods, especially stereo cameras and convolutional neural networks (CNN).

1.1 Problem Statement

The loss of topsoil erosion worldwide amounts to around 23 to 26 billion tons per year (Figure 1.1). This amounts to an annual loss of about one percent of arable land. Different agronomic measures, especially during tillage, help to combat this trend.



Figure 1.1: Soil erosion on sunflower field (photo credit: Elmar Schmaltz, Bundesamt für Wasserwirtschaft, Institut für Kulturtechnik und Bodenwasserhaushalt)

The main factors influencing soil erosion are soil aggregate stability and size, and soil cover [4]. During rain, the impact of rain drops on the soil surface breaks up soil clods and subsequently washes soil away. Additionally, bigger soil aggregates reduce and decelerate this process. Soil cover absorbs the kinetic energy of the rain drop and thereby reduces erosion. In general, bigger soil aggregates and soil cover are favourable conditions. Soil cover improves infiltration and retention of soil moisture and therefore, reduce crop water stress [4]. Additionally, soil cover regenerates soil and increases humus formation, thereby binding carbon in the soil. These benefits have been reported many times ([5], [6] [2]). The goal of tillage is to improve the soil conditions for plant growth by reducing the soil aggregate sizes [7]. The aggregate size distribution can be influenced by the type of tillage implement, the tillage intensity and the soil water content during tillage (Adam and Erbach, 1992; Braunack and McPhee, 1991). This often stands in contrast to the goal of erosion control, which favors bigger soil aggregates.

Most soil erosion models developed since the well-known USLE [8] model include soil cover as factor for soil loss. Soil erosion is influenced by additional factors like slope, rainfall and wind, but these are difficult or even impossible to change while the soil parameters themselves can be improved. Agricultural methods and machines for soil conservation management like direct drilling or soil mulching [9] have been developed. The work in [10] found that overall soil conservation techniques may reduce soil erosion by as much as 80%. National agricultural subsidy programmes often fund the soil protection measures of farmers, for instance the Austrian OEPUL programme [11]. Repeatable measurement methods for the evaluation of the target soil parameters are hardly used in these programs due to the lack of these methods.

The aspects mentioned above lead to a scientific as well as practical need for methods to measure soil parameters, soil cover and roughness. These methods should be repeatable, robust and reliable. The real-time capability and the motion insensitivity of the methods is also of importance, especially when used on moving machines like

tractors or tillage implements.

Soil cover demonstrates that not only soil parameters, also plant parameters are of importance. Plant parameters are of greater short-term interest to the farmer. The overall goal is to optimize yield in an environmentally friendly way. The basis for this and of greatest importance are non-invasive methods to measure the yield, but plant stress and biomass are of interest too. Especially in Austria, grassland is widespread but there is little existing research. Yield estimation and plant classification in heterogeneous grassland are the basis for automated machine control, site-specific fertilizer applications, in-field-logistics and logistic optimization during harvest and objective farm resource planning. From an environmental point of view, site-specific fertilization (Figure 1.2) has the potential to reduce water pollution by nitrates.



Figure 1.2: Nitrogen fertilization on field (photo credit: Innovation Farm Wieselburg, Josephinum Research)

The challenge for yield estimation lies mainly in the heterogeneity of grassland which is why there is no robust method yet proposed. A data-driven machine learning approach is therefore appropriate. In general, no reliable methods for the estimation of site-specific yield and also for plant classification in grassland are available.

1.2 Goals

The overall agronomic goal is to reduce soil erosion and reduce nitrogen pollution while maintaining good plant emergence and high yields. Soil erosion is influenced by weather, soil type and texture, soil cover, soil roughness, root penetration, soil moisture and many more. Not all of these parameters can be captured with a vision system, so we target those that can be measured by cameras and influenced through tillage, these are soil roughness and soil cover. Nitrogen pollution can be prevented by targeted and site-specific fertilization. Plants should only be supplied with as much nitrogen as they

really need, for this purpose their condition, in particular the expected yield, must be determined to infer fertilization maps.

The goal of this thesis was to develop and evaluate a vision system for application on moving agricultural implements that is able to measure the parameters mentioned above, soil cover, soil roughness and grassland yield. The system should include different use cases by using state-of-the-art computer vision methods that offer the potential to reduce soil erosion and to optimize resource usage via precision farming. The three main methods applied for this are described in detail below.

- Methods for measuring **soil cover** allow for estimating the protection against soil erosion, as well as, for quantifying plant emergence. Soil cover reduces the kinetic energy of rain drops impacting the soil surface and thereby decreases soil erosion.
- A method for **soil roughness** measurement evaluates the erosion protection of the soil and enables online machine control for homogeneous plant emergence. A certain level of soil roughness reduces soil wash-off which in turn also prevents erosion.
- Methods to measure **grassland yield and yield composition** enable resource optimized grassland management. Yield estimation in grassland forms the basis for site-specific fertilization which reduces nitrate pollution in the groundwater.

All methods are designed to work on moving agricultural implements but can also be used independently for plant and soil assessment in crop trials. The core technologies used are semantic segmentation and 3-D imaging with stereo cameras.

1.3 Contributions

In this section, the contributions of this thesis are presented grouped by the three open problems to be tackled by a camera system. The main contributions are new methods for soil cover measurement, soil roughness measurement and grassland yield estimation. The requirements of the methods are defined and implemented in a prototype vision system. In detail the contributions are:

Soil cover measurement methods to classify images into the classes soil, living organic matter (plants), dead organic matter (residues) and stone. Implementation and evaluation of two methods based on an entangled random forest (published in [12] and [13]) and an convolutional neural network (published in [14] and [15]).

- Validation of the two proposed methods with the grid method
- State-of-the-art comparison with the method proposed in [16]
- Estimation of the human evaluation error an experiment was conducted employing 10 persons, with each person evaluating identical images.

In Figure 1.3 the process of capturing a soil cover image and the estimated results are shown.



Figure 1.3: Left: Person capturing an image of the soil surface for soil cover estimation; Middle: image of the soil surface; Right: classification of the soil cover components in the image by a CNN

Online soil roughness measurement method and integration into an online machine control system (published in [17], [15], [18] and [19]).

- Integration into a closed loop control system for online machine control
- Statistical comparison of measured soil roughness and soil aggregate size on the field
- Investigation of the effect of the controller
- Investigation of environmental influences on the measurement system and proposed countermeasures
- Statistical comparison of measured soil roughness and resulting field emergence

Figure 1.4 (left) shows a tractor with a seed drill during seeding. The camera on the seed drill captures stereo images Figure 1.4 (middle) and reconstructs the soil surface Figure 1.4 (right) for roughness estimation.



Figure 1.4: Left: Tractor with seed drill combination during seeding; Middle: camera view of the soil surface; Right: 3D stereo reconstruction of the soil surface for roughness estimation

Methods for grassland yield and yield composition measurement and for measuring yield during mowing and swathing (published in [20] and [14]).

- Meadow height measurement system based on a stereo camera, with a new approach to absolute height measurement at the cut-edge.
- Method for estimation of grass/legumes ratio based on a convolutional neural network.
- Methods for segmentation of grassland swaths.
- Method for segmentation of areas of cut grass based on a CNN.
- Machine learning based regression model based on input parameters from the camera system to estimate dry matter yield.
- Statistical comparison of different measurement parameters to grassland yield.

Figure 1.5 shows a tractor during mowing and images of a stereo camera mounted on the mower facing backwards.



Figure 1.5: Left: Tractor and mower with cameras and sensors mounted; Middle: Camera view on the cut edge; Right: top: classification of legumes and grasses in the image, bottom: 3D stereo reconstruction of the scene for plant height measurement.

1.3.1 Publications

The following list gives an overview of publications that were published in the course of this thesis in chronological order starting with the most recent paper. The references [17] and [12] are published papers in peer reviewed scientific journals. The papers [14], [20], [15], and [21] are from peer reviewed conferences. The papers [22], [18], [19] and [13] are in the proceeding of non peer reviewed conferences. At the beginning of each chapter the respective papers are mentioned again.

- [22] **P. Riegler-Nurscher**, J. Prankl, M. Vincze, H. Prankl, "A review on plant and soil properties estimation using ground-based computer vision techniques in the visible spectrum.", *Tagungsband LAND.TECHNIK 2020*, VDI-Bericht 2374, ISBN: 978-3-18-092374-1, p. 353-359, 2020.
- [14] **P. Riegler-Nurscher**, J. Prankl, M. Vincze, "Semantic Segmentation for the Estimation of Plant and Soil Parameters on Agricultural Machines.", *The Sixteenth*

International Conference on Autonomic and Autonomous Systems, September 27, 2020 to October 1, Lisbon, Portugal, ISBN: 978-1-61208-787-0, p. 87-90, 2020.

- [20] **P. Riegler-Nurscher**, J. Prankl, M. Hofinger M., M. Vincze, "Detektion eines Grünlandschwades mit Stereo RGB Kamera", *40. GIL-Jahrestagung*, Digitalisierung für Mensch, Umwelt und Tier, Bonn: Gesellschaft für Informatik e.V., p. 265-270, ISBN: 978-3-88579-693-0, 2020.
- [17] **P. Riegler-Nurscher**, G. Moitzi, J. Prankl, J. Huber, J. Karner, H. Wagen-tristl, M. Vincze, "Machine vision for soil roughness measurement and control of tillage machines during seedbed preparation", *Soil and Tillage Research*, Volume 196, 104351, ISSN 0167-1987, <https://doi.org/10.1016/j.still.2019.104351>, 2020.
- [15] **P. Riegler-Nurscher**, J. Prankl, M. Vincze, "Tillage machine control based on a vision system for soil roughness and soil cover estimation", *Proceedings of the 12th International Conference on Computer Vision Systems (ICVS 2019)*, Thessaloniki, Greece, 23-25 September, 2019, p. 201-210, 2019.
- [21] **P. Riegler-Nurscher**, J. Prankl, J. Karner, "Beurteilung der Schnittqualität von Schneid- und Mähwerken mittels photogrammetrischer Methode", *39. GIL-Jahrestagung*, Digitalisierung für landwirtschaftliche Betriebe in kleinstrukturierten Regionen - ein Widerspruch in sich?, Bonn: Gesellschaft für Informatik e.V., p. 189-192, 2019.
- [18] **P. Riegler-Nurscher**, H. Prankl, G. Moitzi, H. Wagen-tristl, "Robust online soil roughness measurement using stereo vision", *EurAgEng Conference 2018*, Wageningen, Netherland 08 - 12 July, 2018.
- [12] **P. Riegler-Nurscher**, J. Prankl, T. Bauer, P. Strauss, H. Prankl, "A machine learning approach for pixel wise classification of residue and vegetation cover under field conditions", *Biosystems Engineering*, Volume 169, May 2018, p. 188-198, <https://doi.org/10.1016/j.biosystemseng.2018.02.011>, 2018.
- [19] **P. Riegler-Nurscher**, J. Karner, J. Huber, G. Moitzi, H. Wagen-tristl, M. Hofinger, H. Prankl, "A system for online control of a rotary harrow using soil roughness detection based on stereo vision", *Proceedings 75th Conference LAND.TECHNIK - AgEng 2017*, 10-11 November 2017, Hannover, VDI-Berichte Nr. 2300, 2017, ISBN 978-3-18-092300-0, p. 559-565, 2017.
- [13] **P. Riegler-Nurscher**, J. Prankl, T. Bauer, P. Strauss, H. Prankl, "An Integrated Image Analysis System for the Estimation of Soil Cover", *International Conference on Agricultural Engineering CIGR - AgEng 2016*, Aarhus, Denmark, 26 - 29, 2016.

1.4 Results Preview and Outline

The most significant results of the thesis are presented below.

Soil cover estimation is based on area ratios of semantic segmented RGB images. The estimation method based on the entangled random forest achieved an $R^2 = 0.763$ for soil, $R^2 = 0.981$ for living organic matter and $R^2 = 0.722$ for dead organic matter. The method based on a CNN achieved an $R^2 = 0.809$ for soil, $R^2 = 0.965$ for living organic matter and $R^2 = 0.747$ for dead organic matter.

Soil roughness measured with the proposed stereo camera correlates with $R^2 = 0.283$ to the mean weighed aggregate size measured by sieve analysis. A correlation with $R^2 = 0.844$ was found between measured soil roughness and working intensity of a power harrow. When controlling a power harrow based on the measured roughness the mean squared error (between RC roughness in mm) between targeted roughness and actual roughness decreased from $MSE = 2.16$ to $MSE = 1.24$ in the controlled case.

Grassland yield can be estimated with an $R^2 = 0.524$ based on a Support Vector Machine (SVM) model with features from the camera facing backwards onto the cut edge. This thesis discovered that the most important features are plant height and plant composition. Based on an estimation of the swath volume on a rake it is possible to predict yield mass with an $R^2 = 0.412$. An effect of the raking method on the volume-mass ratio was found. The new segmentation method based on CNNs exceeds depth based swath segmentation in terms of accuracy and IoU. It was also found that the composition of the seed correlates less and less with the composition of the plant over time.

Chapter 2, Related Work, gives an overview of agricultural RGB proximal sensing methods followed by prior scholarship done in agricultural sensor technologies and image analysis in our target applications. Lastly, recent developments in the field of semantic segmentation are discussed. Chapter 3, Soil Cover Measurement, presents the two implemented semantic segmentation approaches for soil cover estimation followed by an evaluation of the methods with state of the art metrics. In Chapter 4, the method for Soil Roughness Measurement is presented. The integration of the method into a controller for automatic machine control is also presented and evaluated. The Chapter 5 is about Grassland Yield Measurement. It presents the method of yield estimation and yield composition estimation during mowing and raking. The last Chapter 6 gives an overall conclusion and outlook.

2 Related Work

This chapter gives an overview of prior literature in the area of vision systems for plant and soil parameter estimation. The fields of soil cover and soil roughness measurement, as well as, grassland yield estimation are presented in detail. Recent advancements in semantic segmentation are presented as this computer vision task is very prevalent in this thesis and in agriculture in general.

2.1 Vision Systems for Plant and Soil Parameters on Agricultural Machines

The estimation of plant and soil properties is the basis for the sustainable cultivation of arable land. The use of ground based measurement techniques has the potential of not only mapping parameters but also controlling agricultural machines in real time. One of the most common uses for computer vision on agricultural machines is plant species classification [23], [24]. Various applications can be solved with plant classification. It forms the basis for spot spraying, mechanical weeding, plant phenotyping and yield estimation. The detection rates of the different approaches depend on the number of classes to be differentiated. For weeding, it is common to detect two classes (crop and weed). The estimation and classification of plant parts like leaves [25], stems, seeds [26], ears [27], panicles [28] are used for phenotyping and special applications like yield estimation. Plant row detection [29] is already widely-used in practice for inter-row weeding and automated steering. In contrast, plant pest detection ([30], [31]) is still a field of research and no widespread applications are known. Full-area real-time detection of pests in the field is not yet practical due to the high image resolution requirements. The same challenge concerns the detection of plant diseases ([32], [33]). RGB vision for soil parameters is mostly limited to soil roughness and soil clod size estimation ([34], [35]). Soil cover is composed of living organic matter like plants; dead organic matter like residues; and, stones. Its estimation lies methodologically in between that of plant and soil parameters.

2.2 Soil Cover Measurement

Soil cover is crucial for practising conservative agriculture, protecting soil and improving plant growth. Measuring soil cover accurately is important for planning soil tillage tasks, controlling agricultural implements, or evaluating effects of certain operations on fields. Soil cover is also part of most, if not all, soil erosion models which are presented in [36] and [37].

Soil cover is usually estimated with on-site methods ([38], [39]) like the line-intercept transect method or the lattice screen method. These methods try to achieve an objective systematization of the estimation using regular pattern of sample points. These points are classified manually by a human. These methods are very time consuming and lack reproducibility.

Manual image analysis methods have been introduced to improve reproducibility and usability ([40], [41], [42], [43]). With these methods, the classification task is carried out remotely to the location where the image was captured. This reduces the time on the field and enables a certain reproducibility of the estimation. However, all manual methods are time consuming and very subjective depending on the skills of the evaluator.

The obvious next step was to automate the image analysis. This allows for faster processing and it can be executed easily compared to manual cover estimation [44]. However, the line point intercept method, also known as grid method, has become an established method for validating automated image analysis methods.

Soil cover estimation can be done from satellites (remote sensing) or on the ground (proximal sensing). We focus on proximal sensing with camera systems which utilise color and texture information. Many studies in the field of proximal soil cover estimation focus either only on dead ([45], [46], [47]) or on living soil cover ([48], [49], [50], [51], [52], [53], [54]). There are hardly any methods known in the literature that combine both types of cover like in [16]. Many of these methods are based on simple color thresholding and often need color calibration (e.g. by color tiles)[49].

One way to reduce the effects of light and color variations is to select different color spaces. The work in [55] showed, that color spaces which separate luminance from chrominance work best for soil segmentation tasks.

Common object-based methods rely on sample-based analysis which is very time consuming, because the user has to define meaningful samples for each soil cover type. Object-based methods ([56], [44] and [57]) rely on defining thresholds and these thresholds need sometimes to be adapted for each image [44].

The method proposed in [16] uses object based image analysis and a decision tree to select parameters depending on manual preselected parameters. Although these show promising results, manual parameter adjustment is required and it is often limited to a certain application area like an upper limit of soil cover.

As in all areas of computer vision, machine learning also became popular for soil cover estimation. The work in [58] showed that machine learning based methods outperform classic threshold based methods in classification quality, especially in distinguishing soil from dead organic matter. The main issue with threshold based methods is the manual selection of the threshold and the resulting lack of generalization, especially under different lighting conditions. Machine learning based methods for soil cover estimation can be more illumination tolerant and generalize more when used with large datasets [59].

Early machine learning methods for soil cover classification often use Support Vector Machines (SVM) [60]. The method proposed in [59] uses a SVM for the segmentation of living plants and soil. Methods based on neural networks in a deep architecture like convolutional neural networks, as in [61], look very promising. They can also cope with

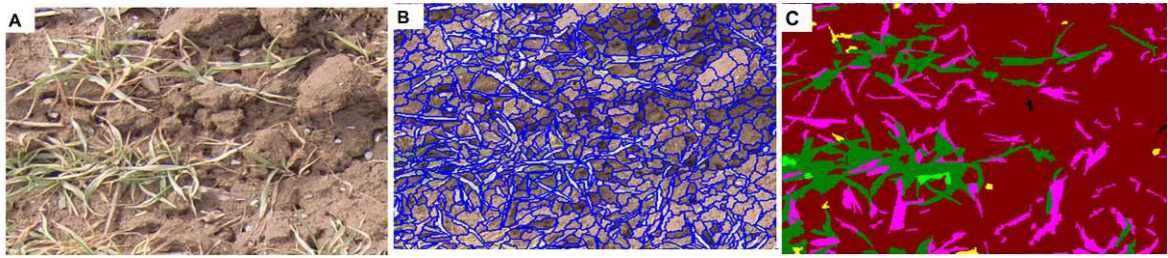


Figure 2.1: Example soil cover segmentation of the method proposed in [16]: Stepwise application of eCognition A) original RGB image B) segmented image C) performed classification with different classes (brown - soil, green - living vegetation, pink - residues, yellow- areas of uncertainty).

different lighting conditions and various soil cover types. However, the method in [61] is in an early prototype stage and has not been sufficiently evaluated.

The main challenge in deep learning, with a large number of parameters in the model, is the need for large datasets to get good generalization and to avoid converging into local minima.

2.3 Soil Roughness Measurement

Soil roughness is an important parameter to quantify the erosivity of soil. Rough soil reduces soil erosion [62] while fine soil and smaller soil aggregates improve plant emergence. Tillage and seedbed preparations improve soil conditions by altering the aggregate sizes [7]. Different studies have investigated the aggregate size distribution and its effects in the seedbed ([62], [63], [64], [65], [66], [67], [68], [69]). The goal of seedbed preparation is to get optimal soil roughness to avoid soil erosion and to get good plant emergence. Soil conditions in fields are in general heterogeneous ([70], [71]). To get homogeneous soil roughness during seedbed preparation, the tillage intensity has to be adapted according to the site-specific conditions. The site specific adoptions of the machine could be based on pre-processed maps or could be calculated in real time based on sensor measurements. No matter which method is used to control the tillage machine, a measurement system is needed. Different methods have been proposed for measuring soil roughness ([72], [73], [74], [75], [76], [77] [78], [79], [80]). Initial attempts to measure soil roughness used pin meters and chains [72]. These variants of contact based methods can't be used in motion and therefore cannot be mounted on moving machines.

Lasers are used in academic settings for measuring soil roughness in-situ ([73], [75], [74]). These laser scanners are very accurate for depth measuring [81], but these sub-millimeter accuracies are not necessary for soil roughness measurements at tillage. Furthermore, the laser scanners are very sensitive to mechanical vibrations and dust.

The Microsoft Kinect™ 3-D camera sensor was used for soil reconstruction in [77]. However, many active sensors like the Microsoft Kinect™ have difficulties operating in bright daylight due to increased noise levels.

Methods for roughness measurements based on multiple views with unknown extrinsic

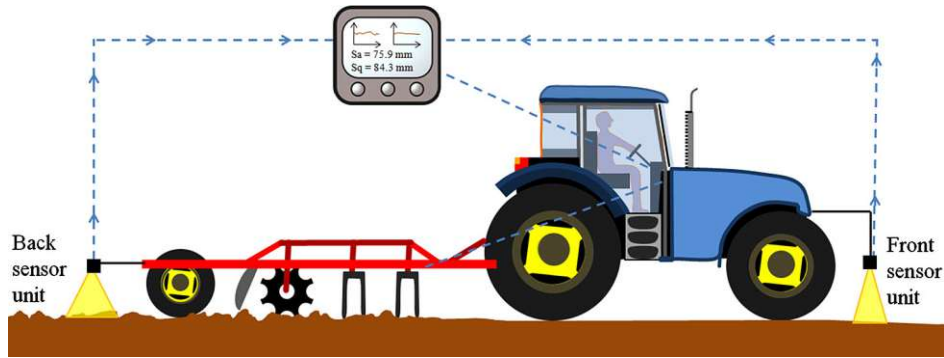


Figure 2.2: Figure of the proposed work of [77]: Microsoft Kinect™ 3-D camera sensors mounted on a tractor with a cultivator for soil roughness measurements

camera positions are used in [78] and [82]. These methods use additional reference patterns on the soil to get an accurate scaling of the reconstructions. However, this makes an application on a tillage machine difficult.

Stereo cameras have no need for reference patterns after calibration. Soil roughness measurements based on stereo cameras have been proposed in [79] and [80]. Stereo cameras have been proven suitable for outdoor applications in agriculture like the Claas CultiCam [83], they work well under changing lighting conditions and fast motion. The color information provided by RGB stereo cameras can be used to detect other soil parameters like soil cover as discussed in the previous section.

Methods to detect clods directly from images, in contrast to those which just measure roughness, have been proposed ([84], [76]). However, it is quite difficult to detect clods directly after tillage and in motion due to the soil aggregates being hard to see. Currently, there is no method known to be able to detect soil clods, of different soil textures and soil colors, during tillage.

2.4 Measuring Grassland Yield

Previous studies tried to measure grassland yield during mowing. The work in [85] used a belt weigher. The grass is loaded onto the conveyor belt which continuously measures its weight. A problem with this setup is the difficulty of weighing while in motion and the complexity of the mechanical system. Another proposed approach is to measure the torque of a conveyor belt and derive the yield from it. In the research in [86], this approach was tested in the lab with two conveyor belts, but led to a strong deviation in the measurement of the torque, especially in their subsequent field tests. In [87], a statistically significant correlation between the power of the conditioner on a mower and mass flux of the grass was found. A similar methodological approach was published in [88] which found a correlation between the conditioner power, force on the conditioner baffle plate and the mass flux. A mixture of grass and lucerne was used and it was shown that the mass flux is not only dependent on the torque of the conditioner but also on grass species, moisture, grass maturity and conditioner intensity. These dependencies did not manifest themselves on the force of the conditioner baffle plate. There are no



Figure 2.3: Images of the method proposed in [94]. Left: on-the-go application as a combined real time measurement of ultrasonic sward height and high precision GPS in movement; Middle: experimental layout; Right: point cloud of the sward height measurements with an ultrasonic sensor

production stage mass flux or yield measurement systems for mowers. Alternatively to measuring torques, the work in [89] tested a pendulum meter to capture the mass distribution on the field. Different plant distributions on the field correlate with different pendulum deflections [90]. The pendulum meter could deliver mass distributions of the plant on a field very cost efficiently by a non-destructive measurement. An integration into existing machines would be possible [91]. This shows that there was research in the field of yield measurement, but no product is currently available on the market. Issues lie in the calibration of the measurement systems and the heterogeneity of grasslands.

Laser scanners enable the measurement of crop heights to estimate the yield. In [92], laser measurements were compared with manual height measurements and georeferenced crop height maps were generated. Laser scanners are expensive and fragile which limits the use on mobile mower implements. The research in [93] used ultrasonic sensors with multi spectral sensors to measure yield mass in grasslands while stationary and in motion. Their research suggests that it is possible to measure grassland yield in heterogeneous fields by combining different sensor systems. It was shown that measurement systems based on ultrasonic sensors reach their limit on high plants ([94], [95], [93]). Combinations of leaf area index and vegetation indices could increase the accuracy of the measurements for higher plant stand. Spectral information alone would not be sufficient for estimation of grass yield [95]. The relation between biomass yield and spectral information for selected legumes and clover grass mixtures was investigated in [96]. The measurements showed that each legume species has its own spectral signature which can be measured accurately ([97], [98]). However, vegetation indices alone did not deliver satisfactory results [96]. The work in [98] showed correlations between vegetation indices (NDVI) and grassland yield. Models based on a combination of spectral information and grass height from ultrasonic sensor measurements [99] showed promising results for yield estimation. The main issues with the approaches presented lies in the difficulty of calibration, the expensive hardware and the susceptibility to errors at high working speeds. These issues prevent the practical application of the measurement systems mentioned in the literature.

In order to improve yield estimation, the species composition in grasslands is also a relevant parameter. Initial studies on the detection of grassland species composition have been presented. Multi-spectral sensors [100] but also RGB imaging methods [101] are used for species classification in grasslands. The most promising preliminary work was presented in [101]. Their proposed method distinguishes clover and grasses in RGB images by using semantic segmentation. Section 2.5 presents state of the art methods for semantic segmentation. We continue to pursue this approach to enable robust detection under real outdoor conditions during movement.

It is generally possible and useful to extend GNSS based yield mapping systems from grain production and precision farming into grassland, because the applications are similar [102]. But in order to help precision agriculture move from the field to the grassland, it is necessary that these technologies are simple and inexpensive, yet versatile [103].

In the field of swath detection, several approaches have been proposed. Current methods use ultrasound, radar or laser sensors, or cameras for this task. Systems based on Laser scanners are already commercially available [104] but expensive and thus not common. These systems are used to steer a tractor or control the driving speed depending on the swath volume, during harvesting with a loading wagon or a baler. Camera-based swath detection enables the utilization of additional information such as texture and color [105]. Especially at small volumes, methods based on 3-D data from laser rangefinders or 3-D cameras have problems with the swath segmentation. The method presented in [105] shows the potential for swath segmentation used on stereo depth data and texture information.

2.5 Semantic Segmentation

Semantic segmentation involves getting a pixel-wise classification mask of an image. In agricultural applications with strong illumination invariances and biological non-rigid objects, simpler methods like thresholding, histogram-based methods, or clustering methods do not achieve the desired classification and segmentation quality. In recent years, machine learning based methods like support vector machines [106], random forests [107] and especially convolutional neural networks (CNNs) are the methods of choice for complex semantic segmentation problems. End to end learning methods like CNNs circumvent the problem of manual feature design.

First CNNs for semantic segmentation were Fully Convolutional network (FCN). There a classical CNN takes an input image and the label map is produced in an additional fully connected layer. The convolutional and pooling layers of the CNN down sample the input resulting in low resolution output label maps. Different approaches like SegNet [108], UNet [109], DeepLab [110], Pyramid Scene Parsing Network (PSPNet) and Dilated Convolutions have been proposed to circumvent this issue. A survey on semantic segmentation based on deep learning methods was published in [111].

We utilized the entangled random forest method [107] and the ERFNet CNN [112] for the task of soil cover estimation (see Chapter 3). For applications in grassland in Chapter 5, we only use the ERFNet method because of its compromise between speed

and accuracy as shown in [113].

2.6 Summary

Methods for soil roughness and soil cover estimation, as well as, for grassland yield estimation have been proposed. However, all the proposed methods are tested under laboratory conditions or to a certain extent under controlled conditions. This is sufficient, among other things, for scientific application and agronomic trial activities. Additionally, a method to control a tillage machine in real time based on soil roughness readings has never been studied. The goal is to get these methods onto agricultural machines for the real time mapping of soil and plant parameters or for the real time control of machines. The robust real time application of computer vision on fields is the prerequisite for broad use in agriculture and the advancement of the field robotics.

3 Soil Cover Measurement

Soil cover consists of living and dead organic matter and stones on the soil surface. A certain amount of soil cover reduces the susceptibility of the soil to erosion. The cover on the soil reduces the amount of soil washed away during rain falls [4]. Soil cover not only increases the availability of plant nutrients, but also supports soil regeneration and enhances humus formation (see [5], [6], [2]). In order to measure the soil cover, an objective and easy to use method is needed. Many national subsidy programs like OEPUL [11] in Austria promote soil cover protection against soil erosion. A camera based method for soil cover estimation could be used to objectively assess the measures taken by the farmer and distribute the subsidies in a more focused manner. It could also be used as a tool to train the farmer to improve his subjective assessment of soil cover during practical use.

Agronomic field trials are another important area of application. The proposed method for measuring soil cover can be used for the quality assessment of tillage, as already published in [114] and [115], or to measure plant emergence.

In order to widely extend the application of this method, the measurement should work with different RGB camera systems, such as those of smart phones or industrial cameras. In contrast to the methods presented in Chapter 4 and 5, only RGB image data and no depth data is used for soil cover measurements. Two different machine learning methods for semantic segmentation were implemented, trained and evaluated. An entangled random forest, as proposed in [107], and a convolutional neural network (CNN), the ERFNet [112] implemented in the bonnet framework [116]. We implemented this method based on the entangled random forest available online at www.soilcover.at.

Parts of this chapter were published in Biosystems Engineering [12], at the ICVS conference 2019 [15] and at the CIGR - AgEng Conference 2016 [13].

3.1 Hardware Setup

Soil cover is defined by its color and less so by its 3-D structure. Hence, the minimum hardware requirements are a RGB camera which can read the soil surface with a resolution of at least 1 mm. Due to the simple hardware requirements and the different possible applications, the soil cover estimation algorithm should work for many different cameras.

Different cameras were used to capture the training data. Images were captured with DSLRs, smartphones and handheld digital cameras (Nikon D50; Nikon D5100; Casio EXZ400, Nokia C700; Nikon D7000; LGD405; Panasonic DMCTZ61, Sony DSC-H9 and Sony DSC-H70).

Different RGB cameras can be used during inference. An Android application was

implemented to simplify the use of the smartphones. The application captures an image with the Android Camera API and sends it to a server which runs the inference and sends the result back.

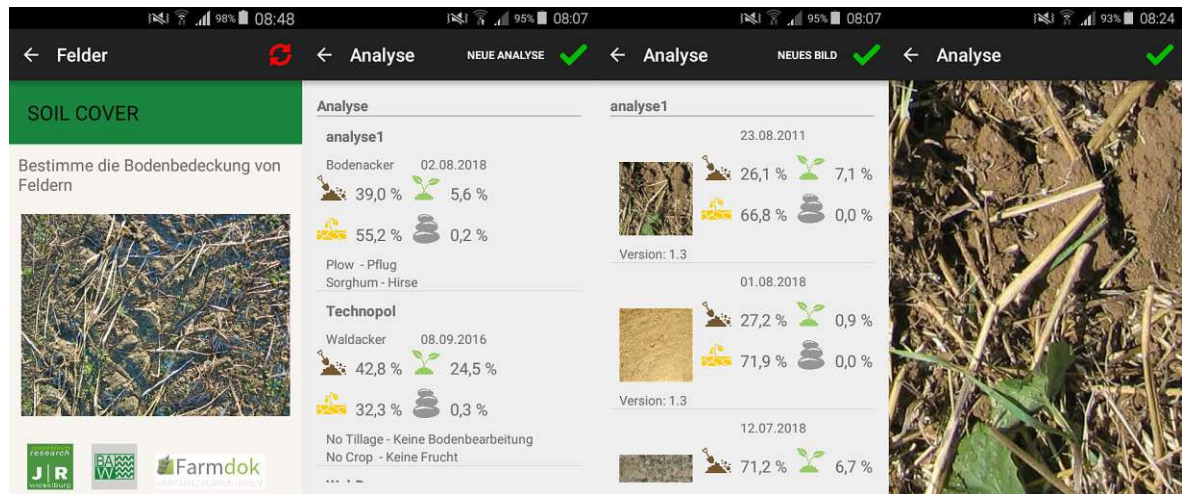


Figure 3.1: Android App for capturing images on smartphones. The applications communicates via a web-service to an inference server.

In order to gather large datasets of soil images, a camera setup with a single board computer was implemented. Figure 3.2 shows the camera setup developed for capturing soil cover images in motion with an IDS RGB UI-3251LE-C-HQ camera (1/1.8" e2v EV76C570 CMOS sensor) and a NVIDIA Jetson Nano single board computer. The camera has an imaging resolution of 1600 x 1200 pixels. A Lensation 5.5 mm lens was used to capture a surface area of 70 x 60 cm from of about 70 cm mounting height above the soil surface. The system runs a gpsd daemon from libgps to label each image with its GNSS position and to read a velocity relative to the soil surface for optimizing the camera exposure time. The handling of the exposure time, dependent on machine velocity, works as described in Section 4.2.4. The Jetson Nano SBC also allows for real time inference of CNNs depending on the image resolution and the network size.



Figure 3.2: Hardware for capturing RGB images on a mobile machine. The system includes a DS RGB UI-3251LE-C-HQ camera and a NVIDIA Jetson Nano for inference of CNN models.

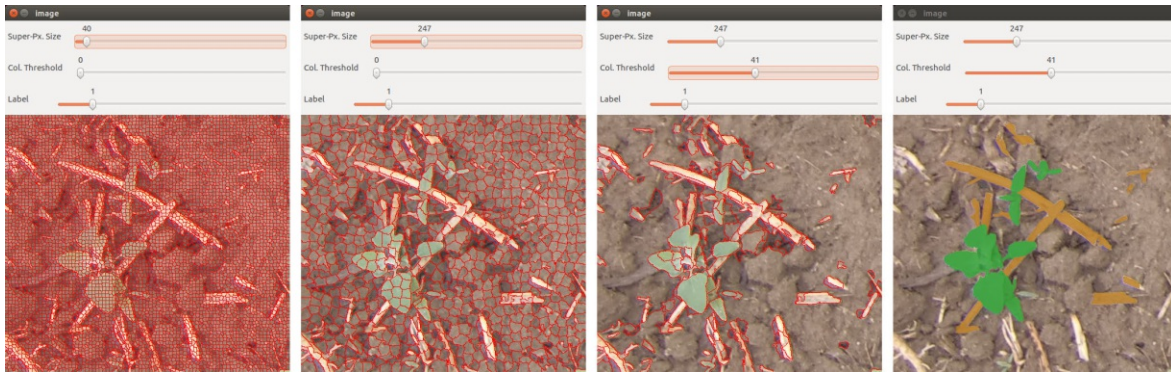


Figure 3.3: Semi-automatic segmentation for ground-truth annotation. A simple interface enables the adjustment of super-pixel size and a color threshold. Finally, the user needs to set labels to pre-segmented areas. From left to right different settings of the pre-segmentation are shown. The right image depicts a semi-transparent overlay of a partially annotated image.

3.2 Dataset

The supervised machine learning approaches utilized in this work are dependent on labelled training images. Training images were captured on the field and a pixel wise class label mask was manually created.

In order to capture a certain range of natural scenarios, images of different residue and vegetation covers were taken. The images were taken from about a height of 1.4 m horizontally to the ground which resulted in a resolution of 1 to 10 pixel/mm. First, the conditions were restricted to diffuse (cloudy) light conditions. The image acquisition was done between May 2015 and October 2016 for crops, catch crops and spontaneous vegetation with diverse plant residues. Finally, a dataset of about 200 different images was created.

3.2.1 Data labelling

We implemented a labelling tool with semi-automatic pre-segmentation to ensure efficient and accurate labelling of the images. The pre-segmentation is based on super-pixel segmentation Simple Linear Iterative Clustering (SLIC), as proposed by [117], and a recursive clustering of nearby super-pixels depending on the color similarity. The user can modify two different parameters: the first parameter specifies the size of the super-pixels and the second one defines the color similarity for recursive nearest neighbor clustering. The tool has an additional option for free drawing with different pencil sizes and flood filling within the label mask. Temporal brightness changes can be controlled with gamma correction to increase visibility in dark or very bright image regions during labelling. With this tool, representative image regions of about 500 x 500 pixel of the 200 training images were labelled.

3.2.2 Public datasets

The method based on the CNN can be pre-trained on similar datasets. Afterwards, the soil cover dataset is used to fine-tune the weights of the model. The ERFNet CNN was pre-trained on the sugar beets dataset [118], which is publicly available. It consists of 12,714 RGB images of sugar beets. The images from the sugar beet dataset have a ground resolution of approximately 3 pixel/mm and capture an area of 24 x 31 cm.

3.3 Entangled Random Forest

The first method implemented is a variant of a Random Forest proposed by [107] and known as entangled random forest. This variant uses entangled features to model the context of individual pixels. These features allow for modelling such that neighboring pixels are more likely to be from the same class. Like all supervised machine learning methods, it uses annotated training data to fit a model of the data by minimizing the error to the training data. During inference, the learned model is applied on new images.

3.3.1 Random Forest

A random forest generally consists of several decision trees. These binary trees make decisions in each node based on a feature and its learned threshold. The trees are built during the training stage by optimizing the information gain in each node. The training of each tree is shown in Figure 3.4. The training method is principally based on the method published in [119]. First, a random sample of images is selected from all training images in a process known as bagging. Starting from the root node down, each node's parameters (feature and threshold) is selected based on random drawings from the range of features and thresholds. The combination of feature and threshold with the best information gain, i.e., which separates the training set best based on the class memberships of all pixels, is selected. Through this process, the tree is gradually built node by node and layer by layer downwards. If a certain information gain is reached in a node, it remains as a leaf node and receives no further child nodes. Furthermore, the training is also stopped if a limiting depth is reached which prevents overfitting of the tree, i.e. the model does fit the training data but fails by classifying new test samples. A sketch of a random tree is shown in Figure 3.5. The building of a tree is repeated N times to generate N trees.

During inference of new test samples, each pixel of the image is sent through each tree downwards and ends up in a leaf node. The leaf node holds a class distribution calculated during training. All class distributions from the trees are combined to vote for a class. The class with the strongest probability is selected and entered into the class mask of the image.

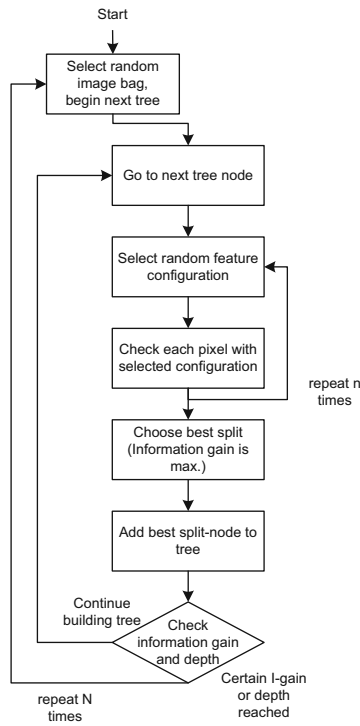


Figure 3.4: Workflow within the training stage of the random forest.

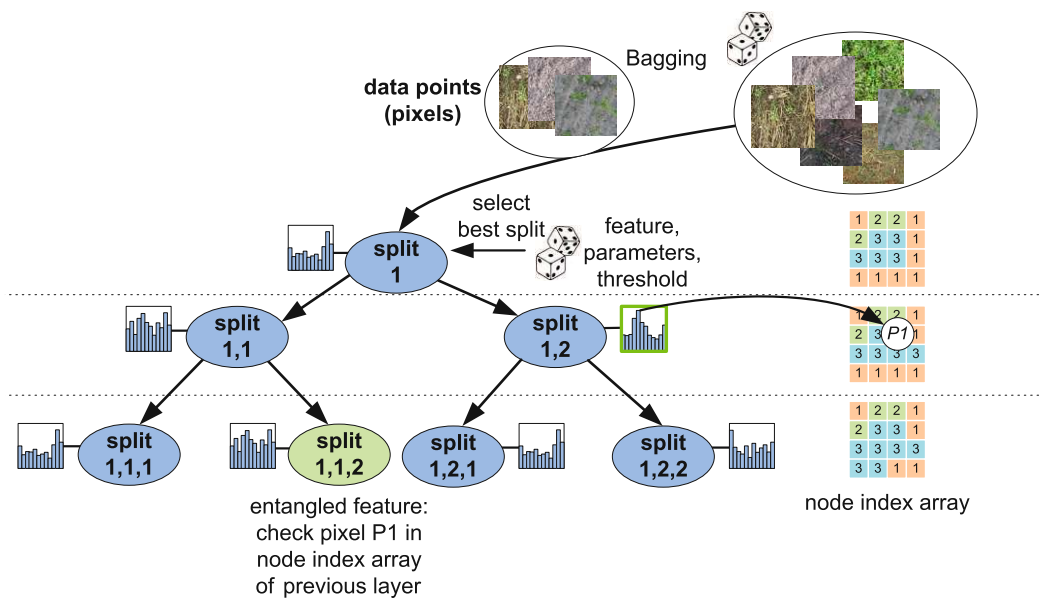


Figure 3.5: Random Forest, consisting of a set of decision trees which provide a leaf node distribution for the learned classes, e.g. soil, organic and inorganic material. The node index array holds the current classification for each pixel and is updated every layer. The initial training samples and the splits within the nodes are selected randomly (symbolized by the dices).

3.3.2 Image features

Each of the nodes splits the data depending on a feature and a threshold of the feature value. We use simple features in contrast to commonly used complex features, for example Textons [120], to reduce computation time.

The aforementioned entangled random forest uses two different types of decision functions: Features that model the color of individual pixels and color differences in the local neighborhood. Hence, these decision functions are able to learn the appearance, the structure and the shape of the soil cover. The second type of feature is the entangled feature; these maximum a posteriori features model class probabilities based on local neighboring pixel classes.

All color features are calculated in the LAB CIE 94 color space. The color channels are separated from brightness, therefore the influence of brightness changes can be reduced or completely skipped.

The simplest feature used is the absolute color feature $f_{color,\{A,B\}}$, where A and B stand for the corresponding channels in the LAB color space. In a region \mathbf{R}_1 at an offset Δ_1 from the pixel \mathbf{x} , the mean color value is computed. The rectangular region \mathbf{R}_1 and the offset Δ_1 are selected randomly during training. The feature for the pixel \mathbf{x} is calculated as in the following Equation 3.1.

$$f_{color,\{A,B\}}(\mathbf{x}; \Delta_1, \mathbf{R}_1) = \bar{I}_{\{A,B\}}(\mathbf{R}_1(\mathbf{x} + \Delta_1)) \quad (3.1)$$

Image gradients between two image regions are formalized by pairwise features $f_{colorDiff,\{L,A,B\}}$. The difference between two randomly selected regions \mathbf{R}_1 and \mathbf{R}_2 at the position Δ_1 and Δ_2 is calculated relative to the pixel \mathbf{x} . Color differences are inherently brightness invariant, therefore pairwise features are calculated on all LAB channels. The shape of the regions is arbitrary, so the features are able to reflect any geometric properties in a supporting area. Figure 3.6 shows the pairwise feature calculation between two points P_1 and P_2 on an image and the corresponding classification result. Pairwise features are calculated by the following Equation 3.2.

$$f_{colorDiff,\{L,A,B\}}(\mathbf{x}; \Delta_1, \mathbf{R}_1, \Delta_2, \mathbf{R}_2) = \bar{I}_{\{L,A,B\}}(\mathbf{R}_1(\mathbf{x} + \Delta_1)) - \bar{I}_{\{L,A,B\}}(\mathbf{R}_2(\mathbf{x} + \Delta_2)) \quad (3.2)$$

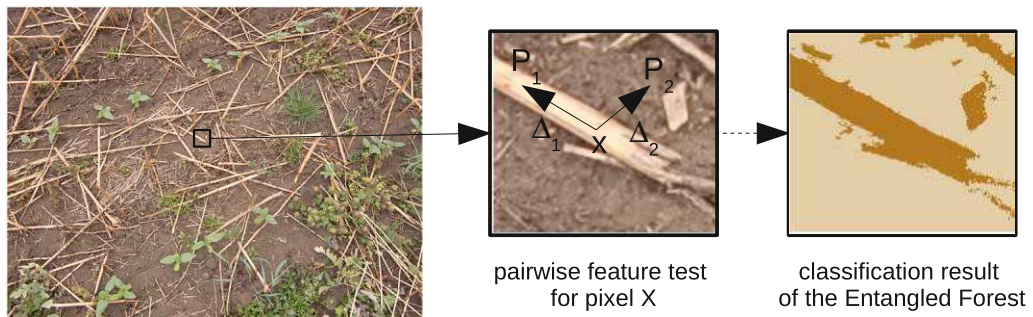


Figure 3.6: Pairwise color comparison (pairwise feature) for a pixel X.

The smoothness of a region \mathbf{R}_1 can be described by variance features $f_{var,\{L,A,B\}}$. Variance is also brightness independent and is calculated on all LAB channels. Variance features can be formalized as presented in Equation 3.3.

$$f_{var,\{L,A,B\}}(\mathbf{x}; \Delta_1, \mathbf{R}_1) = VAR_{\{L,A,B\}}(\mathbf{R}_1(\mathbf{x} + \Delta_1)) \quad (3.3)$$

Image gradients \mathbf{g} can be used to formalize the pointness or the linearness of a region \mathbf{R} around a pixel \mathbf{x} . Small values of $f_{lt,\{L\}}$ indicate linear structures like plant stems. The value of $f_{g,\{L\}}$ gives a measure for the reliability of the value. $f_{lt,\{L\}}$ and $f_{g,\{L\}}$ are formalized by Equation 3.4 and 3.5

$$f_{lt,\{L\}} = VAR_R \left(\arctan \left(\frac{g_y}{g_x} \right) \right) \quad (3.4)$$

$$f_{g,\{L\}} = MEAN_R (\|g_x\| + \|g_y\|) \quad (3.5)$$

The concept of entangled features $f_{MAPClass_{\{A,B\}}}$ is shown in Figure 3.5. The idea behind these features is that neighboring pixels come from the same soil cover type, hence they have the same label with a high probability. This enables the modelling of the context of individual pixels. During testing, the intermediate classification of the layer above is utilized in the current node. We extend the MAPClass entangled features of [107] by a color constraint to get Col-MAPClass features. These features test if the maximum a posteriori (MAP) class of a pixel P_1 is equals to a class C. The posterior class distribution $p(c; n_{P_1})$ of a node P_1 can be accessed because the tree is trained in a breadth first manner. The association between pixel and tree node is maintained in a node index array, as depicted in Figure 3.5. Equation 3.6 formalizes the entangled features. The color term enables a comparison between the mean color in a region R_1 at an offset Δ_1 to a trained threshold color value T_C . This enables a certain smoothing while the additional color term prevents smoothing across object boundaries. The threshold T is selected during training. This value determines whether the classical smoothing, as proposed in [107], or the color constraint option, in the case of a small T value, should be activated.

$$f_{MAPClass_{\{A,B\}}}(x, \Delta_1, P_1, C, T_C) = \begin{cases} \underset{c}{argmax}(c; n_{P_1}) = C \wedge \|\bar{I}_{\{A,B\}}(\mathbf{R}_1(\mathbf{x} + \Delta_1)) - T_C\| < T & 1 \\ otherwise & 0 \end{cases} \quad (3.6)$$

3.3.3 Image pre-processing and parameter settings

Some pre-processing steps are taken before the training and inference of the images. First, the RGB images are converted to the LAB CIE 94 color space. The corresponding gray scale image gradient angle and amplitudes are calculated. From all channels and gradient maps, integral images are calculated. Integral images allow for the fast computation of the features independent of the supporting region size \mathbf{R} . Bagging is applied to select 50% randomly sampled images for the training of each tree. This

bagging approach reduces the correlation between the trees. To counteract the unequal distribution of the classes, the random selection of images is guided towards, the equal selection of each class.

The number of trees is limited to 10 and the building of each tree stops at level 25. Experiments showed that after a depth level of 25 the error rate does not improve, see Figure 3.7. Figure 3.7 shows the F1-Score of a 5 fold cross validation for different feature combinations. The upper bound is the F1-Score for validation where the test-set equals the training-set. Growing is stopped at each node reaching an information gain below 0.01 or if less than 5 points are left in the current node. The features and thresholds are selected in a two step process. First, features and parameters are selected with equal probability. Afterwards, this prior distribution is used to guide the feature selection. This leads to a higher probability of selecting complex features with more parameters and consequently to a higher learning accuracy.

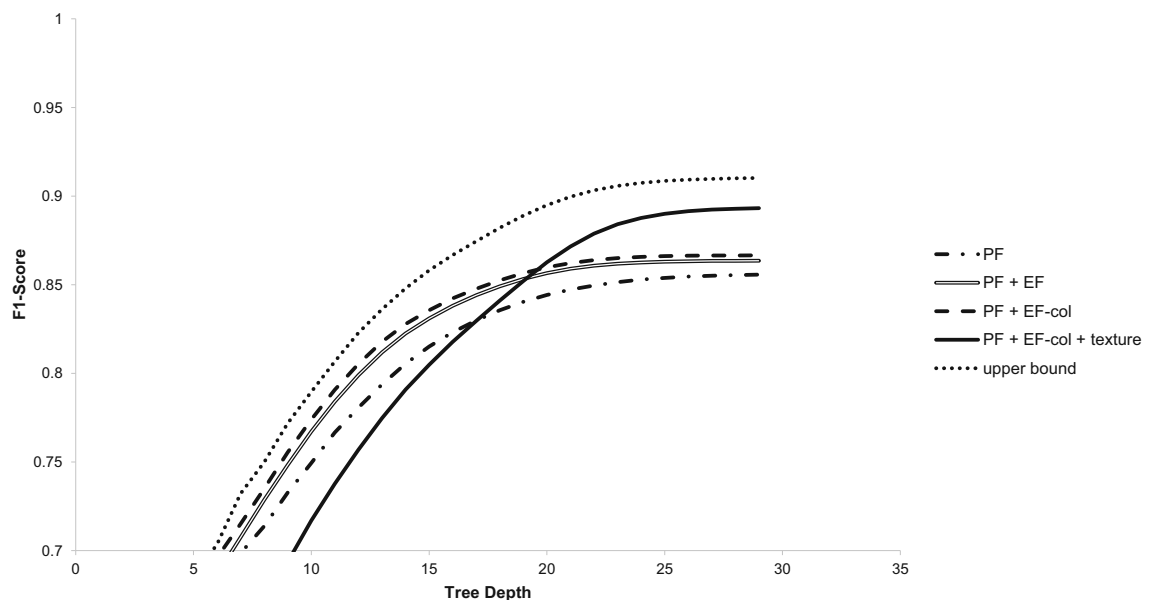


Figure 3.7: Effect of tree depth on F1 score; The curve PF (dashdotted line) is the forest variant with pairwise features, PF + EF (double line) denotes the variant with additional entangled features and PF + EF-col (dashed line) is a forest variant with a color constraint for the entangled features. The variant PF + EF-col + texture (solid line) includes linearness/pointness features. The curve upper bound (dotted line) shows the F1-Score for validation where the test-set equals the training-set.

3.4 Convolutional neural network

A convolutional neural network is a machine learning method using convolutions in the network graph. These convolutions enable the modelling of spatial relations in data and are shift invariant, as required for images. A CNN usually consisting of convolution

layers and pooling to reduce resolution and to extract features and fully connected layers. The weights and biases between the nodes of the network and the convolution kernels are parameters that are adapted during training of the network. Parameters which are manually set are so called hyperparameters, such as layer numbers and structure, learning rate, batch size, etc.

3.4.1 Convolutional neural networks for semantic segmentation

CNNs were initially designed for classification of images and objects in images, but recent methods also focus on segmentation, such as those proposed by Lang et al. [121] and ResNet [122]. The main focus of current research is to reduce the degradation problem of deep architectures and to achieve high resolution pixel outputs. CNNs for semantic segmentation consist of an encoder network followed by a decoder network. The encoder is often based on a CNN for classification to extract semantic information from the image in low resolution discriminative features. On the other hand, the decoder reconstructs a high resolution pixel map as output of the network. Detailed spatial information is lost through the encoder, but deeper semantic information is gained. Several approaches have been proposed to pass detailed information past the encoder to get fine grained label maps, like Pyramid Scene Parsing Network (PSPNet), Full-Resolution Residual Networks or U-nets [109].

CNNs for semantic segmentation are supervised machine learning methods. Labelled training data serves as the basis for training, as for the method presented in Section 3.3. In contrast to our implementation of random forests, dense annotations are necessary for training the CNN.

We selected the proposed ERFNet [112] for our task of segmenting soil cover classes because it provides a compromise between speed and accuracy.

3.4.2 ERFNet

ERFNet follows an Encoder-Decoder architecture, also referred to as U-nets named after the study in [123]. The encoder reduces the spatial dimensions step by step by pooling and the decoder gradually recovers details and spatial resolution. Skip connections forward information directly from each encoder layer to the same level in the decoder. This spatial information in early layers, which are not fed directly through the encoder, help to restore details in the decoder. The information from the previous layer and from the skip connection are concatenated. This enables the network to extract abstract features and reduce computation while preserving fine localization. However, ERFNet does not use long-range skip connections as no empirical benefit was observed [112].

The main element of ERFNet are residual connections with factorized convolutions. Residual connections have limitations in terms of their efficiency and learning capacity. ERFNet introduces so called non-bottleneck-1D (non-bt-1D) layers which combine the benefits of bottleneck and non-bottleneck layers. They are utilized to split 2D filters into a combination of two 1D filters. Figure 3.8 shows these non-bottleneck-1D layers, where two 2D filters are split up into four 1D filters, as implemented in [116]. Non-bt-1D have fewer parameters and are faster than bottleneck layer while the accuracy is the

Table 3.1: Layers of implementation of ERFNet [112] in bonnet framework [116]. The number of output features and resolution are for input images of 512x384.

	Layer	Type	# output features	output resolution
Encoder	1	Downsampler block	8	256x192
	2-3	2 x Non-bt-1D	8	256x192
	4	Downsampler block	16	128x96
	5-8	4 x Non-bt-1D	16	128x96
	9	Downsampler block	64	64x48
	10-13	4 x Non-bt-1D	64	64x48
	14	Downsampler block	64	64x48
	15-18	4 x Non-bt-1D	64	64x48
Decoder	19	Deconvolution (upsampling)	32	128x96
	20-23	4 x Non-bt-1D	32	128x96
	24	Deconvolution (upsampling)	16	256x192
	25-28	4 x Non-bt-1D	16	256x192
	29	Deconvolution (upsampling)	8	512x384
	30-31	2 x Non-bt-1D	8	512x384

same as non-bottleneck layers. For up-sampling, ERFNet uses deconvolution layers with stride 2 to reduce memory and computation requirements.

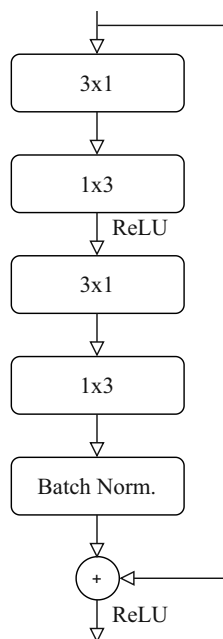


Figure 3.8: Non-bottleneck-1D layer as implemented in the bonnet framework [116].

Table 3.1 shows the layer structure of the ERFNet implemented in the bonnet framework [116].

The implementation uses leaky ReLU as ReLU function. Down-sampling is done by concatenation of max-pooling and 5x5 convolutions of stride 2, like the initial block shown in [124].

Image augmentation extends the training dataset by applying basic image manipulation techniques [125], especially geometric image transformations to improve the generalization of crop images [126]. Deep learning based approaches like Generative Adversarial Networks (GANs) have been frequently used lately. For the training, we applied image flip, gamma functions and gaussian blurring randomly on the image.

3.5 Evaluation

As a result of both methods, we receive a classification mask as well as the percentage of classes in the image which is the agronomically interesting value. Figure 3.9 and Figure 3.10 show example test images of different soil cover scenarios.

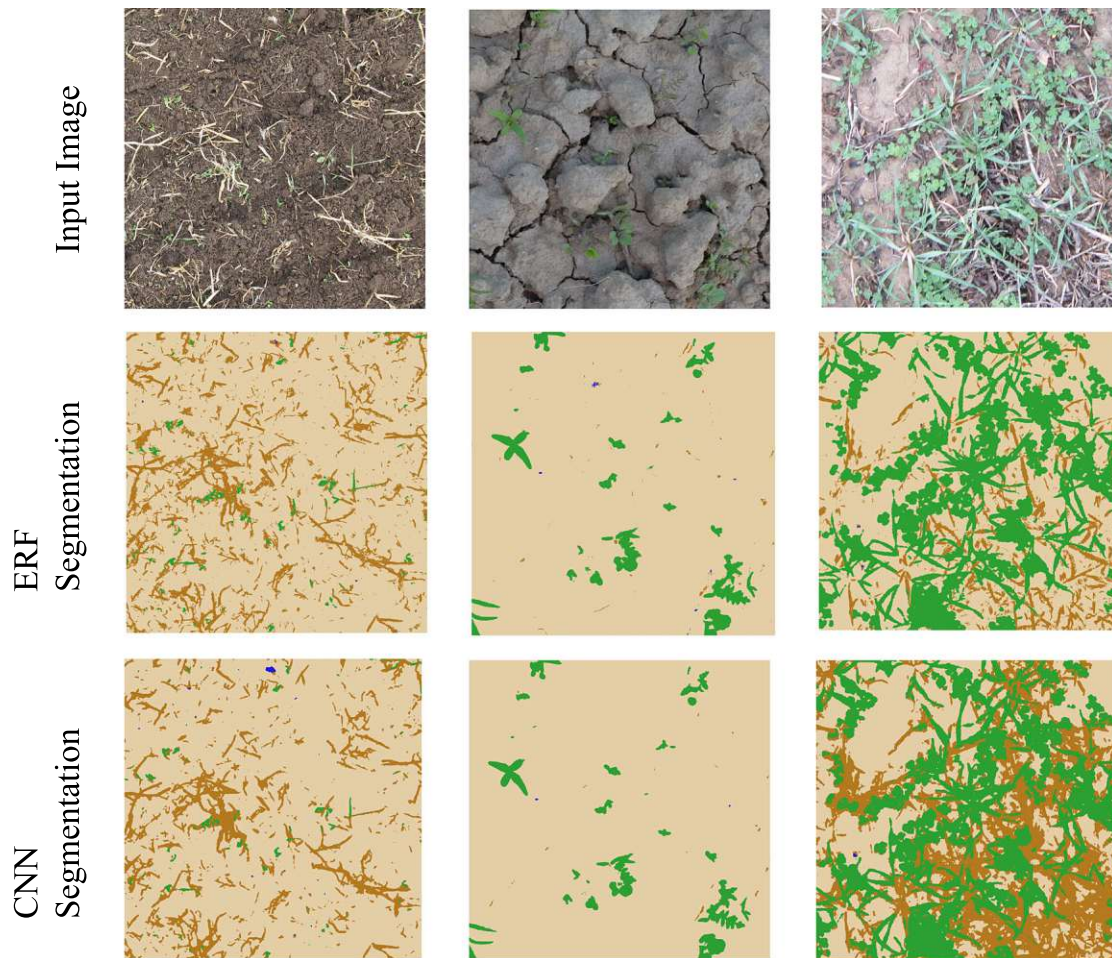


Figure 3.9: Examples of different soils surface covers (top) and the classification results (bottom). living plant ■, residue ■, soil ■, stone ■.

Currently, the evaluation of soil cover estimates can only be compared with manual annotation. The common grid method, as presented in [127], was used to reduce subjectivity. The classical grid method is used directly on the field. Using this method, a human classifies the soil cover at marked points. To select these points systematically, strings with knots at defined intervals are laid down onto the soil surface. The annotator writes down the soil cover class below each knot.

We used this method and adapted it for images, as presented in [16], to evaluate our methods.



Figure 3.10: Examples of different soils surface covers (top) and the classification results (bottom). living plant ■, residue ■, soil ■, stone ■.

3.5.1 Validation with the Grid Method

Soil cover is usually estimated using the grid method. A grid is laid out on the ground and the soil cover is manually classified at grid points [127]. Therefore, a pixel grid is used on an image of soil. A grid of 160 x 160 pixels is overlaid on a test image resulting in 432 crossing points for a 4000 x 3000 pixel image. Each of the points is

manually labelled. For validation, each of these points is also classified by both of the soil cover estimation methods presented. The following graphs show the percentage of soil cover from the manual classification and from the image analysis methods. Figure 3.11 shows the results (percentage of soil cover) of the Entangled Random Forest classification compared to the manual labelling from the grid method. Figure 3.12 shows the classification results from the ERF classification, classification with the ERFNet and the OBIA method proposed in [16].

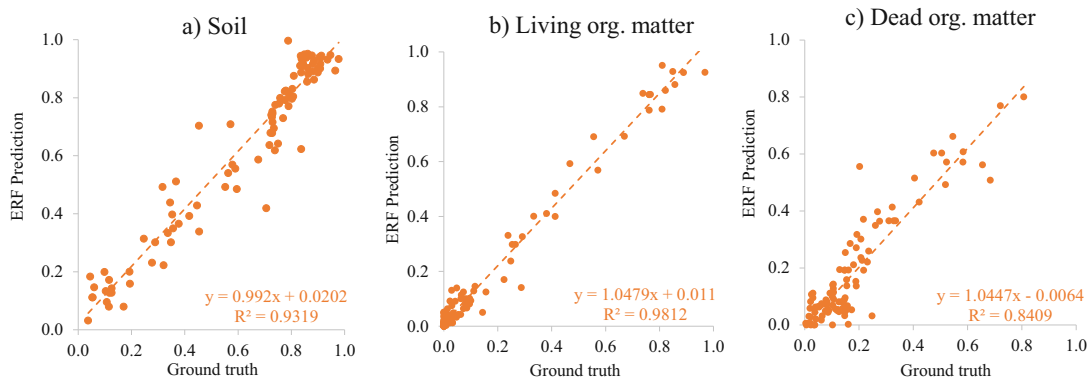


Figure 3.11: Comparison between reference annotations and entangled random forest results using the grid method (●) for (a) soil, (b) living organic matter and (c) dead organic matter ($N = 99$). The regression equation and the coefficient of determination are shown at the bottom of each graph in orange.

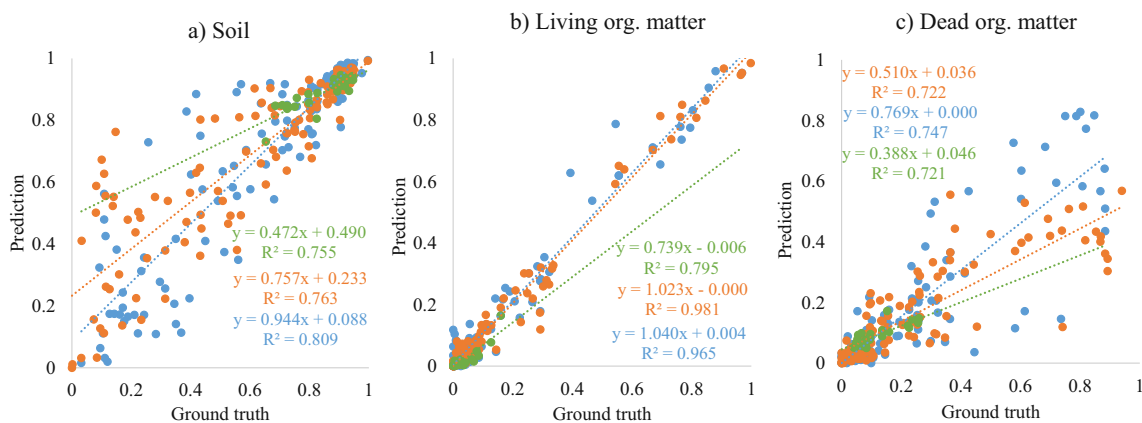


Figure 3.12: Comparison between reference annotations, entangled random forest results (●), ERFNet (●) and the method proposed in [16] (●), for (a) soil, (b) living organic matter and (c) dead organic matter. The regression equations and the coefficients of determination are shown in color for all three methods.

3.5.2 Validation including Human Error

Comparing the soil cover estimation methods with human labels involves a degree of uncertainty. The work in [127] observed strong variations between different evaluators. To quantify the variation between the different annotators, different individuals were asked to label 10 different soil cover images. Each person annotated the same 10 images resulting in 10 soil cover values for each class and image. Table 3.2 shows the standard deviation of percentages of soil cover within the 10 annotators.

Table 3.2: Standard deviation between soil cover estimates of 10 different people.

soil	living organic matter	dead organic matter	stone	mean value
5.5%	4.2%	3.5%	0.7%	3.47%

The variations of the entangled random forest method, as well as, the CNN based method are within the standard deviation of the test persons for living organic matter. The higher variance for the classes soil and dead organic matter can be observed at the human annotators as well as for both algorithms. Challenging images for both methods are images which show a larger amount of direct sunlight (not diffuse). Images with direct sunlight are under-represented in our data set. Direct sunlight produces small shadows around small soil particles. The algorithms recognize these shadows as coverage with residuals. Future research should focus on increasing the number of images with direct sunlight within the training dataset.

3.5.3 Inference Speeds

The inference speeds of the methods are important for real time application. We compared the inference speeds of the CNN ERFNet using different hardware. The results are shown in Table 3.3

Table 3.3: Inference speeds of ERFNet on different devices with an image resolution of 512x384 px.

Hardware	Inference time
UP AI Core X Myriad™ X 2485	268 ms
Intel® Core™ i7-3630QM CPU	190 ms
NVIDIA® Jetson Nano™	166 ms
NVIDIA® GeForce RTX 2080 Ti	6.1 ms

Depending on the application, different inference times must be achieved, especially machine control applications requiring real time inference speeds. We investigated the inference times of four different devices. It should be noted that the NVIDIA® GeForce RTX 2080 Ti was used to train the models and is not well suited for application on agricultural implements.

The NVIDIA® Jetson Nano™ is the most promising inference device for our applications, in terms of its power consumption and its inference speeds. Based on these tests, it was also integrated into the reference implementations of our camera systems.

3.6 Discussion

Soil cover is crucial for sustainable cultivation of arable land. It improves the infiltration and the retention of soil moisture, and increases humus formation. Different national and international subsidy programs focus on sustainable farming and encourage reaching certain soil cover levels on fields. However, the precise and robust measurement of soil cover is the prerequisite for implementing ideas and methods. Currently, no method is known that measures soil cover satisfactorily. Our proposed methods for soil cover measurement focuses on a generic approach to support different camera types while still getting robust results. Both proposed machine learning methods, the entangled random forest, as well as, the method based on the ERFNet CNN are able to distinguish between soil cover types. Both methods are developed with a manageable amount of training data. The method based on CNN uses pre-training on a related dataset to reduce the number of training images needed. However, special cases present a challenge to the methods, especially during strong sunlight, because small soil aggregates cast shadows which can be confused with dead organic matter. One way to solve this problem would be to increase the number of training samples in this corner cases. Additionally, expanding the dataset with new scenarios (e.g. different lighting and soil colors) could further improve the robustness of the system. Using additional data from the near infra-red spectrum could also reduce this problem. The new camera system, for the grassland use case, shown in Section 5.2.2 could be used for this.

The methods were tested under field conditions and validated against manual classifications. Initial tests with low altitude flying unmanned aerial vehicles (UAVs) showed similar results to the ground based images. However, future studies could focus on investigating the limits of the methods when used on UAVs in higher altitudes.

The entangled random forest based method can be accessed at no cost via a web-service at <https://soilcover.josephinum.at>.

4 Soil Roughness Measurement

Soil Roughness is an important factor for managing soils and plants because it influences soil erosion and plant emergence. During seeding and tillage, the goal is to break apart soil so that it is fine enough for strong plant emergence, but not too fine as to protect against capping and erosion. The type of tillage implement, the tillage intensity, the working depth, the soil type and the soil water content during the tillage process can all influence the aggregate size distribution in the seedbed (see [7] and [68]). Soil conditions might change in spatially and temporally significant ways in a field [70]. New methods for measuring soil roughness in motion while controlling a tillage machine are needed in order to be able to control the soil roughness and to get a more uniform roughness at changing soil conditions.

Such methods relieve the driver, especially those without agronomic knowledge, of the strenuous task of changing the machine parameters manually during tillage. Using headlights on the camera system enables farmers to operate tractors at night which is very challenging during regular manual operation. The ongoing trend of automation is leading to soil roughness control serving as the basis for seeding with robots. Additionally, such a system can be used for documentation of the soil roughness, for example mapping for subsequent precision farming applications. A roughness measurement system also provides metrics to evaluate the work quality of different tillage machines.

The measurement system is integrated into an ISOBUS Class III application to control a tractor's PTO speed and driving speed in order to vary tillage intensity.

Parts of this chapter were published in Soil and Tillage Research [17] and presented at the ICVS conference 2019 [15], at the EurAgEng Conference 2018 [18] and at the VDI Land.Technik conference 2017 [19].

4.1 Hardware Setup

Depth information is needed in order to measure soil roughness and there is currently no method for measuring the aggregate size from a single image in motion. The difficulty lies in the rich random texture which makes it difficult to recognize individual aggregates. Laser based systems are quite expensive and not as mechanically robust as cameras. Due to the rich texture of soil and plants, passive stereo camera systems are a good choice to perform this method. Stereo matching is computationally intensive and time consuming if calculated on a CPU. Therefore, the computation hardware requirements are quite high.

4.1.1 Hardware Requirements

The measurement system will be mounted on a tillage machine and therefore has a short ground clearance of about 25 cm to 50 cm. Based on findings in [128], differences in height of about 3 mm should be detectable. The depth resolution can be calculated by Equation 4.1. The observed area should be as big as possible, which requires a short focal length, to capture a representative area. This method should enable additional night-time operation.

The depth resolution Δz of a stereo camera setup can be calculated by

$$\Delta z = \frac{z^2}{f \cdot b} \Delta d \quad (4.1)$$

f ... focal length

b ... the baseline

d ... disparity

4.1.2 Test Setup

For the roughness measurements two different cameras are used: a LI-OV580-STEREO RGB stereo camera and a LI-USB30-V024STEREO stereo camera, which are both from Leopard Imaging Inc. The RGB camera LI-OV580-STEREO has two OmniVision color CMOS 4M image sensors with a resolution of 2688 x 1520 and an optical format of 1/3". The lenses have a focal length f of 2.8 mm. The baseline b of the stereo camera is 80 mm. In order to achieve real-time calculation with the color camera, a reduced resolution of 640x480 are used for stereo processing. The grayscale camera, LI-USB30-V024STEREO, uses Aptina MT9V024 Global Shutter WVGA sensors with a resolution of 640 x 480. The lenses have a focal length f of 2.35 mm. The baseline b of the stereo camera is 50 mm.

The computations were done on an ODROID XU4 single board computer with Samsung Exynos5422 Cortex™-A15 2Ghz and Cortex™-A7 Octa core CPUs with 2Gbyte LPDDR3 RAM.



Figure 4.1: LI-OV580-STEREO (left), LI-USB30-V024STEREO (right)

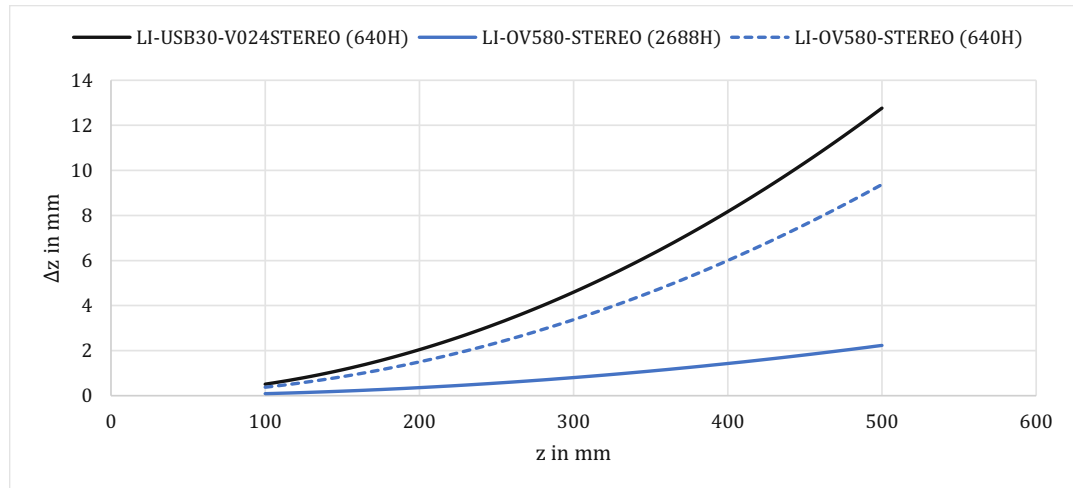


Figure 4.2: Depth resolution Δz depending on z for LI-USB30-V024STEREO and LI-OV580-STEREO cameras with full resolution of 2688 pixel and reduced resolution of 640 pixel (assuming no sub-pixel accuracy).

Using Equation 4.1 ($z = 270$ mm, assuming no sub-pixel accuracy), a Δz of 0.65 mm was estimated for the LI-OV580-STEREO camera ($2 \mu\text{m}$ pixel size) and a Δz of 3.72 mm for the LI-USB30-V024STEREO camera ($6 \mu\text{m}$ pixel size). For additional empirical validation of the height measurement, ten reference heights from a plane surface were measured using a height gauge. Ten heights between 140 and 270 mm were selected. A mean error, between height gauge readings and camera depth measurements, of 1.55 mm was estimated.

4.2 Roughness Estimation

The goal of the roughness estimation process is to calculate the roughness indices of a surface. After initially calibrating the stereo camera, the processing is done in a roughness estimation pipeline. The rich texture of soil surfaces allow for a simple and efficient block matching algorithm. External influences on the roughness estimation are taken into account during processing. Details of the stereo processing can be found in [129].

4.2.1 Camera Calibration

The stereo camera must first be calibrated in order to perform stereo matching. This is done once initially for the stereo camera. Afterwards, the parameters of the camera set up should remain the same. Therefore, there are special requirements for the mechanical connection of the two cameras in order to avoid changes in the position and the orientation of the cameras over time.

The camera calibration is done by the calibration method implemented in OpenCV [130]. First, the intrinsic parameters of each of the two cameras are calibrated and afterwards the extrinsic stereo parameters are also calibrated. Due to the strong positive

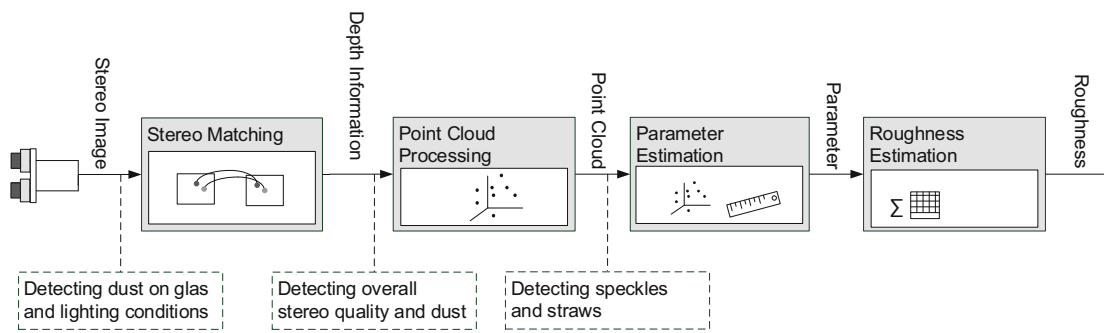


Figure 4.3: Estimation process from the camera images to the final roughness value.

distortion of the lenses, the radial distortion parameters were only calibrated up to the 3rd coefficient to receive converging results.

4.2.2 Roughness Estimation Pipeline

The estimation process consists of several consecutive steps. The necessary steps performed by the roughness estimation pipeline are shown in Figure 4.3. The individual steps are explained in detail below.

First, the lighting conditions are evaluated to adapt exposure time, further details on how to handle the lighting conditions are described in Section 4.2.4. After image rectification based on the parameters of the camera calibration, the stereo matching is performed by the block matching algorithm implemented in OpenCV [130]. The block matching delivers a disparity map as output. The density of the disparity map $q = [0,1]$ is used as a quality parameter. Missing parts in the disparity map are interpolated by a multi-scale pyramid based approach. Each interpolated pixel is filled by the local average corresponding to a lower level of the pyramid. This efficient interpolation process is necessary to get a dense representation in order to calculate the surface normals.

Using the camera parameters, it is possible to generate a 3-D point cloud based on the results of the disparity map. Speckles and residues are removed during the point cloud processing step. Parameters of the point cloud like height measurements and surface normals are used to calculate the roughness indices. Finally, a weighted mean combines several roughness measurements to smooth the roughness values.

4.2.3 Roughness Indices

Several roughness indices have been proposed over the years. Studies in [79] have shown that the RC index and the average angle of normals (AoN) are suitable for describing soil after tillage. Figure 4.4 illustrates the underlying concepts of these two indices. The *RC* index is based on the standard deviation of height measurements:

$$RC = \sqrt{\langle (z(x,y) - \langle z(x,y) \rangle_{X,Y})^2 \rangle_{X,Y}} \quad (4.2)$$

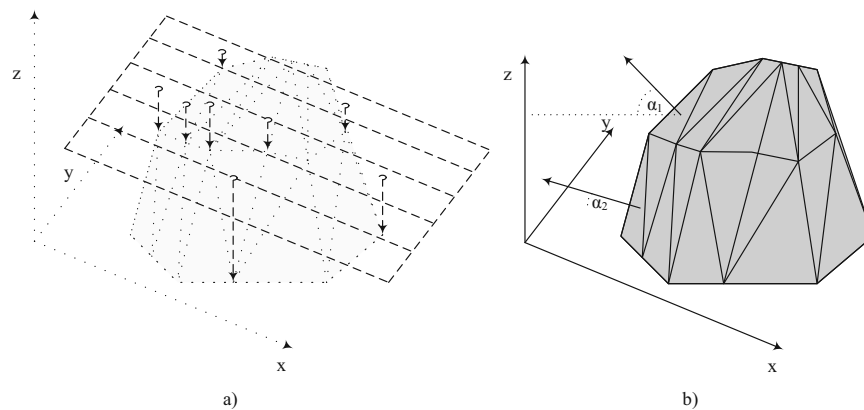


Figure 4.4: Height measurements for calculation of the RC indices, b) angles of the surface normals for calculation of the average angle of normals (AoN).

In order to accommodate directed roughness patterns like furrows, directed roughness indices were introduced in [80]. The index RC_X , see Equation 4.3, and RC_Y represent the roughness in the X and Y directions separately which makes it possible to exclude directed patterns from the measurement. A perfectly flat surface would have a RC index of 0 mm. For a drawn tillage machine, this results in a higher roughness transversal to the driving direction RC_Y compared to the roughness along the driving direction RC_X .

$$RC_X = \langle \sqrt{\langle (z(x,y) - \langle z(x,y) \rangle_X)^2 \rangle_X} \rangle_Y \quad (4.3)$$

The AoN index is based on the surface normals of the point cloud. The average surface normal of the overall surface is calculated as presented in the following equation:

$$AoN = \frac{1}{N} \sum_{i=1}^N \arccos \left(\frac{S_{i,Z}}{\|S_i\|} \right) \quad (4.4)$$

A perfectly flat surface would have an AoN of 0° .

4.2.4 Handling of External Influences

Two important parameters for cameras in motion are the exposure time and the resulting motion blur. To avoid motion blurring at faster speeds, the exposure time has to be controlled. Many modern agricultural implements are connected to the tractor via ISOBUS. The ISOBUS provides the tractor's drive and operating data to the implement, including the driving speed $v_{machine}$. Equation 4.5 shows the the maximum possible exposure time $t_{s,max}$ for a given $v_{machine}$.

$t_{s,max}$ is influenced by the vertical spatial image resolution of $res_{vertical} = 480$ and $s_{vertical} = 0.5$ m. The tolerance factor $\eta = [1,n]$ in Equation 4.5 gives the overlap of pixels exposed to the same area. In order to avoid motion blur, η should be kept low at

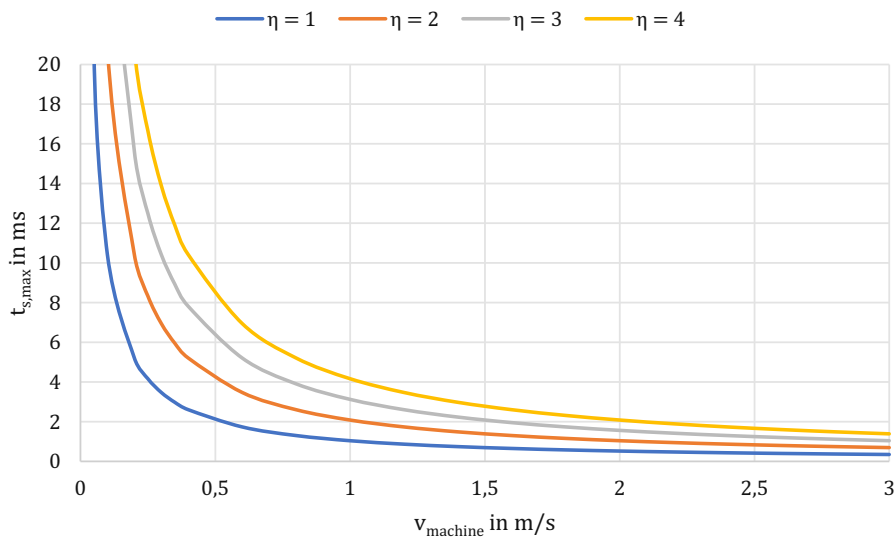


Figure 4.5: Maximum exposure time $t_{s,max}$ for different driving speeds $v_{machine}$ and overlapping factors η .

a value of 1. Figure 4.5 shows the maximum exposure time $t_{s,max}$ at increasing driving speeds $v_{machine}$ for different overlapping factors η .

$$t_{s,max} = \frac{\frac{s_{vertical}}{res_{vertical}}}{v_{machine}} \cdot \eta \quad (4.5)$$

The brightness of the image, its mean intensity, is the basis for calculating the actual exposure time t_s under the limiting $t_{s,max}$. Additionally, the peak value of the image histogram is considered to avoid over- and under-exposure.

The exposure time t_s is selected to keep both values brightness and peak value of the image histogram within a certain range. In order to increase the range of lighting, additional LED- headlights are mounted next to the camera to illuminate the soil. The headlights allow for light independent operation even during nighttime. In general, less motion blur should be favored over brighter images in order to detect enough image correspondences by the stereo matching algorithm.

In order to detect deposits on the camera glass, the left and right camera image are compared. Deposits on the camera are usually asymmetric between the left and right images. Both images are taken with the same parameters, so strong differences in the images, estimated by the mean gray value, are attributed to deposits on the glass. Furthermore, the density of the disparity map q also decreases by the covered image area. Therefore, a distinction must be made between deposits on the glass and general matching problems (e.g., by dust in the atmosphere etc.). A low q value indicates dust in the atmosphere between the camera and the soil surface if other causes, such as residues, exposure or deposits on the glass can be excluded.

One major influence on the roughness measurement are residues like straws on the soil and, in general, soil cover. In order to avoid the influence of residues at the roughness

measurement, we exclude pixels containing soil cover from the roughness estimation. These pixels are masked as invalid and excluded from the calculation for the RC and AoN indices. As presented in Chapter 3, soil cover can be detected from RGB images. When using the RGB stereo camera, we can use the labels *living organic matter* (plants), *dead organic matter* (residues) and *stone* to mask invalid points on the disparity map.

If a gray-scale stereo camera is used, an approach based on geometric filtering is used. In this process, the 3-D point cloud is filtered to detect unexpected spikes which differ in their geometric properties from soil aggregates. The geometric properties are defined by height in proportion to their cross-sections. The simple metric for the geometric filtering can be implemented very efficiently and is favored for real-time application.

A status message is sent to the user if any of the influences mentioned above exceeds a certain threshold.

4.3 Online Machine Control

The goal of controlling tillage machines, in this case power harrows, is to achieve a desired roughness over the entire field. The operator selects the desired roughness R_{ref} based on his agronomic knowledge in order to till the soil at a roughness which is sufficient for good field emergence, yet not too fine as to protect against capping and erosion. The control algorithm tries to achieve a uniform roughness of R_{ref} at changing soil conditions over the entire field.

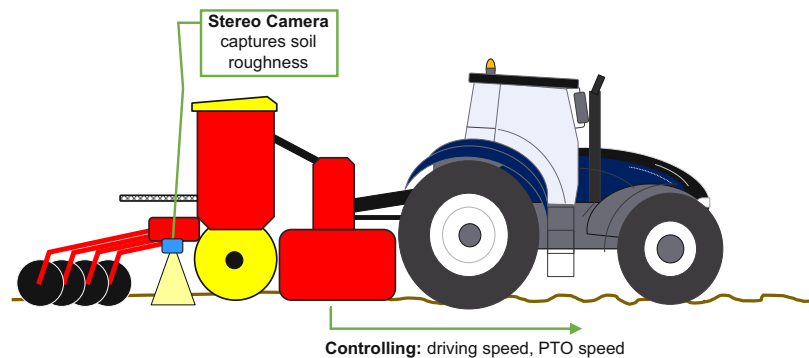


Figure 4.6: Concept of the tillage machine control.

The operator inputs the desired roughness R_{ref} into the system. During tillage, the camera system continuously estimates the current roughness R_{act} and adapts the tractor driving speed v and the PTO speed n_{PTO} accordingly. Figure 4.6 shows a concept of the overall control system. In Figure 4.7, the control loop is shown.

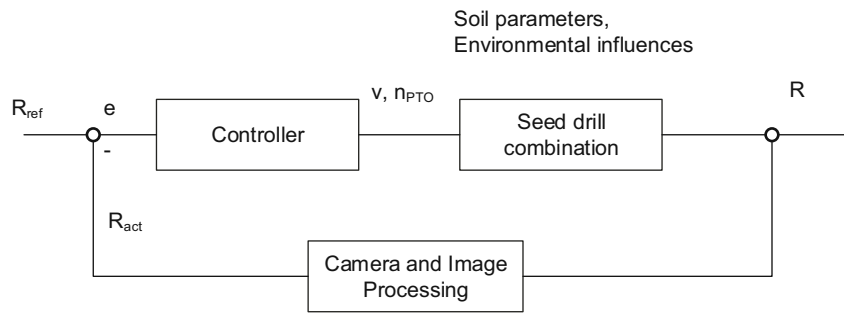


Figure 4.7: Closed loop controller.

The control algorithm is implemented in the ECU of the power harrow. The camera systems sends the current roughness R_{act} and status information via CAN bus to the ECU. The ECU calculates the desired speeds (v and n_{PTO}) and sends them to the tractor via ISOBUS Class III. Using ISOBUS Class III, it is now possible for the first time to control the tractor with an implement. An additional hysteresis avoids speed changes for only small soil roughness changes.

In order to get a functioning system a headland management system is implemented. After lowering the linkage of the tractor on the headland, the roughness estimation is stopped for 1 meter to measure only the tilled soil.

Figure 4.8 shows the final camera set up with additional headlights mounted on the power harrow.

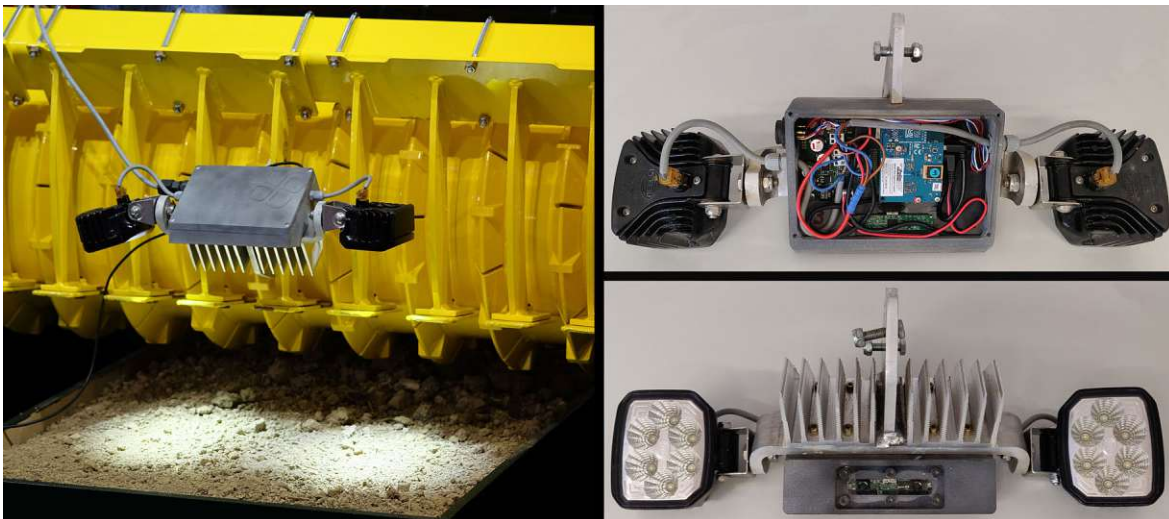


Figure 4.8: Camera mounted on the power harrow (left). Detailed view of stereo camera with Odroid XU4 and headlights (right)

4.4 Field Trials

The field trials of the roughness measurement camera system took place during the 2016 and 2017 farming seasons. The goal of the trials was to estimate the relation between the mean clod diameter of the soil and the stereo vision roughness. An additional goal was to evaluate the working intensity of the power harrow and the controller effect. External influences were also evaluated in order to determine if they had any practicability effects.

The main field trials were carried out in three experiments in Lower Austria with different soil conditions (silty loam, loamy silt and clay loam). Table 4.1 shows an overview of the experiments.

4.4.1 Soil sampling and sieve analysis

We validated the camera system by comparing the estimated roughness to the mean-weight diameter of the soil aggregates. Estimating the mean aggregate size was based on soil sampling and a subsequent sieve analysis.

The soil samples were taken to the same location where the camera images were taken before. Before taking soil samples, stereo images were captured from the corresponding location.

The soil samples were carefully taken with a spade from an area of 25 x 25 cm from the seedbed horizon (0-6 cm +/- 1 cm) and air-dried afterwards. This was done by the people from the Versuchswirtschaft Gross Enzersdorf (VWG) of BOKU.

A vibratory sieve shaker (Retsch AS 200, Haan, Germany) was used for the sieve analysis. The mesh sizes of the different sieves were 40 mm, 20 mm, 10 mm, 5 mm, 2.5 mm, 1.25 mm and < 1.25 mm. The amplitude of the shaker was set to 1.3 mm with an acceleration of gravity of 18.75 g. The different fractions were weighted with a precision scale with an accuracy of 0.1 grams.

The mean-weight diameter (MWD) is the standard method for comparing seedbed and soil aggregate composition. The MWD is calculated with the Equation 4.6 by weighting each diameter by its gravimetric proportion.

$$MWD = \frac{\sum_i n_i \cdot d_i}{\sum n_i} \quad (4.6)$$

MWD ... mean-weight diameter (mm)

n_i ... weight of the sieving fraction in the mesh size *i* (g)

d_i ... mean of the mesh size *i* (mm)

4.4.2 Experimental Design

Three main field experiments were carried out to test the roughness measurements and the controller system. The first experiment focused on getting different seedbed aggregate compositions. This was done by using three different primary tillage implements with different working intensities. Different driving speeds during seeding resulted in an additional gradation of the aggregate sizes. A seeding combination based on a seed

drill (Pöttinger Vitasem 302 ADD) and a power harrow (Pöttinger Lion 303.12 Classic, Grieskirchen, Austria) were used for all the experiments. During the first and second experiments, the soil roughness measured by the camera was compared to the results of the sieve analysis.

Experiment 1 was designed as a randomized block trial with three repetitions for each variant. The plots had a length of 140 m and a width of 6 m (two working widths). During seeding, the PTO speed of the power harrow was set to a constant 540 1 min^{-1} . The variants differed in their primary mode of tillage operation and the driving speeds during seeding. Details are shown in Table 4.1.

In contrast to experiment 1, the PTO speed was varied for experiment 2. As in experiment 1, different primary tillage machines were used in experiment 2. It was expected that due to the varying PTO speed and primary mode of tillage the range of soil aggregate sizes would increase. The randomized block trial in experiment 2 consisted of three repetitions for each variant. Each plot had a length of 200 m and a width of 6 m. The PTO speed was changed between 300, 540 and 1000 1 min^{-1} during seeding. The moisture conditions at experiment 2 during November were nearly at field capacity.

The controller was tested during the third experiment. The soil roughness estimation from the stereo camera system served as the basis to control the PTO speed of the tractor. The tractor was a New Holland T7.270 (United Kingdom) which is capable of executing ISOBUS Class III commands from the implement and in our case the ECU of the power harrow. Two variants were compared in experiment 3: Seedbed preparation with controlled PTO speed based on estimated soil roughness and seedbed preparation with constant PTO speed (550 1 min^{-1} uncontrolled). During the experiment, the controller tries to generate a requested soil roughness, in this case $RC = 14$, by varying the PTO speed between 550 and 1000 1 min^{-1} . Primary tillage was done with a mouldboard plough in experiment 3. The two variants were alternated side by side in tracks of about 400 m. The controlled and constant variants were placed adjacent to each other and can be compared based on the assumption that they have very similar soil locally. The effect of the controlled variant is measured by calculating the mean squared error between the requested soil roughness R_{Ref} and the actual measured soil roughness R_{act} at n points along all tracks (see Equation 4.7).

$$MSE = \frac{1}{n} \sum (R_{ref} - R_{act})^2 \quad (4.7)$$

Table 4.1: Field Experiments design

	Experiment 1	Experiment 2	Experiment 3
Primary tillage	Mouldboard plough Cultivator shallow Cultivator deep	Mouldboard plough Cultivator Short disc harrow	Mouldboard plough
Seedbed preparation	Power harrow PTO: 540 (285) rpm	Power harrow PTO: 300 (158) rpm, 540 (285) rpm, 1000 (527) rpm	Power harrow PTO: between 550 (290) rpm and 1000 (527) rpm Field-test of the control-approach
Forward driving speed	1.1 $m s^{-1}$, 1.7 $m s^{-1}$, 2.2 $m s^{-1}$	1.4 $m s^{-1}$	1.3 $m s^{-1}$

The characteristics of the soils at all three trial sites are shown in Table 4.2.

Table 4.2: Characteristics of the soils on the trial sites.

	Experiment 1	Experiment 2	Experiment 3
	Gross Enzersdorf (48°12'1.79"N, 16°33'33.75"E)	Krummnussbaum (48°12'4.22"N, 15°9'31.29"E)	St. Leonhard am Forst (48°7'32.95"N, 15°17'29.37"E)
Sand ($g kg^{-1}$)	36	10	25
Silt ($g kg^{-1}$)	44	70	40
Clay ($g kg^{-1}$)	20	20	35
Organic matter ($g kg^{-1}$)	13.3	12.8	12.0
Volumetric water content ($m^3 m^{-3}$)			
0-5 cm	0.24	0.34	0.37
5-10 cm	0.22	0.38	0.41
Soil type (FAO, 2006)	Calcaric Chernozem	Cambisol	Stagnosol

The moisture data was estimated using meteorological and soil property data with the method proposed in [131].

Among others, the work in [62] and [132] investigated the relation between soil aggregate sizes and plant emergence. In general, finer soil aggregates lead to better plant emergence. We have conducted an additional experiment to assess this idea. During the seeding of winter wheat, the roughness was recorded on a field in Wieselburg (48°8'36.19"N, 15°8'38.75"E). The soil characteristics of this experiment are shown in Table 4.3. When plant emergence started, the method for soil cover estimation based on the entangled random forest presented in Chapter 3 was used to develop geo-referenced mapping of the living plant cover. A GoPro HERO5 Black camera was used to capture images for the soil cover estimation.

Table 4.3: Characteristics of the soils from the additional experiment to observe plant emergence.

	Sand	Silt	Clay
Minimum	7.1	52.6	13.5
Maximum	15.4	78.2	32
Mean	10.39	68.97	20.65
Median	9.6	69.3	20.6

4.5 Evaluation

In order to be able to evaluate the system, different parameters were recorded and compared. First, the aggregate size distribution in the seedbed was measured with the sieve analysis to get insights into the soil aggregate composition and the effects of tillage. The roughness measurement system was evaluated by investigating the relation between MWD and the roughness estimated with the stereo vision system. To evaluate the overall system, the effect of the controller was shown. External influences on the roughness measurement system were investigated. A first attempt to demonstrate the relation between soil roughness and field emergence was made by comparing the roughness measurement results with the soil cover measurements from Chapter 3.

4.5.1 Aggregate Size Distribution in the Seedbed

Besides the mean weighted diameter, the fractions obtained from the sieve analysis give a more detailed understanding of the aggregate size distribution, as shown in Tables 4.4 and 4.5. The results show that the primary tillage influences the aggregate size even after seeding (see Table 4.4). Tilling with a mouldboard plough resulted in a coarser seedbed compared to tilling with a cultivator after the seedbed was prepared with a power harrow.

The variant using a plough for primary tilling resulted in 34.4 % of aggregate sizes > 20 mm after the seedbed preparation. The weight of the fraction > 20 mm was statistically higher for the ploughing rather than the cultivator treatment. The fraction of aggregates < 2.5 mm was 24.2 % after ploughing and statistically significantly lower than for the variants using a cultivator during primary tilling. In experiment 2, the disc harrow resulted in a finer seedbed compared to the mouldboard plough and cultivator treatments. Table 4.5 shows the effect of the PTO speed which resulted in finer seedbeds with higher PTO speeds. This difference is clearly visible in the MWDs in the last row of Table 4.5.

In order to obtain a better subjective assessment of the aggregate sizes and their corresponding roughness measurements, Figure 4.9 shows sample images with RC and AoN values in ascending order.

Table 4.4: Soil aggregate fractions (%) and mean-weight diameter in the seedbed horizon, different tillage operations (n=36), experiment 1.

	Plough	Cultivator deep	Cultivator shallow
>40 mm	18.3 ^c	10.3 ^b	4.9 ^a
20 - 40 mm	16.1 ^b	12.3 ^a	12.6 ^a
10 - 20 mm	15.9 ^a	15.9 ^a	16.3 ^a
5 - 10 mm	14.4 ^a	15.4 ^a	15.8 ^a
2.5 - 5 mm	11.3 ^a	13.5 ^b	15.0 ^c
1.25 - 2.5 mm	9.1 ^a	11.9 ^b	12.3 ^b
<1.25 mm	15.1 ^a	20.6 ^b	22.1 ^b
MWD (mm)	19.9^c	14.3^b	11.3^a

Table 4.5: Soil aggregate fractions (%) and mean-weight diameter in the seedbed horizon, different primary tillage operations (plough, cultivator and disc harrow) and PTO speed of the power harrow (300, 540, 1000 1 min^{-1}), n=3, experiment 2.

	Plough			Cultivator			Short disc harrow		
	300.0	540.0	1000.0	300.0	540.0	1000.0	300.0	540.0	1000.0
>40mm	57.8 ^b	23.2 ^a	19.4 ^a	37.6 ^b	23.3 ^b	5.9 ^a	19.8	21.3	8.8
20 - 40 mm	13.6 ^a	27.7 ^b	23.7 ^b	29.3 ^a	29.7 ^a	36.5 ^a	26.0	27.9	26.4
10 - 20 mm	13.0 ^a	22.0 ^b	21.3 ^b	17.5 ^a	23.3 ^b	27.4 ^b	22.2	21.3	24.1
5 - 10 mm	7.3 ^a	12.9 ^b	15.6 ^b	8.6 ^a	12.4 ^{ab}	16.3 ^b	15.1	14.2	17.3
2.5 - 5 mm	4.2 ^a	7.3 ^b	9.9 ^b	3.7 ^a	6.1 ^b	7.8 ^b	8.4	8.1	10.8
1.25 - 2.5 mm	2.2 ^a	3.7 ^{ab}	5.5 ^b	1.5 ^a	2.6 ^{ab}	3.2 ^b	4.5 ^a	3.9 ^a	6.7 ^b
<1.25 mm	2.0 ^a	3.1 ^{ab}	4.5 ^b	1.9 ^a	2.6 ^a	2.9 ^a	3.9 ^a	3.4 ^a	6.0 ^b
MWD (mm)	41.5^b	26.9^a	23.6^a	34.8^c	27.6^b	20.2^a	24.6	25.8	18.7

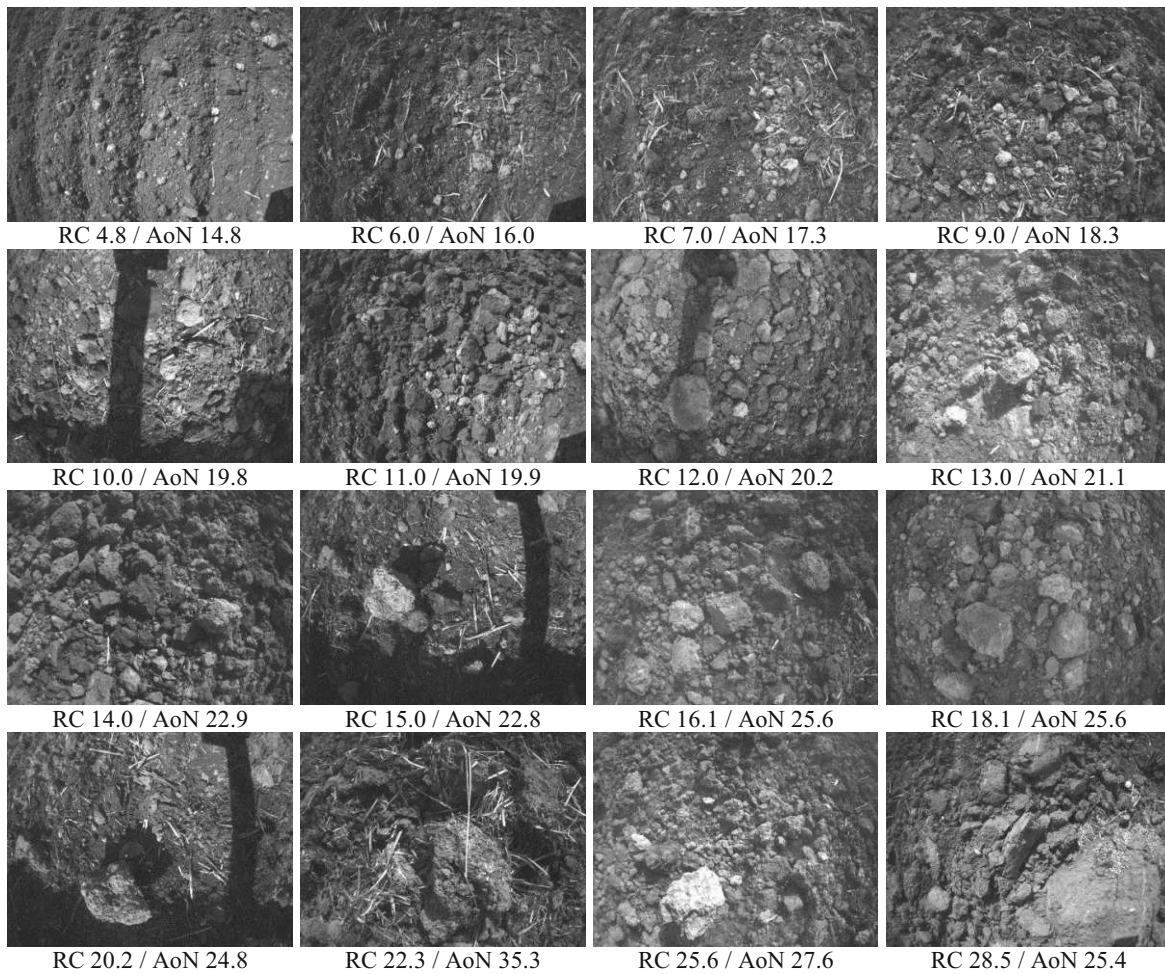


Figure 4.9: Sample images with corresponding RC and AoN values in ascending order.

4.5.2 Relations between Mean Clod Diameter and Stereo Vision Roughness

The correlation between the mean-weight diameter and the estimated roughness was calculated in order to validate the roughness measurement system. Figure 4.10 emphasises the difference between the soil roughness and the soil aggregate size estimation (MWD). The roughness only considers the soil surface, whereas the MWD is calculated from the soil volume leading to a certain discrepancy between the two measurements.

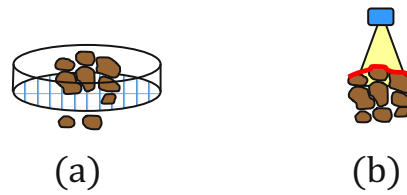


Figure 4.10: (a) Estimation of the mean weighted diameter of the soil aggregates. (b) Estimation of the soil surface roughness.

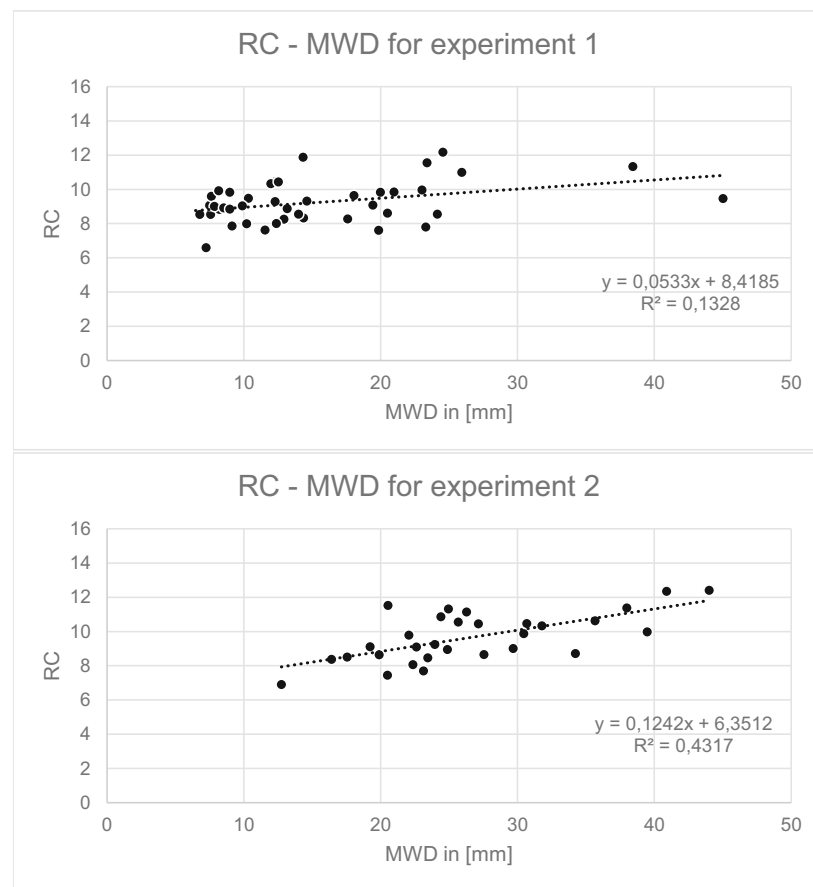


Figure 4.11: Measured results of MWD (x-axis) and RC roughness from the stereo camera (y-axis) for experiment 1 (top) and experiment 2 (bottom).

The Pearson correlation between RC and MWD was 0.55 in experiment 1 and 0.66 in experiment 2. The Pearson correlation between AoN and MWD was 0.455 in experiment 1 and 0.449 in experiment 2. The RC index resulted in a better correlation to the MWD than the AoN index, therefore the RC index was used to develop the controller. 52

samples were used for experiment 1 and 30 samples were used for experiment 2. The relationship between RC and MWD for experiment 1 was $RC = 0.14 \cdot MWD + 7.90$ with $R^2 = 0.30$. Experiment 2 resulted in a relationship of $RC = 0.12 \cdot MWD + 6.35$ with $R^2 = 0.43$. Figure 4.11 shows the relationship between the measurements for the MWD and the soil roughness estimation for all the sieve-analysis sample points.

4.5.3 Working Intensity and Controller Effect

The working intensity of a power harrow is influenced by the working depth, the tool geometry, the driving speed and the PTO speed. The working depth should be kept constant depending on soil and seeding requirements. The driving speed and the PTO speed can be influenced during the preparation of the seedbed. In order to further examine the results of experiment 2, the relationship between the working intensity, manifested by tine loops per meter, and the roughness RC is investigated.

The tine loops per meter l_{tine} in (1/s) are defined by the following equations. The power harrow, using the same configuration from the experiments, translates the rotations of the PTO n_{PTO} to tine rotations n_{tine} by a factor of 1.89474. In order to better understand this concept, the tine loops per meter are depicted in Figure 4.12 on the right.

$$l_{tine} = \frac{n_{tine}}{60 \cdot v_{machine}} \quad (4.8)$$

$$n_{tine} = n_{PTO} / 1.89474 \quad (4.9)$$

$v_{machine}$... Driving speed of the power harrow in m/s

n_{PTO} ... PTO rotation speed in RPM

Figure 4.12 shows the RC values for different tine loop per meter values l_{tine} from experiment 2. The driving speed $v_{machine}$ was recorded with a GNSS receiver. A higher number of tine loops per meter increases the work intensity which produces a finer seedbed (Roughness RC) as observed by the stereo camera system. The driving speed was targeted at 1.4 m s^{-1} and varied from 1.1 m s^{-1} to 2.8 m s^{-1} . The PTO speed n_{PTO} was intentionally varied between 300 rpm and 1000 rpm, as depicted in Table 4.1, resulting in 1 to 7 tine loops per meter.

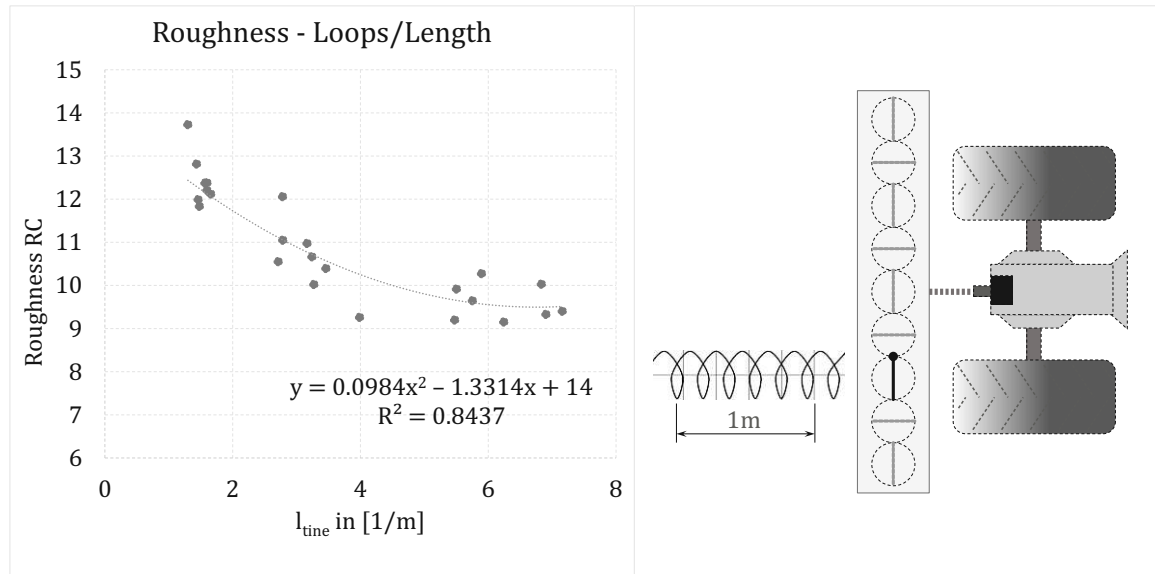


Figure 4.12: Roughness RC results at different working intensities (tine loops per m). Each point equals the mean value of one track. $N = 24$.

The third experiment focused on quantifying the controller effect. The controller changes the tillage intensity by changing the PTO speed n_{PTO} to achieve a desired soil roughness R_{ref} on the entire field. Initial tests on the field in experiment 3 showed that a roughness value of $RC = 14$ would be appropriate for the conditions and the seeds. It was also observed that the roughness varied greatly across the field. As mentioned in Section 4.4.2, the mean squared error between the desired roughness R_{ref} and the actual roughness R_{act} is appropriate to describe the homogeneity of the roughness on the field and is a good variable to validate the controller. The tracks with constant PTO speeds of $n_{PTO} = 550$ resulted in an error of $MSE = 2.16$ deviating from desired roughness of $RC = 14$. On the other hand, the tracks with PTO controller resulted in an lower error of $MSE = 1.24$. This clearly indicates that the controller better achieves the targeted value of $RC = 14$.

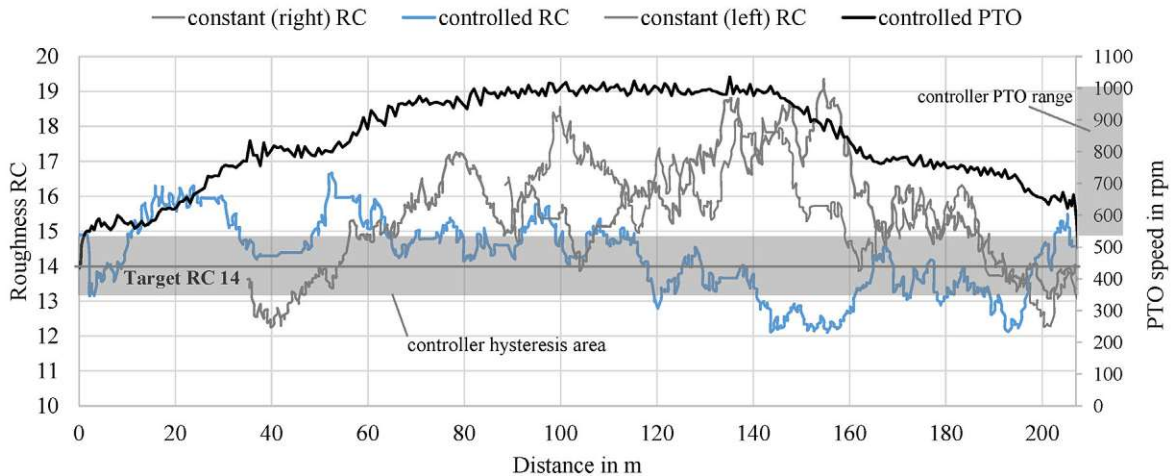


Figure 4.13: Controller effect: The diagram shows three tracks side by side, with the controlled track in the middle (blue line), and two tracks with constant PTO speeds (gray line). The PTO speed of the controlled track is shown by a black line. The target value for the controller was $RC = 14$.

In order to investigate the controller in more detail, Figure 4.13 shows three example tracks of experiment 3. In this figure, you can see the RC values for a controlled track in blue and the neighboring tracks with a constant PTO speed in gray. The black line shows the PTO speed n_{PTO} of the controlled track and its corresponding y-axis is on the right side of the figure. The PTO speed n_{PTO} was recorded from the ISOBUS. The hysteresis area (between $RC = 13.2$ and 14.8) is shown as the rectangle with grey shading around the target RC value $R_{ref} = 14$. The control range of n_{PTO} is depicted by the rectangle with grey shading on the right y-axis. In general, the field is rougher at the beginning and reaches its maximum roughness at about 150 m along the track. After the maximum point of roughness, the roughness decreases rapidly as finer soils are reached. This is especially visible in the tracks with a constant PTO speed where the power harrow decreases the soil roughness by a constant value and the soil heterogeneities are clearly visible. The controller starts by gradually increasing the PTO speed to the maximum speed of 1000 rpm at 80 m within the track. At the point where the soil gets finer (about 150 m), the PTO speed decreases again. The changes to the PTO speed are always visible when the RC value of the controlled track R_{act} leaves the hysteresis area.

In general, the RC value for the controlled track moves closer to around $R_{ref} = 14$, which also manifests itself in the lower MSE for the controlled tracks in general. Additionally, the RC reaches its maximum at 52.2 m when the $RC = 16.6$ which is much lower than the maximum roughness of the constant tracks which reach an RC of 19.4 at 154 m.

4.5.4 Effects of External Influences

During roughness measurement environmental influences are especially prevalent. Figure 4.14 shows several images with external influences and includes the corresponding disparity map with the resulting quality value q . In general, dust formation completely covering the soil did not occur during the field trials because most dust is generated by the tines. The tines are separated from the camera by the packer roller, so almost no dust is drawn into the camera's field of view. An additional air drag removes dust during driving. The camera is mounted roughly 30 cm above the soil surface resulting in a very low air volume between the soil and the camera where dust can form. Stronger dust formation occurred when the machine was stationary and the PTO was activated, as shown in Figure 4.14 b. Roughness measurements were weighted inverse to the quality parameter q . Dust formation is the main factor, in addition to residues, which influences this parameter. If the quality parameter falls below a certain value, a message is sent to the user.

Motion blurring was avoided by using the Equation 4.5 in Section 4.2.4.

The residue handling methods of Section 4.2.4 were not investigated separately, therefore they are integrated in the results of Section 4.5.2.

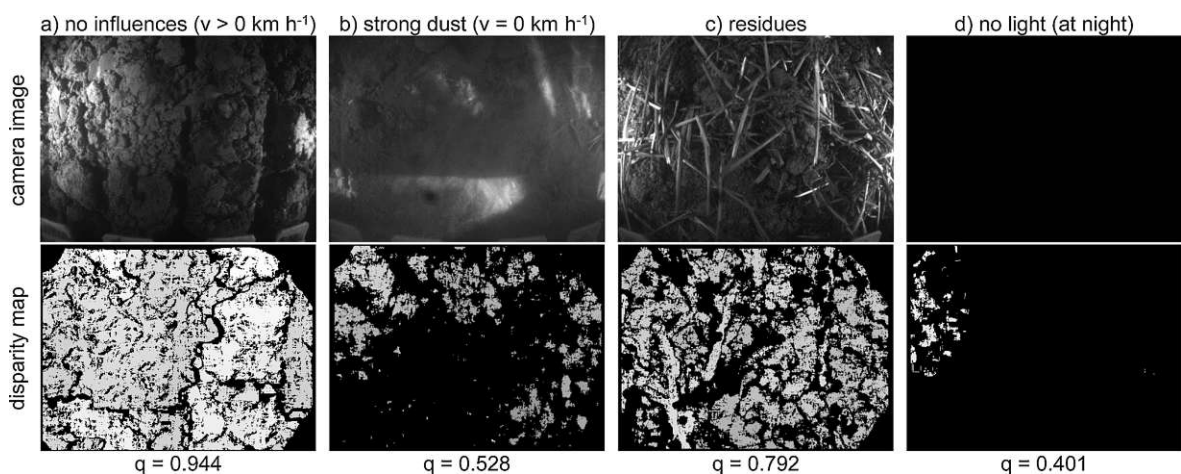


Figure 4.14: Effects of external influences on the image and the corresponding q values.

a) shows a regular example with good stereo detection quality during motion. b) is an example on the same field where dust was created deliberately by stopping the tractor turning the tines on the stand. No drafts remove the dust atmosphere. c) shows a scenario with residues in which peaks are removed from the disparity map to reduce their influence. d) is an image taken during night without lighting. This scenario is also detected by image parameter mean brightness.

4.5.5 Soil Roughness and Plant Emergence on a Field

As the investigations in [62] and [132] showed, increased soil roughness decreases plant emergence, so an additional experiment was carried out to assess this claim further.

Figure 4.15 shows the soil roughness recorded and geo-referenced on the left and the plant emergence mapped on the right. The plant emergence was mapped using the random forest method for soil cover estimation in Chapter 3. A correlation between increasing roughness and decreasing plant emergence is visible. This additional experiment was done only to confirm the connection already known from the literature between soil roughness and plant emergence. The research was not further deepened.

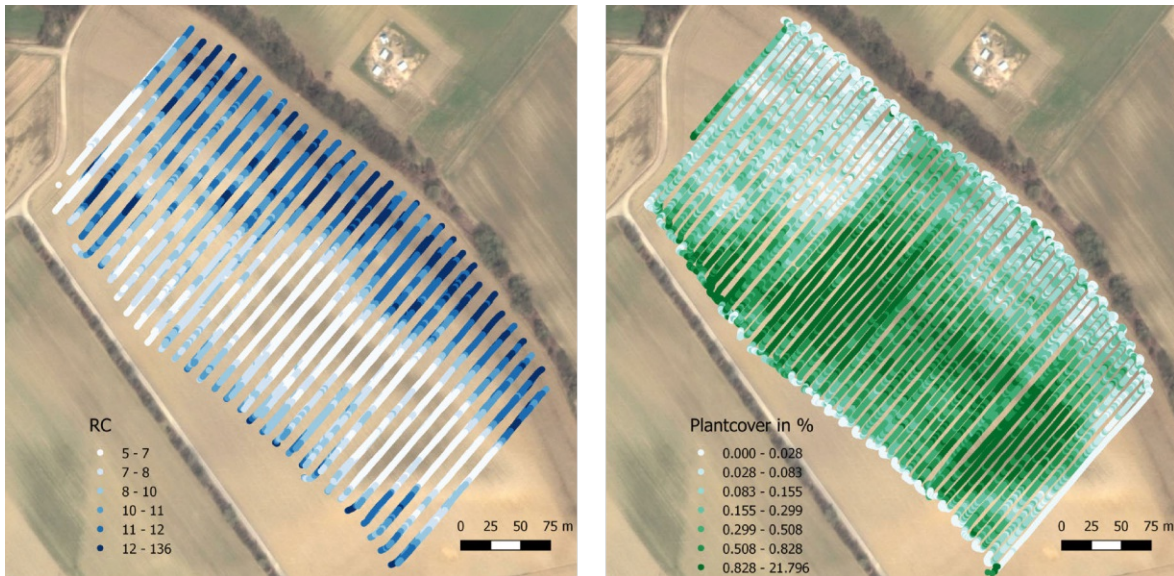


Figure 4.15: Soil roughness (left) and soil cover (right) estimated on a field in Wieselburg (AT). Soil cover is estimated with the entangled random forest method presented in Chapter 3.

4.6 Discussion

The relationship between RC and MWD was $RC = 0.14 \cdot MWD + 7.90$ with $R^2 = 0.30$ for experiment 1 and $RC = 0.12 \cdot MWD + 6.35$ with $R^2 = 0.43$ for experiment 2.

The work published in [76] used a laser profile meter to estimate the height standard deviation of soil surfaces (SD) which is very similar to the RC roughness. They estimated a linear relationship of $SD = 1.88 \cdot MWD - 4.02$ and $R^2 = 0.63$ between the MWD and their roughness measurement SD. There are some possible explanations as to why our results do not show strong correlations between MWD and roughness.

First, the different offsets between the first and second experiment might be explained by different moisture contents and different soil types. This assumption was supported by the work in [132] which also observed influence of soil moisture and soil type on MWD, especially with high tool speeds. This work further explains that the re-compaction and levelling after the packer roller of the power harrow pushes larger soil aggregates into the soil which results in a weak influence on the soil roughness. The range of soil roughness is reduced and leads to lower slopes in the MWD-RC plot which reduces the coefficient of determination R^2 at higher variances. The mounting position of the

camera directly behind the packer roller is not the optimal position, but other mounting positions are difficult to implement due to structural and practical constraints.

An additional effect influencing the measurement mentioned in [133] is sorting whereby larger soil aggregates are brought to the surface. This, in turn, is not nearly as a strong effect at moisture conditions observed in experiment 2 but also in experiment 1. However, it would describe the offset of the MWD-RC relationship between experiment 1 and 2 and should be investigated in further studies. An increasing moisture content in experiment 2 also led to a certain kneading effect which causes no reduction in soil roughness despite an increased tillage intensity.

From a geometric perspective, two additional factors influence the relation between MWD and roughness. The first factor is the observed area. Soil roughness is defined by the observed area and the resolution of the height measurements. The work in [80] proposes observing a minimum length of 841 mm of soil surface. Due to the limited number of mounting positions on the power harrow and the limitations of the camera optics, the observed soil area is only about 500 mm in length for our camera system. A sliding mean window was used, as described in Section 4.2.2, to increase the observed area. The 3-D point clouds were not matched due to limited processing resources. Each recording is considered separately and only the roughness values are averaged by the mean window. Future research could focus on reconstructing a closed recording of the soil surface by matching the individual point clouds at high frame rates. The second factor concerns the considerable curvature of the soil surface independent of the soil aggregates. In turn, this effect is not so influential due to a small observed soil surface area.

We observed that increased tillage intensity, defined by the number of power harrow tine loops per length, decreases soil roughness (see Figure 4.12) and vice versa. A correlation between tillage intensity and MWD was observed in the previous works [132] and [134]. This supports the claim that tillage intensity can be controlled by an estimated roughness to achieve a homogeneous seedbed in terms of MWD.

Tests with the controller showed promising results, as presented in Section 4.5.3. A homogenization of the soil roughness was observable. Additionally, it was possible to reach the desired RC roughness index within the hysteresis area. A clear effect was also visible subjectively and confirmed the assumption of the farmer on-site.

Therefore, the correlation between tillage intensity and roughness, and the results of the controller tests show that the stereo camera based roughness estimation works with sufficient accuracy to control the power harrow.

Remaining issues concern the controller overshooting the desired roughness by a certain degree which can be attributed to a large window ($n = 11$) of the sliding window averaging. In addition, this leads to the controller neglecting small-scale soil heterogeneities on a field. The controller is also delayed by computational delays and the inertia of the tractor. The next generation of camera systems, used for the use case in Chapter 5, use a different single board computer to reduce the computation time.

The relationship between the plant emergence and soil aggregate sizes has already been investigated in [132] and [62]. Smaller soil aggregate sizes improve plant emergence while increasing soil erosion [82]. We can also observe the effect of roughness on plant emergence, as shown in Figure 4.15.

The last topic to discuss is how to select the reference roughness R_{ref} of the controller. The optimal seedbed roughness is dependent on different parameters like soil moisture, soil type, plant species (kernel size), expected rain, etc. A simple way of operating the system could be to prepare a small area of the field manually by an experienced driver who tries to achieve the optimal seedbed based on his own experience. The observed mean roughness in this area can then be used as a target by the controller for the entire field. The remaining field will then be prepared according to the selected R_{ref} .

5 Grassland Yield Measurement

Currently, precision farming is mostly limited to arable farming due to the economic advantages of market crops and the optimal production conditions which are less common in permanent grassland. Precision farming technology is hardly used in grassland although it occupies large areas domestically and internationally. Across Austria, permanent grassland occupies more than 1.3 million hectares or 48% of the total agricultural land. With an area of 3.44 billion hectares, the proportion of permanent pasture land worldwide is roughly twice the size of arable land. Grassland is mostly found where agriculture is not possible due to technical challenges or an unsuitable climate. Arable precision farming methods have an enormous potential to be used for site-specific management of grassland. Biomass sensors for mowers are advantageous because they can detect yields directly before the biomass is harvested or at any point during the vegetation period [102]. On the other hand, estimating grassland parameters in a non-destructive fashion is challenging because of the heterogeneous composition of grassland. The variability of grassland ranges from favorable locations that are intensively used to alpine pastures with many different plant families, genera and species. The latest trends and developments in agriculture, especially in precision dairy farming, demand the precise measurement of grassland parameters.



Figure 5.1: Tractor and mower with cameras and sensors mounted, and people taking yield samples in the background.

The approaches for yield estimation and family or genus classification presented in this chapter serve as the foundation for automated machine control, site-specific applications, in-field-logistics and logistic optimization during harvest, and objective

farm resource planning. Automated machine control applications can involve steering along swath or controlling the tractor's driving speeds during harvesting with baler or silage trailers depending on the yield values. Applying nutrients at specific sites could reduce the negative environmental impacts and reduce the expenditures. Site-specific grass seeding has already been partially tested and applied in practice.

This chapter demonstrates how the RGB stereo camera system can be used to measure grassland. Different parameters have to be measured to estimate yields in heterogeneous grasslands. This chapter also focuses on the RGB stereo camera systems and builds on the methods presented in the previous chapters, but other sensors were used for the sake of comparison, including a near infra-red camera (850 nm), a laser scanner, an ultrasonic sensor and a pendulum-meter. The cameras and sensors were mounted on a mower and on a rake. The standing plants can be measured and classified on the mower while the the volume of the plant swath can be estimated on the rake.

Parts of this Chapter are published in [14] and in [20].

5.1 Grassland harvesting

Harvesting grassland consists of several steps also known as a harvesting chain. Different machines can carry out the harvesting, for example a diagram of silage harvesting is shown in Figure 5.2. Here, the mower (Figure 5.2 (a)) is followed by the rake (Figure 5.2 (b)) to form swaths which are then removed by a baler ((Figure 5.2 (c))), forage harvester or loader wagon. In order to produce hay, the grass is tedded several times after mowing (a) to reduce the water content. Silage harvesting also often involves tedding to reduce the water content of the grass.

Harvesting results in the grass having four different types: Grass stock (uncut plants), mown grass, swath and grass turf. Estimating the yield of mown grass is difficult due to the small volume of the spread grass. The grass stock can only be observed during mowing. The swath can be observed during raking and removal with a baler, forage harvester or loader wagon. The grass turf is visible during all steps.

Keeping these different grass types in mind, we decided to mount the camera system on the mower and on the rake. The camera attached to the rake can also be used on balers, forage harvesters or loader wagons to capture the swath with the same viewing angle but in the reverse viewing direction.

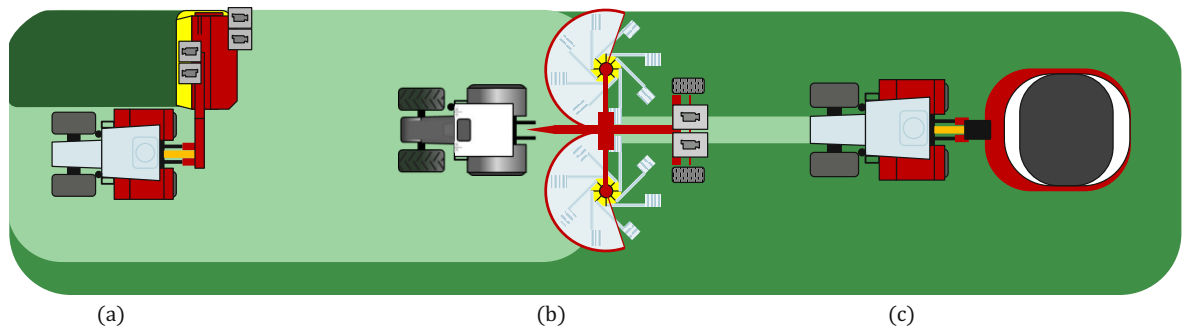


Figure 5.2: Typical tasks during harvesting grassland. Grass is mown (a), followed by raking after drying (b) and removal by a baler (c). Grass sward ■, mown grass and swath ■ and grass turf ■

5.2 Hardware Setup

The aim is to find hardware that can fulfil the tasks from Chapter 3 and 4, as well as, the grassland tasks. The additional hardware requirements for the grassland tasks are presented below, followed by the actually built and used camera setup. Additional sensors, already used for grassland and new to be used in grassland, were tested. The tractor's 12-V on-board voltage powered the camera system and all the sensors which required an external power supply.

5.2.1 Hardware Requirements

The hardware requirements are nearly identical to those used for the roughness measurements. There is more flexibility surrounding the mounting position of the stereo camera for plant measurement. However, an additional mounting structure is generally necessary on the mower. Mounting the camera on the rake is less complex. For both applications, the sensors and cameras can be mounted about 1.5 m above the ground. The geometries given in the Chapter 4 are sufficient for the field of view required for the swath measurement. A wider field of view on the mower would provide larger measurement area, but is hard to implement due to the strong parallax.

Classifying the family and the genus of plants requires a higher camera resolution than for the roughness estimation, because the fine structures of the plants must be recognizable. The image resolution is within the ranges of the required ranges for the soil cover estimation utilized in Chapter 3.

Mowing is usually done in daylight, so the use of additional lighting was not required but can be investigated in future research. The tractor's driving speeds are often quite high during mowing, but the optical flow is not much higher because of the camera's increased distance from ground unlike during the roughness measurement. As a result, the quantum efficiency and the shutter time are within similar ranges.

An additional near-infrared (NIR) camera is not required but is interesting to include for scientific purposes.



Figure 5.3: Cameras setup with two RGB cameras, a NIR camera and a NVIDIA Jetson Nano for inference of CNN models.

5.2.2 Test Setup

The following paragraphs describe the camera setup that was built based on the requirements. In order to, add the additional NIR sensor, all cameras must be synchronized. The cameras used for the roughness measurement from Leopard Imaging Inc. do not support external triggering. Therefore, a new camera system, shown in Figure 5.3, was set up which is backwards compatible enabling it to run the roughness and soil cover estimations. The camera system consists of two IDS RGB UI-3251LE-C-HQ cameras (1/1.8" e2v EV76C570 CMOS sensor with 1600 x 1200 pixel), an IDS UI-3241LE-NIR-GL camera (1/1.8" e2v EV76C661 CMOS monochrome sensor with enhanced NIR sensitivity and 1280 x 1024 pixel) mounted between the RGB cameras and a NVIDIA Jetson Nano single board computer. The cameras used 4.5 mm lenses for a wide field of view required on agronomic machines. The housing was 3-D printed with plugs for connecting USB 3.0 and Ethernet devices as well as a power supply of 12 V. The camera was controlled using a ssh connection based on the Ethernet or WIFI connection. The images were stored on the system with additional GPS tagging in the image capturing mode. Depending on the accuracy requirements, a GPS receiver or a RTK based precision GPS system can be connected to the system.

5.2.3 Additional Sensors

Additional sensors mentioned in the literature for use in grassland were integrated.

Three different sensors were integrated to measure sward height:

- Ultrasonic Sensor HC-SR04 (40 kHz sound waves)
- Radar Evaluation Board SiRad Easy® from Silicon Radar GmbH(122 GHz radar)
- SICK Mid range distance laser sensor DT35-B15851 (850 nm near-infra-red)

These three sensors each provided a distance value but with a different measurement principle. All sensors were mounted at the same height above the ground. The distance

readings of the sensors were averaged over one second so they could be synchronized with the camera images which are stored at 1 FPS.

Ultrasonic sensors have already been used for sward height measurements [99]. The ultrasound-based sensor is relatively inexpensive. The sensor measures the transmission time of an ultrasonic signal reflected from an object. The roughness of the object and the sensor frequency determine if an object reflects or defuses the signal [99]. The ultrasonic Sensor HC-SR04 is connected to an Arduino micro-controller to convert the analog voltage output of the sensor and to transfer it via USB to a computer. The sensor has a field of view of about 15° .

Radar from satellites is already used to measure plant height. Radar from short distances (proximal sensing) has rarely been used to measure plants heights. A method for height determination in cereals is commercially available [135]. We used a 122 GHz FMCW / CW radar sensor, commonly used in the automotive sector, to measure sward height in grasslands. When mounted, the sensor with the radar lens has an opening angle of about 8° . Electromagnetic waves of about 120 GHz are reflected by water. The grass sward mostly consists of water and can be read as a water volume by the radar sensor. The sensor integrated into the evaluation board returns a distance spectrum. The first bump is extracted from the signal after a certain threshold in the distance spectrum is reached. The internal housing of the sensor also reflects the signal, but can easily be filtered by calibrating it or reaching the minimum distance threshold. The distance of the first strong reflection (first bump) is stored as the sward height. The sensor is connected via a serial connection over USB. The protocol for reading the data was implemented in the C++ application.

The last sensor tested to measure sward height was a single point laser sensor in the 850 nm near-infra-red band. The sensor returns distance measurements at a very fast rate. The laser signal is reflected on the surface of the plants. Only the first target of the sensor is stored and processed for the purposes of this study. The laser sensor has an opening angle of 0.43° and therefore has the smallest field of view of all the distance sensors, which often results in the soil being detected in thin plant swards.

Drawing on the concept discussed in [89], we integrated a pendulum-meter into the mower protection sleeve, as shown in Figure 5.4. The pendulum-meter is utilized because in denser and higher plants the protection sleeve of the mower is deflected more strongly when the driving speed is also taken into account. The pendulum-meter consists of a rod which is attached to a ball bearing and is also connect to the mower protection sleeve. The deflection of the rod is measured at the pivot point of the ball bearing with a potentiometer. The angle of deflection is digitized and passed on to the logging computer with an Arduino micro-controller. The driving speed is logged with the GNSS RTK receiver mounted on the mower.



Figure 5.4: Pendulum-meter on the mower for measuring the deflection angle of the mower cloth.

The NIR image and the red channel of the RGB image can be used to calculate the Normalized Differential Vegetation Index (NDVI) of different image areas. The NDVI usually captures the biomass of an area very well. The camera system individually changes the exposure times for the RGB sensor and for the NIR sensor. In order to eliminate the influence of the different exposure times, the NIR channel is divided by the exposure time of the *NIR* sensor t_{expNIR} , and the *Red* channel is divided by the exposure time of the RGB sensor t_{expRGB} . A more detailed investigation of the spectral reflection in the IR band was deliberately omitted from this project. The equation for calculation of the NDVI is as follows:

$$NDVI = \frac{\frac{NIR}{t_{expNIR}} - \frac{Red}{t_{expRGB}}}{\frac{NIR}{t_{expNIR}} + \frac{Red}{t_{expRGB}}} \quad (5.1)$$

The mower was equipped with the two camera systems described in Section 5.2.2 and one of each of the additional sensors mentioned above in Section 5.2.3. One of the cameras was facing forwards towards the grass sward and one camera was facing backwards on the cut edge between the grass sward, the grass turf and the mown grass. Figure 5.5 depicts the mounting positions of the sensors and both camera systems.

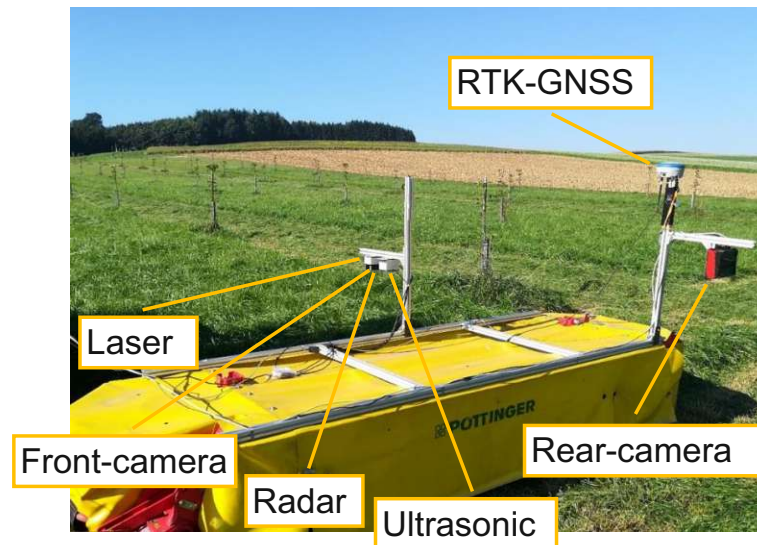


Figure 5.5: Sensors on the mower: one camera setup is facing forwards towards the grass sward and one backwards on the cut edge. The sensors are mounted on the front facing downwards into the grass sward.

Additionally, a single plane LiDAR sensor SICK LMS111 estimates the volume of swaths on the rake. The LiDAR sensor measures distances point-wise with a 905 nm near-infra-red light. The rotation of the scanner provides planar sampling. The LMS111 was configured with a 25 Hz and 0.25° angular resolution. The sensor's field of view is -45° to 225° and has a working range of 0.5 m to 20 m. The laser scanner is connected through Ethernet to a recording computer.

5.3 Volume Estimation during Mowing

Sward volume is one important factor in order to estimate yields in grasslands. Due to the given geometries of the camera's field of view, it is not possible to observe the entire mowing area with the stereo camera system, therefore the mean volume per area is a more tangible feature. This involves measuring the mean sward height which is then multiplied over the observed area.

The sensors were mounted on a disc mower NOVADISC 350 from Pöttinger. As mentioned above, one camera is facing forwards and one camera is facing backwards. Figure 5.6 illustrates a bird's eye view of the mower and tractor.

Each image and sensor reading is stored with a time-stamp and a GNSS location. The entire system is integrated into a Linux application written in C++ which records all image (Section 5.2.2) and sensor data (Section 5.2.3) and synchronizes it with the GNSS RTK readings.

The sward height measurement is based on the same calibration procedure and stereo pipeline as described in Section 4.2.1 and Section 4.2.2. Additional handling of the exposure time is based on method in Section 4.2.4. The 3-D point cloud, which is

returned from the stereo pipeline estimates the mean and median height of the forward facing camera. The median height appears to be more robust towards outliers in the 3-D reconstruction of the grass sward surface. The front facing camera and all the distance sensors (ultrasonic, radar and laser sensor) run into the same issue of only measuring the distance between the sensor and the top of the grass sward. The known ground distance serves as the basis to calculate the plant height. When changing the mounting height, the sensors must be calibrated on a flat surface to measure the ground distance again. Additionally, accelerations from driving disturb the sensor mounting arms resulting in an error which a distance measurement to the top of the grass sward cannot compensate for.

Figure 5.7 shows sample images captured from the front facing camera. The left or right image of the stereo camera can be directly used to classify a plant's family or genus as discussed in Section 5.4. The yellow protective cloth of the mower is often also visible in the image, so it must be recognized and masked. This issue is discussed later on in this section.

In summary, the processing pipeline starts with segmenting the image into regions followed by stereo processing, height estimation and family and genus classification. Finally, the yield is estimated by a machine learning method based on the features mentioned in the previous step.

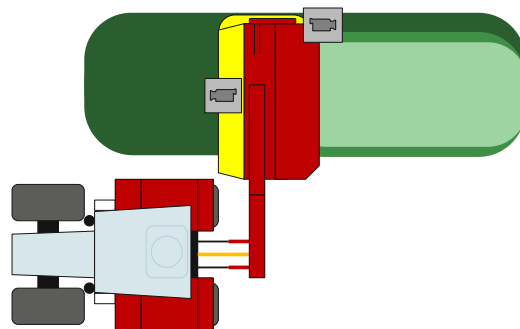


Figure 5.6: Cameras on the mower. One camera is facing forwards towards the grass sward and one camera is facing backwards on the edge between the grass sward ■, the grass turf ■ and the mown grass ■

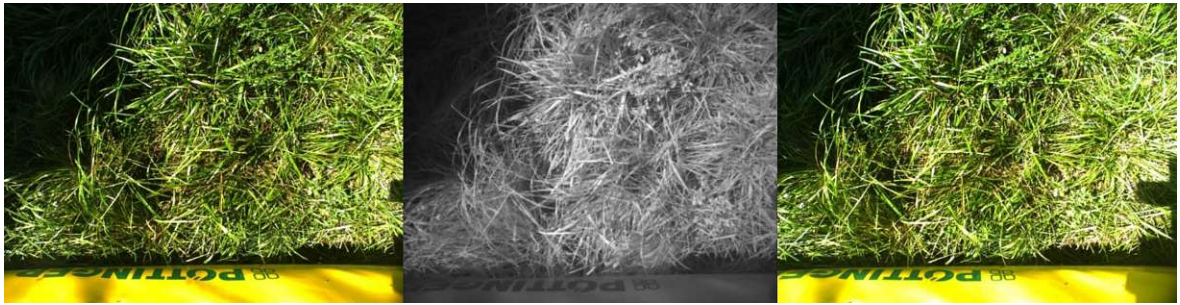


Figure 5.7: Front camera view on the mower. Left RGB Image, NIR image (850 nm) and Right RGB image.

The backwards facing camera observes different areas, such as the grass sward, the grass turf, and the mown grass. This setup offers additional viewing angles unlike the front facing camera.

First, it is possible to investigate the grass turf's condition and density. The NDVI of the grass turf area is a simple metric making it easy to calculate. Future research should focus on how to measure the grass turf area in more detail using more complex algorithms like semantic segmentation.

The absolute height of the grass can be calculated by viewing the grass turf and measuring the distance between the top of the grass sward and the cut height which is represented by the top of the grass turf. By calculating the absolute height, the sensor height calibration becomes obsolete. Additionally, the errors from the acceleration of the mounting arm are eliminated, because the height difference between the grass turf and the grass sward can be calculated from each recording. The visible area of the grass sward can be used to classify the family or the genus of plants as discussed in the Section 5.4. Figure 5.8 shows an example image of the backwards facing camera.



Figure 5.8: Rearview camera on the mower with grass sward (uncut plants), the grass turf, the mower cloth and the mown grass visible. Left RGB image, NIR image (850 nm) and right RGB image.

In order to take advantage of the rearview, the individual areas must be distinguished. First, a static mapping of the different image areas was configured, but this setup is error-prone when measuring different grass heights, measuring on the headland, or

driving in a curve. The mower cloth must be segmented independently of the static mapping due to its dynamic movement. By adding the mower cloth class to the semantic segmentation, it can be filtered in the same step from the images. The class mower cloth is excluded from all calculations (grass turf height, family and genus classification and NDVI estimation).

To be able to deal with all these situations, we decided to train a CNN which can detect and segment areas of the grass sward, the grass turf, the mower cloth and the mown grass within the image. The segmentation CNN is based on the bonnet framework [116] as presented in Section 3.4.2. 382 train images captured during the field trials between 2018 and 2020 were annotated. The data annotation tool described in Section 3.2.1 in the free drawing mode was used. The images were rescaled to a resolution of 512 x 384 pixels. It has been observed that a higher resolution is not necessary for this task. Example images are shown in Figure 5.9. The segmentation distinguishes between the classes grass sward (uncut plants), the grass turf, the mower cloth and the mown grass. A rectangular sub-area of the first class, grass sward, is segmented to estimate family and genus composition (see Section 5.4).

A train/test set split was performed to evaluate the accuracy of the method. The results depending on the pre-training are shown in Section 5.8.5. The rearview camera also implicitly integrates the full image segmentation into the yield estimation results (Section 5.8.1 and 5.8.3).

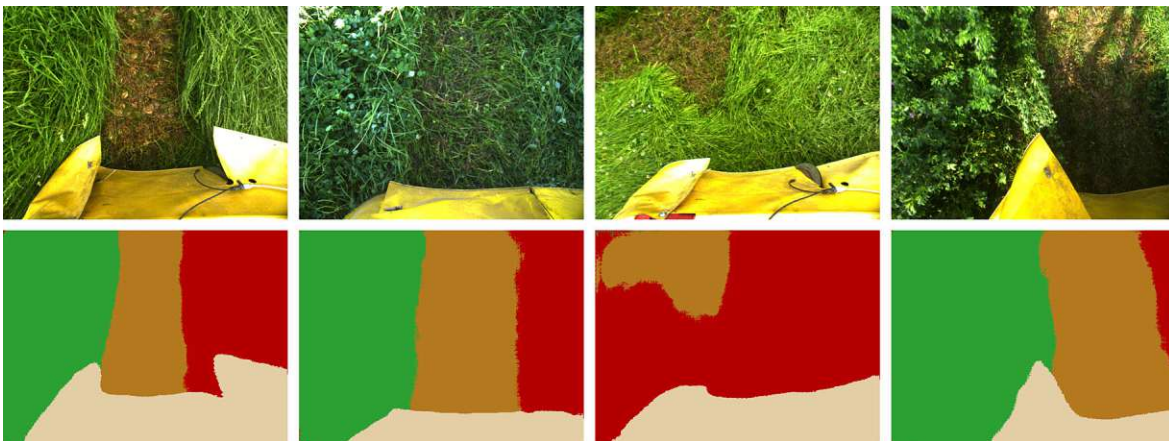


Figure 5.9: Results of the rearview mower segmentation: four examples of images captured on the mower and their corresponding segmentation results below. grass sward (uncut plants) ■, grass turf ■, mower cloth ■, mown grass ■.

5.4 Family and Genus Classification

The composition of a plant's family and genus in grasslands is an additional parameter to measure yield. Additionally, it can be used for targeted livestock feeding. At feeding, the legume portion can be controlled when dosing the amount of protein. A geo-referenced map of legumes can provide information about the nutrient supply and the general

condition of the soil. From an agronomic point of view, the proportion of grasses and legumes is interesting. Additionally, rumex is also an issue in grasslands, because it is considered a weed and must be removed with herbicides or mechanically. It is also interesting to distinguish clover from other legumes.

The information summarized above results in the following classes: soil, grasses, legumes, clover, rumex and other. Later on, this chapter will also demonstrate the challenges in differentiating legumes, clover and rumex due to limited data and the similarities between the classes.

The same method utilized for the soil cover estimation was applied to semantically segment the grassland images. The segmentation is based on the implementation of ERFNet [112] in bonnet framework [116] which is further described in Section 3.4.2.3.4.2.

5.4.1 Dataset

A grass clover dataset is publicly available and consists of artificially created images [136]. The images are random mosaics of single cut out plants (grass and clover). An example image is shown in Figure 5.10 (right). We tested the effect of pre-training with this dataset on the ERFNet.

The sugar beet dataset [118] was the second dataset for pre-training the ERFNet. It consists of images of sugar beets and weeds on different soils (see Figure 5.10 on the left).



Figure 5.10: Example images of the sugar Beet dataset [118] (left) and the grass clover image dataset [136] (right).

Manually annotated images were utilized to fine tune the family and genus classification model. The training images were recorded during the field trials, as shown in Table 5.1. Figure 5.11 shows two example images which were challenging to classify due to motion blurring and strong sunlight. In general, our dataset was created while the mower was in motion, under varied lighting conditions, and under other everyday conditions. Rumex hardly occurred in the trials and is therefore not represented in the dataset. In order for an application to detect rumex, the dataset would have to be expanded for the specific plant.

1030 training images with a resolution of 512 x 384 pixels were cropped from the randomly selected high resolution images (1600 x 1200 pixels). In contrast to the full image segmentation of the sample images in Figure 5.9, the full geometric resolution was used to detect the fine details of the plants. The same tool for data annotation described in Section 3.2.1 was used.

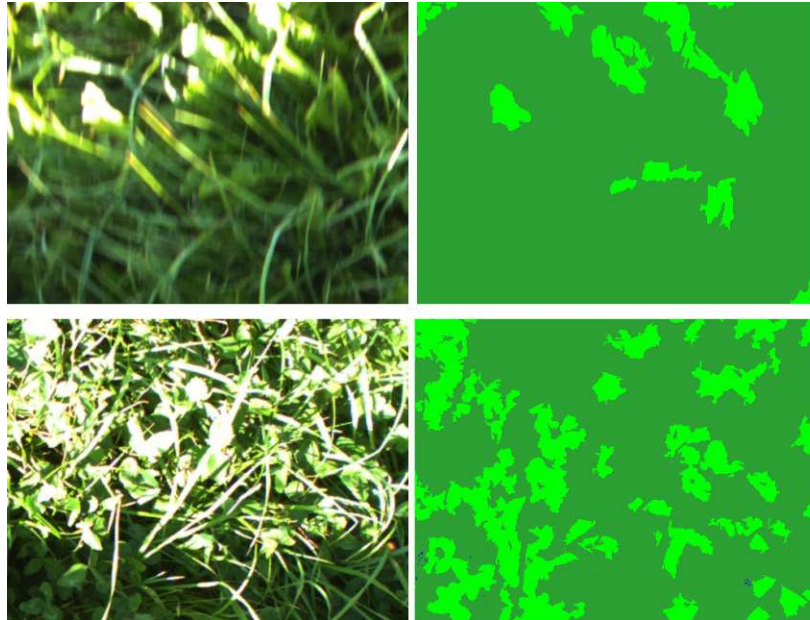


Figure 5.11: Two challenging examples from our grass legumes dataset with motion blurring (top) and high dynamic range by direct sunlight (bottom). Grass ■ and legumes ■

5.5 Yield Modelling from Sensors on the Mower

The cameras on the mower estimate several parameters which serve as a basis to model the dry matter yield. The main parameters estimated by the camera are the median and mean sward height, the grass-legumes ratio and the NDVI of the grass sward. These parameters are the features used for simple machine learning models. Regression machine learning methods try to model the dry yield as an output variable. Four different machine learning methods, implemented in the scikit-learn library [137], were tested.

One of the methods was the support vector regression (SVR) with an RBF kernel. The epsilon value of the epsilon-SVR model was set to 0.2. The regularization parameter C was set to 0.3 and is inversely proportional to the regularization. The implementation in scikit-learn is based on libsvm and works well for a low number of trainings samples [137], as is the case for our purposes.

An ordinary least squares Linear Regression with normalization was also tested. Non-linearities cannot be mapped, but over-fitting can be prevented.

The multi-layer perceptron regression was tested with one hidden layer with three units and a tanh activation function. An Adam [138] solver with a constant learning rate of 0.01 was used to optimize the weight. The other parameters like momentum and beta of the Adam solver were all set to the default values in scikit-learn.

The last method tested was the K-Nearest neighbors regression from scikit-learn. The regression was used with $k = 2$ meaning that the two nearest neighbors in the feature space were uniformly applied to calculate the regression value.

The evaluation results of the different machine learning methods are presented in Section 5.8.3.

5.6 Swath Measurement

Automatic guidance along swaths, yield estimation, machine control and logistics optimization in hay and silage harvesting all require the precise location and measurement of grassland swaths. Current methods use ultrasonic, radar, laser sensors, or cameras for this task. The central delivery rake Top 652 C from Pöttinger was used for all trials. We mounted the camera presented in Section 5.2.2 on the back of the rake backwards facing, as shown in Figure 5.12, to record sensor data.

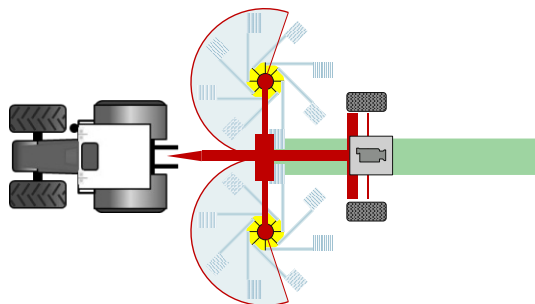


Figure 5.12: Cameras on the rake facing towards the swath.

The swath and the grass turf are in the camera and laser scanner's fields of view. The camera is tilted approximately 45° towards the swath in the direction of the travel so that a longer swath section can be captured. This allows for the swath to be recorded in a closed stream up to a certain speed. Figure 5.13 shows the camera, the laser scanner and the RTK GNSS receiver mounted on the back of the rake.

When recording the swath, it should be noted that the entire working width is not always used due to driving which is either imprecise or near boundary areas. This leads to smaller swath volumes and does not reflect the characteristics of the local plants. This is important to note when plant-related conclusions are to be drawn. However, an accurate GNSS can reduce this error.

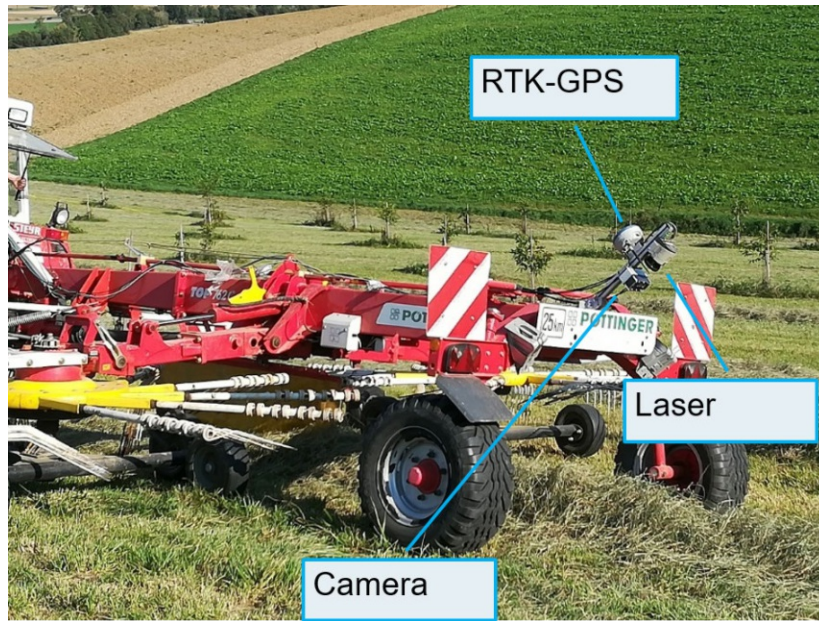


Figure 5.13: Cameras and sensors mounted on the rake facing backwards on the swath.

Figure 5.14 shows the left RGB Image, the NIR image (850 nm) and the right RGB image of a swath.

The NIR image and the red channel of the RGB image can be used, as with the grass images, to calculate the NDVI (Equation 5.1). In this case, the NDVI can only be calculated on the grass turf next to the swath.



Figure 5.14: Captured images of a swath. Left RGB image, NIR image (850 nm) and right RGB image.

The swath volume estimation is based on the same calibration procedure and stereo pipeline described in Sections 4.2.1 and 4.2.2. However, the swath must first be segmented to calculate the volume. Two different methods have been utilized to complete this task. The first method, described in Section 5.6.1, segments based on depth information. The second method, mentioned in Section 5.6.2, segments using the left RGB image and a CNN. In both cases, the images are rescaled to 1/3 of the original image size before stereo processing. Preliminary tests have shown that a 1/3

image resolution is sufficiently accurate to measure swaths. Higher resolution images of the swath do not depict the relevant structures. After segmenting the point cloud, the swath's points are used to estimate the volume against the plane. The volume is calculated by triangulating the points based on the PCLs OrganizedFastMesh method [139]. The ground plane estimation uses the PCL SAC segmentation [139] in which the swath area is left out.

5.6.1 Swath Segmentation based on Depth Information

The 3-D point cloud of the stereo camera can be used to segment the swath. During calculations, it is assumed that in most cases the swath is in the middle of the image and the area next to the swath is flat. The various steps needed to perform this process, from image acquisition to swath segmentation, are shown in Figure 5.16.

The first step involves removing the center region from the point cloud where the swath is expected to be located. The remaining points are used to estimate a plane which approximates the ground surface and is estimated using the Point Cloud Library (PCL) SAC segmentation [139]. Then, the point cloud is rotated so that the ground surface lies in the XY plane. Finally, all points above the plane are marked in the original point cloud. If needed, the marked 3-D points can be projected back onto the image using the extrinsic and intrinsic camera parameters.

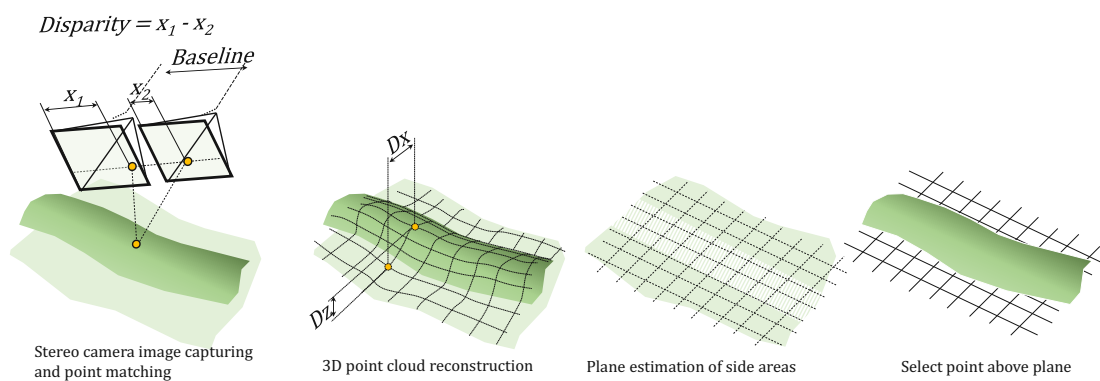


Figure 5.15: Swath segmentation based on depth data

5.6.2 Swath Segmentation based on RGB Image

Methods based on 3-D data from laser scanners or 3-D cameras have problems segmenting the swath, especially for small swath volumes. The second method is based on the assumption that the swath can also be segmented by the left RGB color image of the stereo camera. The advantage here is that it produces a higher resolution of more distant areas. The depth resolution declines quadratically with the distance, and the resolution of the color image declines linearly.

For this purpose, 189 training images were annotated into the classes swath and no swath. The tool for data annotation described in Section 3.2.1 was used. The most

efficient way of annotating the images was to use the free draw feature with flood filling. The images were resized to 512 x 384 pixels for training.

For the semantic segmentation of the swath images, the same method from the other semantic segmentation tasks was applied. Section 3.4.2 presents further details about this method. The network was pre-trained again in two versions of the sugar beet dataset [118] and of the grass clover dataset [136]. The effect of using different datasets for pre-training is presented in Section 5.8.5.

5.6.3 Swath Volume Estimation based on Laser Scanner Data

The laser scanner SICK LMS111 returns data in an array of points described by their angle, distance and reflectance. Based on the angle and distance, the points can be located relative to the sensor's center. Figure 5.16 shows a plot of the points in gray. In order to estimate the swath cross-sectional area, a similar approach to the 3-D volume determination from Section 5.6.1 was implemented in 2-D.

The ground line was estimated using a Hough Transformation for lines implemented in OpenCV [130]. During this process, the middle area, where the swath is expected, is left out. This step corresponds with the plane determination in the 3-D case.

The points are approximated using Catmull-Rom splines implemented in the Boost Libraries [140]. Spline approximations are used for simple estimation of the area under the swath contour. This is similar to the triangulation method applied in the 3-D case.

In the final steps, the area under the contour is summed up starting from the center outwards until the contour intersects the ground line. This step is represented by the black lines and the red dots in Figure 5.16.

The measured data can be mapped using the GNSS data to get a swath thickness map. In order to get the entire yield volume, the swath cross-sections have to be integrated over the distance travelled from the GNSS data.

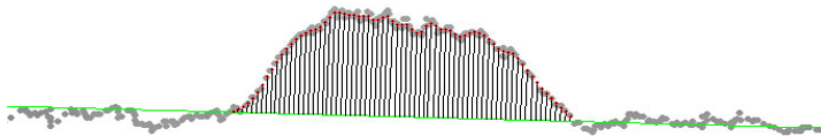


Figure 5.16: The swath cross-sectional area estimation from the laser scanner data. Points from the LMS111 are shown in gray. The approximated ground plane is shown in green and all points above the plane after the first intersection are marked in red.

5.7 Field Trials

Yield samples were taken continuously over three seasons at harvesting. The mower was equipped with different sensors which captured data continuously. Afterwards, random samples were taken to measure the yield per m^2 . A tarpaulin and the suspension scale Kern CH 15K20 with a weight range of 15 kg and a readability of 20 g were used as shown in Figure 5.17. A mowing width was weighed over a length of 1 m or 2 m in the case of the lower yields. Additional moisture samples were taken to estimate the gravimetric moisture content using the Darr method (gravimetric moisture measuring method). About four moisture samples were taken at different points in time during one field trial. The interpolated gravimetric moisture contents were used to calculate the dry matter yield of each sample point. These sample points were combined with the sensor and the camera data to form the training data for the yield model which is described in Section 5.5.



Figure 5.17: Manual weighing at random sample points during mowing and raking with the suspension scale and a tarpaulin.

In order to eliminate inaccuracies caused by the human rather than automatic driving, a tractor with RTK was used during the 2020 season. Table 5.1 shows the harvest times and locations where yield samples were taken after mowing. The dry matter content of the harvest shown in Table 5.1 is based on moisture samples taken during weighing with the suspension scale. The amount of moisture in the harvested grass is mainly determined by the time elapsed after cutting and less so by the heterogeneities of the field. It is assumed that the volume measured by the camera is more dependent on the plant's dry matter content than its moisture content. If standing grass were dried, the measured height would hardly change. The situation is different for swaths because dried grasses have a smaller masses, so they are less compressed by their own mass, as a result dry swaths have a greater volume. Therefore, the moisture content must be known to estimate the mass of swaths.

Table 5.1: Field trials mower

Date	Location	Number of Sample Points	Family/Genus composition	Dry matter content of harvest (%)
2018-08-22	Wieselburg	5	grass-clover silage	0.326
2018-09-18	St. Leonhard am Forst	41	lucerne and grass-clover hay	0.252
2018-10-10	Wieselburg	44	lucerne and grass-clover silage	0.254
2019-05-07	Wieselburg	38	grass-clover silage	0.200
2019-06-12	St. Leonhard am Forst	26	grass-clover hay	0.300
2019-06-15	Wieselburg	15	grass-clover silage	0.267
2019-07-23	Wieselburg	35	lucerne and grass-clover silage	0.380
2019-08-28	Wieselburg	26	lucerne silage	0.285
2020-05-07	Wieselburg	25	grass-clover silage	0.187
2020-06-11	Wieselburg	38	lucerne and grass-clover silage	0.229
2020-07-20	Wieselburg	26	grass-clover silage	0.271
2020-08-26	Wieselburg	31	grass-clover silage	0.291
2020-10-06	Wieselburg	32	grass-clover silage	0.177
Number of Sample Points		382		

Trials with the rake were conducted during the 2018 and 2019 seasons. Most of the trials were conducted directly after a mower trial (2018-08-22, 2018-09-20, 2019-05-08, 2019-06-15 and 2019-07-23). The rake trials were similar to the mower trials. The rake equipped with the sensors captures sensor and GNSS data when raking occurs behind the machine. Afterwards, the random sample points were selected. The tarpaulin and the suspension scale Kern CH 15K20 were used again to weigh one or two meters of the swath depending on its overall expected yield. Additional dry matter samples were taken to calculate the dry matter yield. Table 5.2 shows the harvest times and locations where yield samples were taken after raking.

The driving path during raking produces certain inaccuracies just like during mowing. In principle, it is best to drive into fully mowed areas, but this was unfortunately not possible due to logistical conditions during harvesting. Instead, lane to lane driving was carried out potentially resulting in a smaller swath due to inaccurate driving and a distortion of the actual yield measured. However, accurate GNSS measurements can factor out these tracking inaccuracies, which may also occur in later practical applications.

Table 5.2: Field trials rake

Date	Location	Number of Sample Points	Family/Genus composition	Dry matter content of harvest (%)
2018-07-31	St. Leonhard am Forst	23	grass-clover hay	0.9142
2018-08-22	Wieselburg	9	grass-clover silage	0.5010
2018-09-20	St. Leonhard am Forst	35	lucerne and grass-clover hay	0.8735
2019-05-08	Wieselburg	22	grass-clover silage	0.3518
2019-06-05	St. Leonhard am Forst	28	grass-clover hay	0.8938
2019-06-15	Wieselburg	11	grass-clover silage	
2019-07-23	Wieselburg	15	grass-clover silage	0.5049
Total Number of Sample Points		143		

5.7.1 Outdoor Laboratory Test for Swath Measurement

An additional test was conducted to evaluate the accuracy of the swath volume estimation in a controlled environment. A flat and mown grass area was used for this purpose. The camera was mounted on a stand to obtain the same point of view as on the rake. Hay was manually applied step by step to the area to form a swath. The hay was taken from pressed square bales. Each time hay was added, the mass was weighed with the suspension scale which made it possible to eliminate the error caused by the ground plane detection and to only focus on the volume measurement. Additionally, multiple repetitions make it possible to determine the random noise in the measurement. Figure 5.18 shows the camera mounted on the stand with a swath in front.



Figure 5.18: Outdoor laboratory trial to validate the swath measurement system under controlled conditions.

5.7.2 Comparison of different Raking Concepts

In 2019, students compared the ability of a star-wheel rake (Sitrex 7.10 m), a band-rake (Zagroda 2.20 m) and a side delivery rotary rake (Pöttinger 3.00 m) to swathe grasslands. The purpose of this experiment was to find out if the raking system influences the ability of the camera to estimate mass. They used the camera system proposed in this thesis to measure the swaths. The trials took place in St. Leonhard am Hornerwald (AT) at two fields ($48^{\circ}36'3.2328''N$, $15^{\circ}32'41.5104''E$ and $48^{\circ}36'8.3592''N$, $15^{\circ}32'56.4648''E$) on the 21st September 2019. The fields were planted with rye and red clover, which served as the nurse crop and were subsequently mown. Camera images of the swaths were taken with the stand shown in Figure 5.18. In all trials, the masses of the swath at $N=24$ random sample points were measured with the Kern CH 15K20 suspension scale. The dry mass samples were taken as a subset of the sample sites. The results of the trial are shown in Section 5.8.4 and in Figure 5.33.

5.7.3 Targeted Trial for Grassland Family and Genus Classification

In 2020, a plot trial with different clover-grass mixtures was conducted. During the trial, eight different variants in 3 m strip plots were seeded. 32 measurement points, four for each variant, were selected. The 8 clover-grass mixtures from SAATBAU LINZ eGen are shown in Table 5.3. The plots were mown three times with a hedge trimmer at the measurement points, on 2020-05-27, 2020-08-12 and 2020-10-06. An area of 60 x 50 cm was mown for each sample point which corresponds to the field of view of the camera. The yield per square meter can be calculated from the sample area.

Table 5.3: Clover-grass mixtures seeded for the targeted family and genus classification trial.

Mixture	Legumes (%)	Grasses (%)
FUTTERPROFI EI (Triple N - Schaumann instead of EI)	50	50
GRÜNLANDPROFI B	14	86
GRÜNLANDPROFI EB	10	90
FUTTERPROFI WM	26	74
FUTTERPROFI KM	33	67
Lucerne	100	0
FUTTERPROFI LR	50	50
Red clover	100	0

The measurements were taken weekly in 2020 with the camera system presented in Section 5.2.2 at the 32 measurement points. The camera was mounted on the stand shown in Figure 5.18 but also with a perpendicular point of view as on the mower.

In order to evaluate the experiment, the mass fractions of legumes, grasses and others were weighed directly after mowing. Correlations between camera-predicted ratios of grasses and legumes and the weighted ratios were calculated, as well as, the correlation between the camera-predicted ratios and the ratios of the seeds of the mixtures as presented in Table 5.3.

5.8 Evaluation

The evaluation of the measurement systems is based on weighing yield or the yield fractions (for estimation of grass/legumes ratio). The yield of the dry matter served as the basis for all the evaluations. As described, the dry matter samples were taken during weighing and packed into air tight freezer bags. The correlations between the weighed yields and the sensor or model data are described by the Pearson correlation coefficient or by the R^2 , which equals the squared Pearson correlation coefficient for linear models.

First, the separate evaluation of the different features captured by the camera and other sensors is presented. The second section examines the evaluation results of the family and genus composition classification. This is followed by an evaluation of the combination of the features into an overall yield model. The fourth section deals with the evaluation of the swath's segmentation and yield estimation. Finally, the effect of pre-training on the different applications of semantic segmentation is examined.

5.8.1 Volume Estimation and other Sensor Measurements during Mowing

The evaluation of the method for plant height measurement described in Section 5.3 is based on data from the field trials. During the field trials listed in Table 5.1, the camera

and sensor data was recorded and the yield was measured at random sample points.

The following Figure 5.19 shows the features estimated by the camera mounted on the front of the mower and facing downward. Each field trial is color-coded which reveals the field trial biases. The R^2 value for each feature is calculated from the difference of the feature and the dry matter yield from all field trials combined.

The first figure in the top left corner shows the median height grass sward estimated by the camera. The value was manually calibrated by measuring the camera height on the field. In contrast, the mean height grass sward, shown in the top right corner, is more prone to outliers. The NDVI (Equation 5.1) of the grass sward in the bottom left corner is calculated from the entire image excluding the mower cloth. The legumes portion in the bottom right corner is estimated based on the method from Section 5.4. Additionally, the trials on the lucerne fields with a high portion of legumes are visible on the right side of the figure.

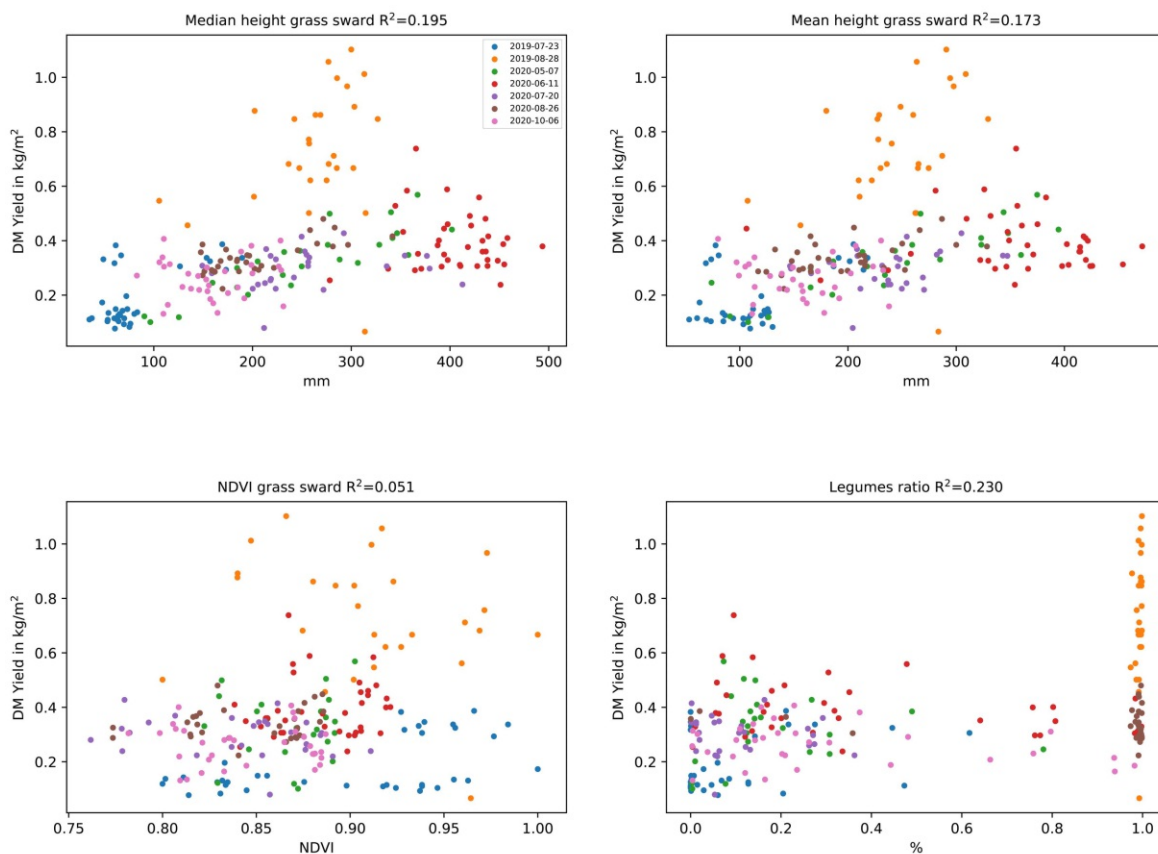


Figure 5.19: Data from the forwards facing camera. The individual field trials are color coded.

Figure 5.20 shows the estimated features of the backwards facing camera on the cut edge.

The figure on the top left and the top right show the mean and median distances to the grass sward from the camera. Both values are calculated from the area classified as

grass sward. Unlike the front facing camera, they are not manually height calibrated.

The median and mean sward heights shown in the middle left and right, respectively, are calculated as the height difference between the areas classified as grass sward and grass turf. As expected, these two features correlate best to the dry matter yield compared to the others.

The legumes portion is estimated by cropping a patch from the region classified as grass sward. The NDVI of the grass sward is estimated from the region classified as grass sward. Neither the legumes portion nor the NDVI correlate to the yield.

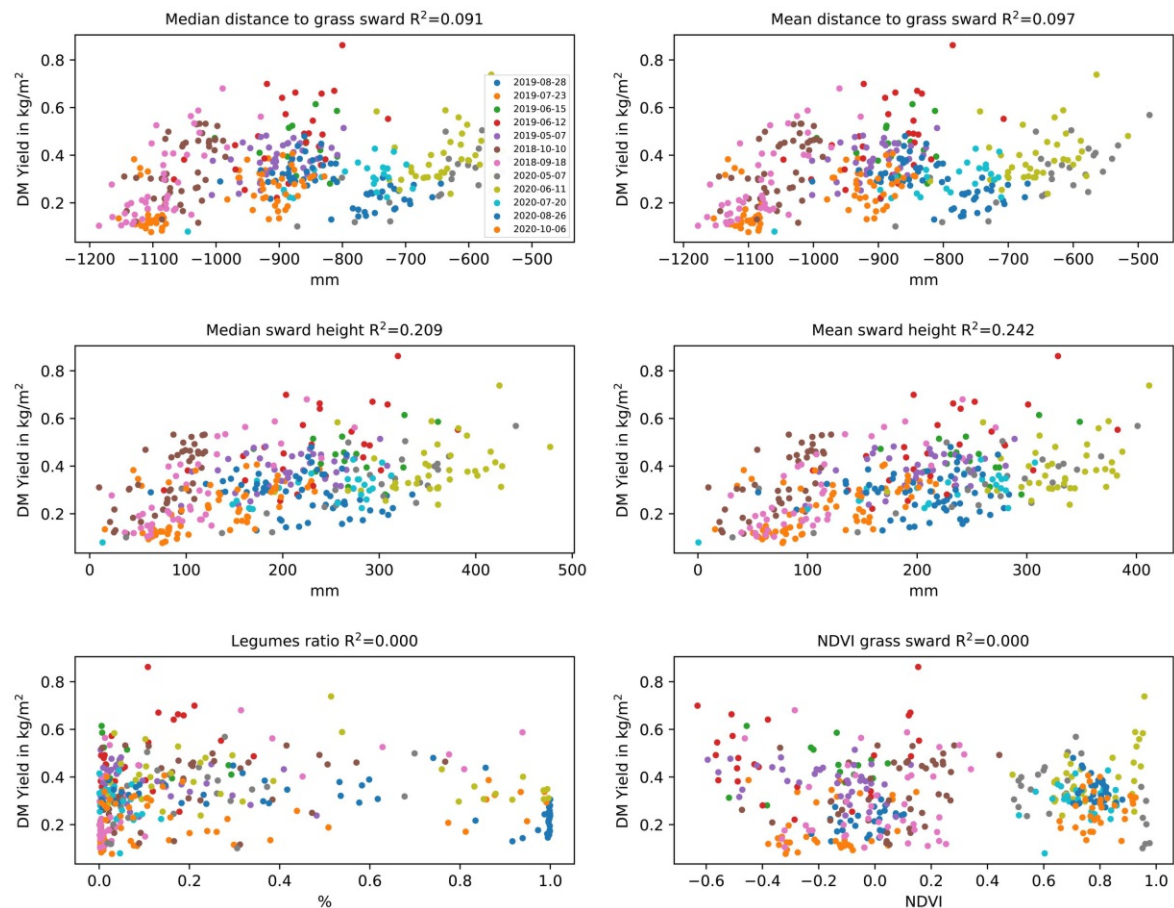


Figure 5.20: Data from the backwards facing camera. The individual field trials are color coded.

The additional sensors, which were not integrated into a yield model, barely resulted in a more accurate yield prediction. The measurements of the sensors described in Section 5.2.3 were compared by applying the Pearson correlation to the dry matter yield measured during the trials. The Pearson correlation between the angle of mower cloth and the weighed mass was 0.1611 during the trial on 2019-08-28 and 0.5162 on 2020-05-07.

The Pearson correlation of the distance sensors radar, ultrasound and laser are presented in Table 5.4. The laser sensor achieved the best correlation with the dry

matter yield followed by the radar sensor.

Table 5.4: Correlation between sensor readings and the weight of the dry matter yield (DMY) for the sensors on the mower.

Date	Pearson Correlation of DMY to Radar	Pearson Correlation of DMY to Ultrasound	Pearson Correlation of DMY to Laser
2020-05-07	0.5490	0.1615	0.7210
2020-06-11	0.0835	0.0879	0.0434
2020-07-20	0.1073	0.4118	0.0822
2020-08-26	0.8128	0.2455	0.7833
2020-10-06	0.1026	0.4162	0.4254

5.8.2 Grassland Family and Genus Classification

The method for estimating the family and genus composition presented in Section 5.4 enables color images to be segmented into grasses and legumes. The quality of the grass/legume ratio predictor is evaluated in different ways. One way is to split the annotated images into training and test sets. An example image of grass/legume segmentation is shown in Figure 5.21. The figure shows the RGB image, the prediction and the ground truth labels of the test image.

The accuracy of the prediction on the test set was 0.8522 and the intersection over union (IoU) was 0.5374.

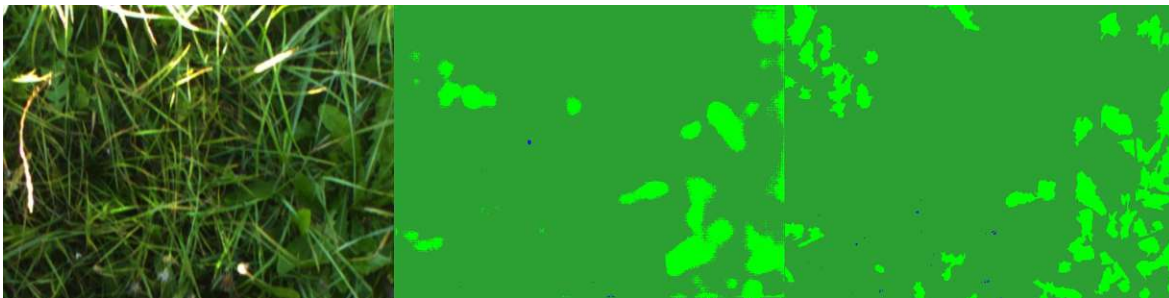


Figure 5.21: Example image (left), prediction (middle) and ground truth labels (right). Grass ■ and legumes ■

The second method to evaluate the segmentation of grasses and legumes was the targeted trial for grassland family and genus classification described in Section 5.7.3. The images were taken in weekly intervals until the grassland was cut. The grasses and legumes portions were weighed at the sample points.

Figure 5.22 shows the correlation of the grasses/legumes ratio with the weighed ratio in the seeds. It can be observed that the correlation decreases with the temporal distance to sowing.

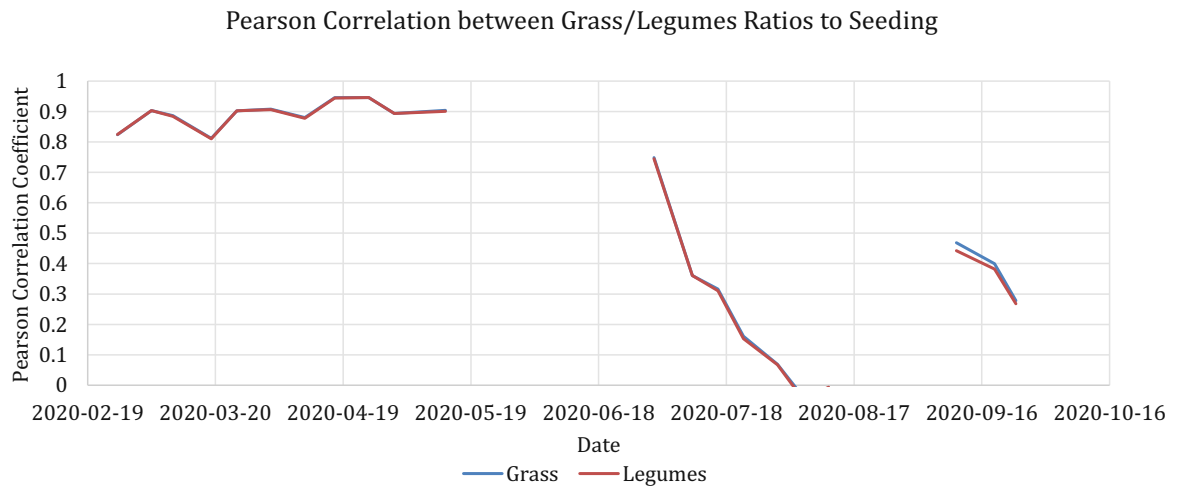


Figure 5.22: The Pearson correlation between the legumes-grass ratios at seeding and at the weekly measurements.

Figure 5.23 shows the correlation of the CNN estimated 2D grasses/legumes ratio to the weighted dry matter mass ratios during harvesting. As one would expect the correlation increases the closer the harvest date comes. Just before harvesting, the best correlation is expected as it is the measurement of the ratio from the image closest to the reference measurement by weighing the ratios.

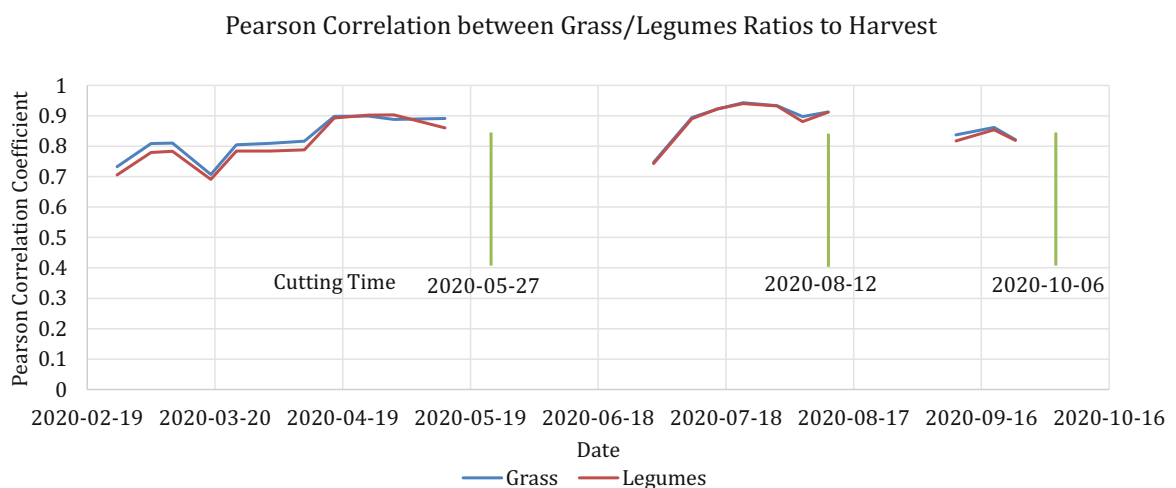


Figure 5.23: The Pearson correlation between the legumes-grass ratios at harvest and at the preceding weekly measurements.

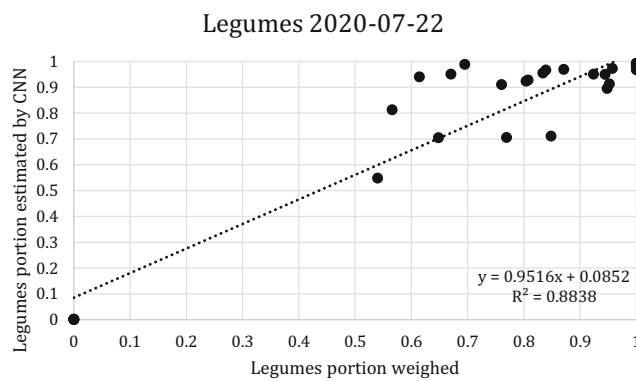


Figure 5.24: An example plot of the weighed and the predicted legumes portions for 2020-07-22.

The plot shown in Figure 5.24 is from one prediction day (2020-07-22) and allows for a deeper investigation of the prediction quality. Pure variants with only grasses or only legumes can be detected well. In-between values tend to be overestimated in the direction the more frequent plant type. For example, the proportion of legumes will be overestimated if they are more common during testing.

5.8.3 A yield estimation model

The different features estimated by the camera system are integrated into simple machine learning models. Details about the models are presented in Section 5.5. The models were evaluated with different train and test set splits and different feature combinations. Careful consideration should be taken as to whether the NDVI data is useful because it requires an additional NIR camera and should be left out if possible.

Figure 5.25 shows the results of the dry matter yield predictions based on the different machine learning models. The features used in the models are from the front camera. The machine learning methods include the support vector regression in the top left, the linear regression in the top right, the multi-layer perceptron regression in the bottom left and the k-nearest-neighbor regression in the bottom right. The x-axis shows the yield measured by the weighing scale (ground truth) and the y-axis displays the value predicted by the model. Training was repeated 100 times and during each trial five samples were randomly shifted to the test set. The SVM regression consistently achieved the best results.

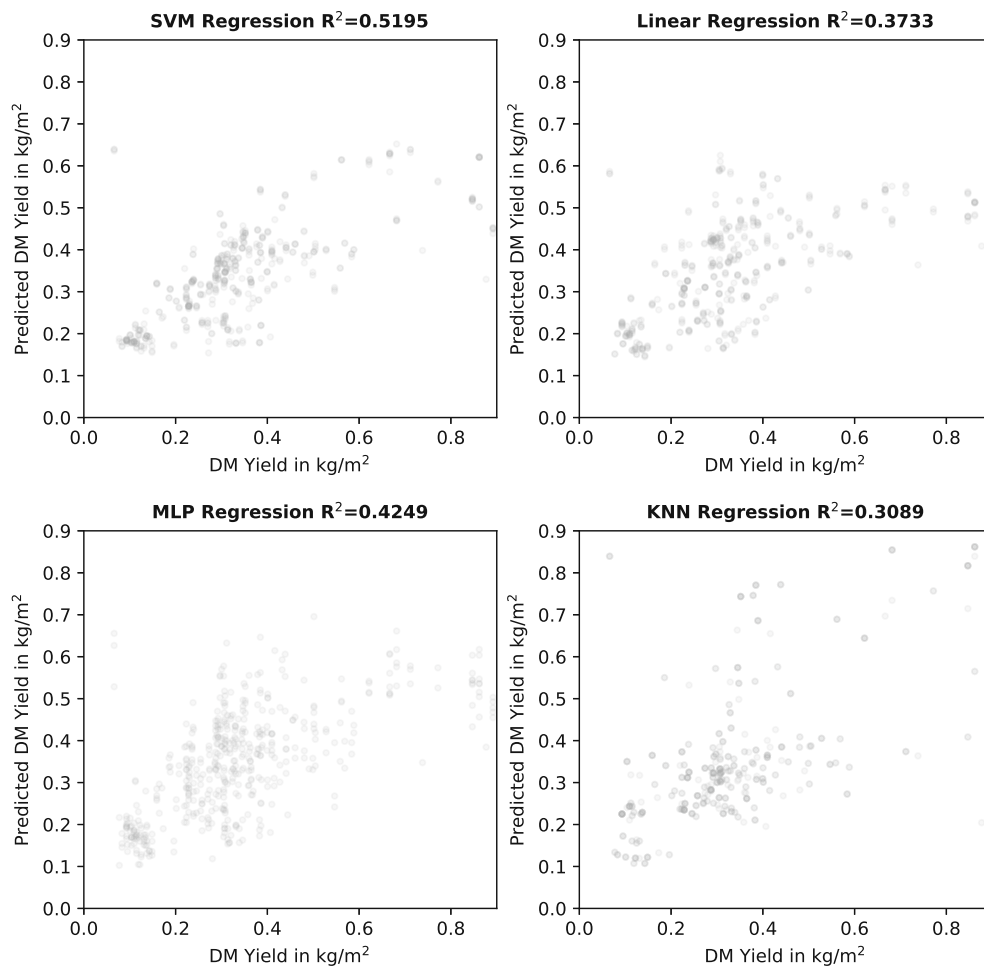


Figure 5.25: The predicted dry matter yields from the features estimated by the front facing camera. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis.

In Figure 5.26 the results of the SVM-Regression model without the feature NDVI are visible. The evaluation shows, that the model predicts the yield almost as well as in the previous model with NDVI (Figure 5.25).

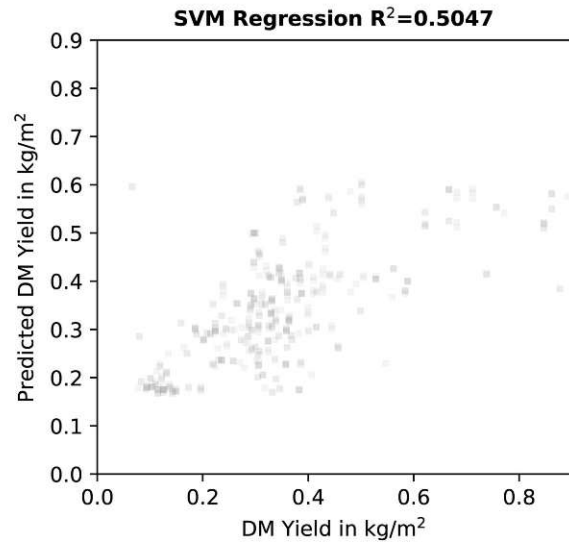


Figure 5.26: The predicted dry matter yields from the features estimated by the front facing camera without the NDVI. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis.

The model's predictive ability decreases significantly when only using the height values without the grass/legumes ratio information as shown in Figure 5.27.

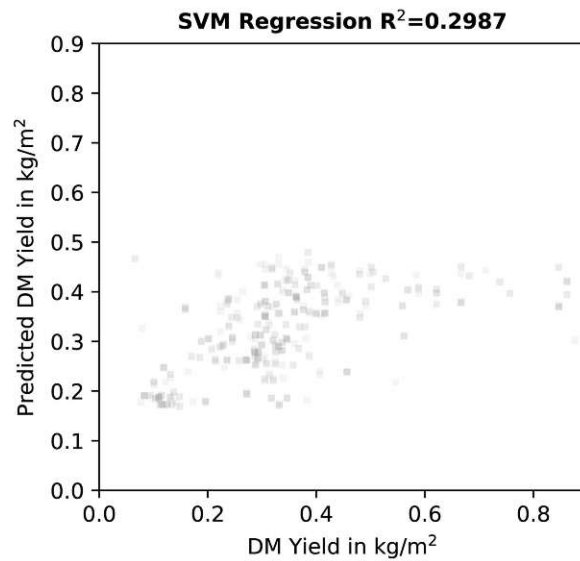


Figure 5.27: The predicted dry matter yields based solely on the height measurements of the front facing camera. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis.

The grass sward area visible in the rear facing camera is smaller. The predicted dry matter yields from the features estimated by the rear facing camera are visible in

Figure 5.28. The R^2 of the prediction equals 0.5241 for the SVM. This supports the assumption that a less visible sward area does not affect the results.

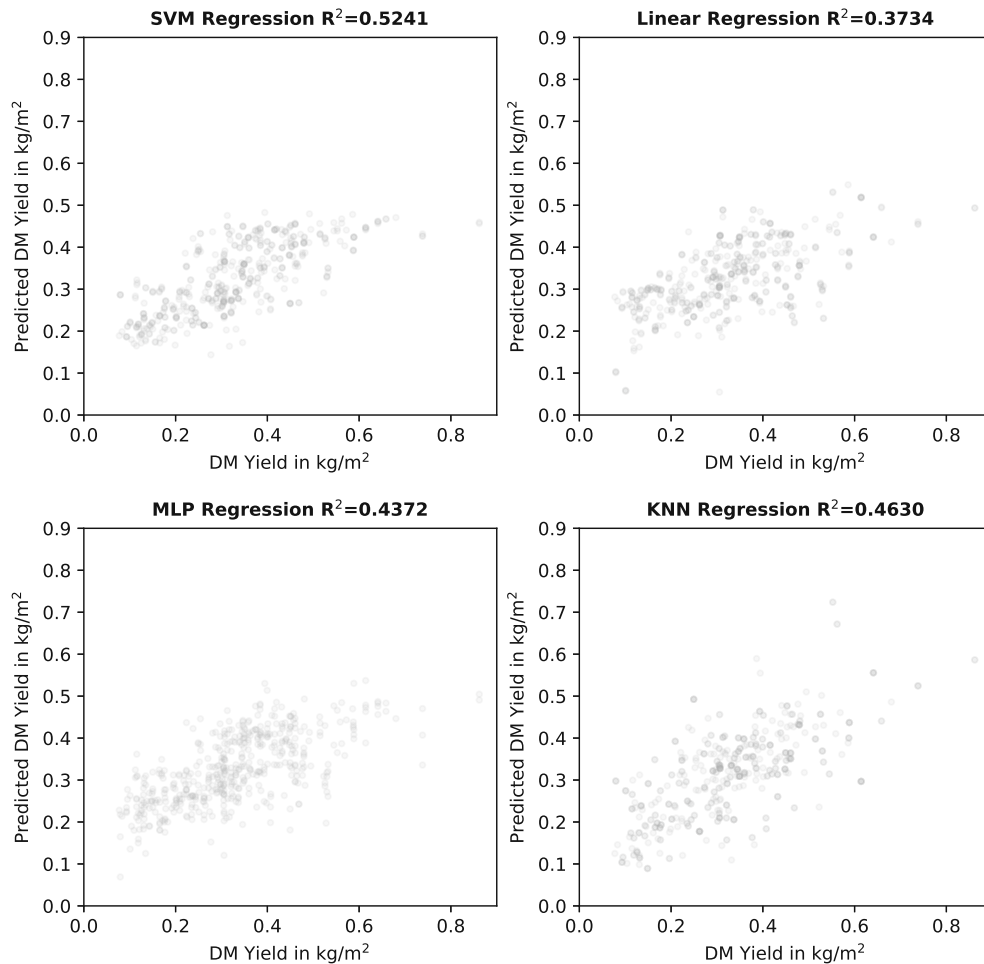


Figure 5.28: The predicted dry matter yields from the features estimated by the rear facing camera. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis.

The training was repeated 100 times for all the results shown above and during each trial 5 samples were randomly shifted to the test set. As a result, the field bias is not taken into account. As shown in Figure 5.20 there is a certain field bias in the data of the different trials (Table 5.1). Therefore, the training was repeated with randomly shifting entire field trials, from which all sample points were included, to the test set. This results in the prediction quality decreasing significantly, as shown in Figure 5.29 and the R^2 value decreasing towards 0.2933.

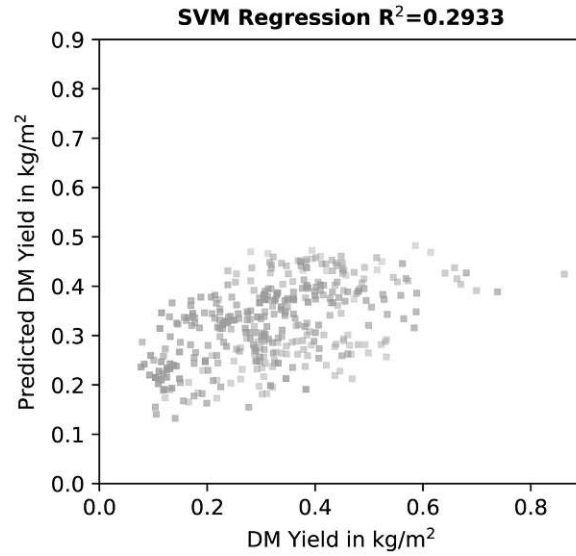


Figure 5.29: The predicted dry matter yields from the features estimated by the rear facing camera after shifting an entire field trials from the training set to the test set. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis.

Figure 5.30 shows the prediction results when only using the height values without the grass/legumes ratio information and without the NDVI data. These results were again based on a test set of randomly selected fields from which all sample points were included. The randomly selected fields were drawn from the entire field trial dataset. In this case R^2 drops in that case to 0.1771.

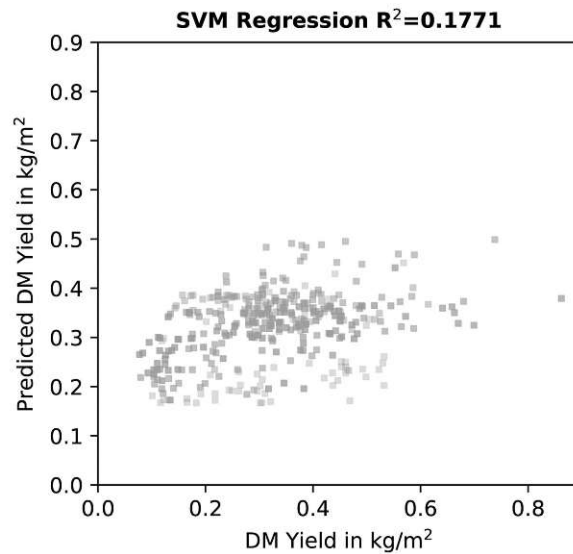


Figure 5.30: The predicted dry matter yields only based on height measurements estimated by the rear facing camera after shifting entire field trials from the training set to the test set. The ground truth points are listed on the x-axis and the predicted points are listed on the y-axis.

5.8.4 Swath Segmentation and Volume Estimation

For the evaluation of the swath segmentation, 39 test samples from different scenarios were evaluated based on their accuracy. The accuracy enables a comparison between the number of correctly classified pixels to the total number of pixels. The test samples are not included in the training data of the CNN segmentation.

This evaluation resulted in an accuracy of 0.965 for the segmentation based on the ERFNet CNN and an accuracy of 0.820 for the segmentation based on the depth data. Figure 5.31 shows some sample images from the test data. For swaths with low volume or a low height, as in example images 4 and 5, segmentations based on the depth data resulted in a larger error.



Figure 5.31: The swath segmentation results: the sample images followed by the manual labels, the segmentation results based on the depth data and the segmentation results based on the CNN

The aforementioned swath segmentation was integrated into the swath volume estimation. The results demonstrated that CNN was the superior segmentation method which led to it being adopted for all subsequent volume estimations. To evaluate the volume estimation method, the data from the field trials, listed in Table 5.2, was used. Random samples were taken on the field, weighed and their dry matter content was estimated. There is no efficient and accurate method to measure the swath volume besides methods involving cameras and lasers, therefore the estimated swath volume of the camera is compared to the yield mass. The results of the Pearson correlations between the estimated dry matter yield and the measured volumes of the swath are shown in Table 5.5. The swath mass can be predicted relatively consistently with a mean Pearson correlation of 0.6420.

Table 5.5: The correlation between the estimated volume by the camera and the weight yield mass for the rake field trials.

Date	Location	Pearson correlation between camera volume and weighed mass	Family/Genus composition
2018-07-31	St. Leonhard am Forst	0.5679	grass-clover hay
2018-08-22	Wieselburg	0.6475	grass-clover silage
2018-09-20	St. Leonhard am Forst	0.7845	lucerne and grass-clover hay
2019-05-08	Wieselburg	0.5873	grass-clover silage
2019-06-05	St. Leonhard am Forst	0.5746	grass-clover hay
2019-06-15	Wieselburg	0.6679	grass-clover silage
2019-07-23	Wieselburg	0.6640	grass-clover silage

Due to technical difficulties, the laser data from the LMS111 was only collected on 2018-07-31 and on 2018-09-20. The Pearson correlation between the mass and the volume measured by the laser was 0.7348 on 2018-07-31 and 0.7173 on 2018-09-20.

To test the accuracy of the swath volume estimation method, an outdoor laboratory test was conducted which is described in Section 5.7.1. A hay swath was manually placed on an area of mown grass, weighed and the volume was estimated with the camera system. Figure 5.32 illustrates the relationship between the volume and mass of the swath. The influence of the camera resolution was also investigated. Figure 5.32 (a) shows the volume estimation compared to mass based on a third of the original image resolution and Figure 5.32 (b) does the same based on a half of the original image resolution. These figures indicate that the error of the volume estimation increases as the resolution decreases. The linear relation between the volume estimated by the camera V_{camera} and the mass weighted m was $m = 19.06 \cdot V_{camera} + 0.25$ with $R^2 = 0.89$ for one third of the original resolution (533 x 400 pixels) and $m = 19.03 \cdot V_{camera} + 0.30$ with $R^2 = 0.95$ for the half resolution (800 x 600 pixels).

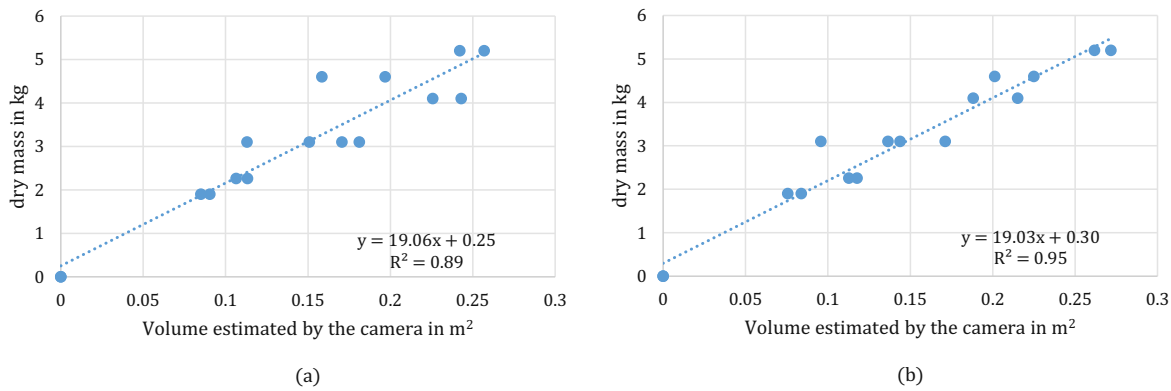


Figure 5.32: Results of the outdoor laboratory test for the swath volume measurement. The estimation with a third of the original image resolution (a) and the estimation with half of the original image resolution (b).

The swathing concept can influence the relationship between the swath volume and the swath mass. The experiment described in Section 5.7.2 was conducted to investigate this aspect. Figure 5.33 presents the results of 24 sample points of the experiment with a central delivery rake, a sun-wheel and a band-rake. The correlation between the estimated volume and the weighed swath mass was especially strong for the sun-wheel with an R^2 of 0.95. The measurements of the swath after the central delivery rake, as used in all other experiments, resulted in the lowest correlation of only $R^2 = 0.51$.

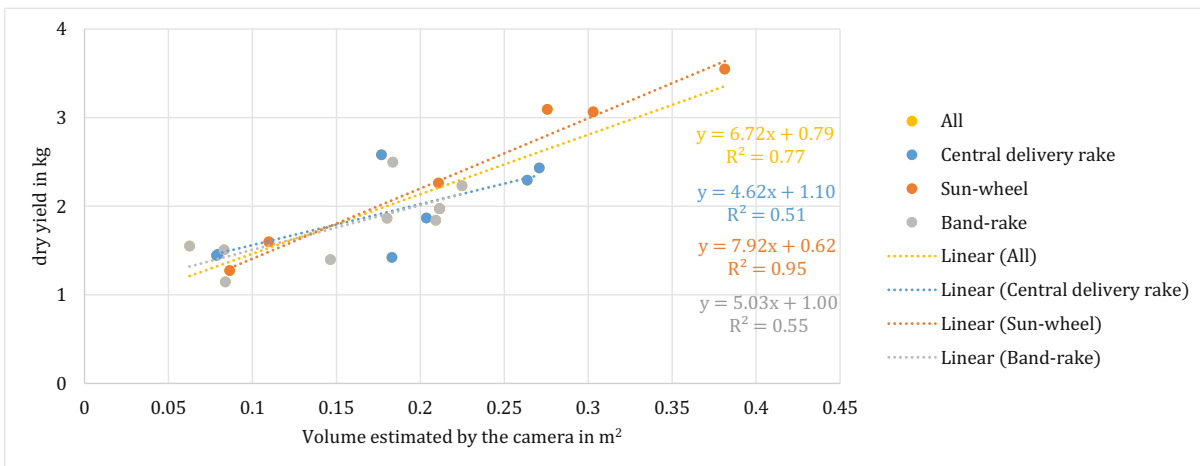


Figure 5.33: The relationship between the estimated swath volume and the weighed swath mass for the different rake concepts

5.8.5 Pre-training for Semantic Segmentation

During the pre-training of the ERFNet CNN, the encoder weights were transferred to the new model as initial weights. It is expected that the information stored in the

encoder features can be transferred to a new model to perform similar tasks. The CNN for the semantic segmentation is integrated into the grass/clover segmentation (Section 5.4), the swath segmentation (Section 5.6) and into the cut area segmentation (Section 5.3).

All three tasks were trained with different initial encoder weights. They were initialized with the weights of the semi-artificial grass clover image dataset [136], the sugar beet dataset [118] and with random weights (none).

The quality of the segmentation in Table 5.6 is defined by its accuracy and its IoU. The results show that pretraining improves the segmentation quality. The clover grass dataset especially improves the accuracy and the IoU.

Table 5.6: Model accuracy and IoU, with and without pre-training on the test datasets

<i>Application</i>	<i>Pre-training</i>	<i>Accuracy</i>	<i>IoU</i>
(grass/legumes ratio estimation)	sugar beet dataset	0.8522	0.5374
	clover grass dataset	0.8859	0.4480
	none	0.8462	0.5288
swath detection	sugar beet dataset	0.9653	0.9313
	clover grass dataset	0.9734	0.9470
	none	0.9604	0.9221
cut segmentation	sugar beet dataset	0.9106	0.7903
	clover grass dataset	0.9340	0.8312
	none	0.9286	0.8241

The inference time of all the models are stated in Table 3.3 of Chapter 3 for the soil cover estimation.

5.9 Discussion

In the following section the technical, environmental and agronomic aspects influencing the different measurements are discussed. The first subsection discusses the volume and yield estimation of the mower followed by the family and genus classification. Lastly, the swath segmentation and the swath volume estimation and the technical aspect of the pre-training on semantic segmentation are discussed.

5.9.1 Volume Estimation during Mowing and Yield Modelling

The yield estimation on the mower can only be evaluated using yield samples. Many different external influences can affect measuring from the grass sward height measurement to the grass/legumes ratio to the yield.

Yield sampling is one source of error which is influenced by the mowing width and sampling area, the GNSS Accuracy, the error of the weighing scale, wind forces on the grass bag hanging on the weighing scale and incorrect moisture samples. The mowing width and the sampling area error were reduced in the 2020 field trials by

using RTK based tractor steering and a hedge trimmer to cut the grass swath more precisely. However, it is believed that this still has the greatest influence on sampling. GNSS Accuracy hardly plays a role under the assumption that the local grassland is homogenous. The error of the weighing scale can be neglected by a mean sample size of 7 kg and a readable resolution of 20 g (0.29 %). The influence of the moisture samples is also considered to be low. A yield sample naturally dries over time. The moisture samples are taken less frequently than the yield samples and therefore are interpolated. The drying effect must be represented in the interpolation between the moisture samples.

Depending on the sensor type, there different errors affect the plant height measurement. Besides the known height measurement accuracies of the sensors, plant height measurement causes additional errors that must be taken into account. One difficulty is the geometry of grasses. Grasses are tall and narrow and are therefore difficult to capture by stereo matching from a short distance. This geometry also makes measuring distances with radar and ultrasound difficult due to the wide measuring lobes. The laser distance measurement is barely affected by this. The fundamental question of how many measurement points must be recorded on the surface of the stand remains unanswered and must be investigated in more precise laboratory tests.

Lodging is one of the main issues that influences the height measurements of all distance sensors (camera, radar, laser and ultrasonic). When the plants reach a certain height, they are pushed down by their own mass or by the wind and the rain. This distorts the plant height measurement very strongly and leads to outliers. Future work could introduce lodging as an additional class in cut segmentation. However, the difficulty lies in distinguishing between lodging and cut grass.

There are also features which cannot be captured by the sensors like plant and stem density and mass differences under the visible leaf surface. It was expected that the pendulum sensor could cover these aspects. However, this does not work robustly enough and needs to be investigated in more detail in the future.

The NDVI, which can be calculated by the data from our camera setup, saturates at higher soil cover values. The literature points out that the NDVI is not good at predicting grassland yields. Calculating different indices would require a more complex camera setup. However, the 850 nm NIR camera can be used for other applications in future.

Overall, we were able to predict yield during mowing with $R^2 = 0.5241$ with field bias and $R^2 = 0.2933$ over several fields. The work in [99] was able to predict dry matter yield with $R^2 = 0.75$ by using ultrasonic sensors and $R^2 = 0.82$ by combining ultrasonic sensors with data from a FieldSpec spectrometer (350 to 2500 nm). However, their field trial took place on one field where no field bias can occur and they increased yield variance by using different N-fertilizer applications. In general, higher differences in yields lead to a better ability to predict. The method in [141] predicted the biomass in homogeneous Italian ryegrass meadows with $R^2 = 0.56$. They calculated a multiple linear regression of two visible color indices. The work in [142] used hyper-spectral data in the range of 350-900 nm to predict grassland yield with $R^2 = 0.62$. The work in [142] used partial least squares regression (PLS) models.

5.9.2 Family and Genus Classification

The classification of the grass and the legumes in the grassland images resulted in an accuracy of 0.8522 and an IoU of 0.5374 for pre-training with the sugar beet dataset on the ERFNet.

As observed in the evaluation, the class values tend to be overestimated in the direction of the more prevalent plant type. This can be explained in part by the labelling process. Typically, the labelling begins with the less prevalent portion, e.g. with grasses in a lucerne field image. Afterwards, the image labelling is completed by filling the rest of the image with the class legumes. In that case, some grasses or darker regions get labelled as legumes. This results in a certain bias towards the more present class in the image and an S-curve in the prediction-ground truth plot.

This research confirmed the expected trend of a higher correlation of the harvest dry-mass to the images captured closer in time to the harvest. However, it was surprising that the correlation with the grass/legume content during seeding dropped so quickly.

Our work on grass/legumes segmentation can be most closely compared with the work in [101]. It achieved a prediction of the clover dry matter fraction with a standard deviation of 7.8%. This surpasses our results, although it should be noted that we do not have access to such homogeneous grasslands and our images were taken under strongly fluctuating light conditions. The effect of the different light conditions is visible in the images taken under low exposure times. Areas deeper in the grass sward are not exposed to enough light and appear black, especially with small clover plants. This effect could only be reduced to a limited extent by a gamma function in the image augmentation. The under-exposed areas (pixel value = 0) no longer carry any texture information. Therefore, the clover mass is still underestimated in the underexposed images.

5.9.3 Swath Segmentation and Volume Estimation

The segmentation of swaths with a CNN from an RGB image achieves an accuracy of 0.9734 and an IoU of 0.9470. Swaths with a low volume or a low height, as in the example images 4 and 5 in Figure 5.31, result in a larger error by segmentation based on the depth data. This can be explained by the lower depth resolution of the stereo setup compared to the image resolution. In general, the lower accuracy of the segmentation using the depth data can be explained by the lower depth resolution, the unevenness of the ground and errors in the plane estimation. Overall, it has been shown that the swath segmentation based on the depth data does not work as accurately as the segmentation based on the RGB image data.

When applying the CNN based segmentation on the point cloud to estimate the volume, a mean Pearson correlation of 0.6420 to the weighed swath mass was achieved. This indicates that either errors occur in the volume measurement or that the swath mass cannot be directly deduced from the volume. Uneven ground has a strong influence on the volume measurement because it can result in an inaccurate calculation of the ground plane.

The outdoor swath measurement laboratory test showed that the volume measurement

of the swath delivers strong correlations with the weighed mass. For these tests, the swath was formed by hand from pre-pressed hay balls which enabled for excluding the inaccuracies caused by raking. An even ground plane was used to avoid errors during ground plane estimation.

The tests to compare the different raking concepts showed similar correlations for central delivery rake and the band-rake. The sun-wheel produced strong volume-mass correlations which indicates that the volume-mass ratio is strongly influenced by the type of rake used. Future research should examine the effect of the different rakes in more detail with the specific aim of predicting the mass accurately even when using central delivery rakes.

Additionally, the swath volume estimation method was tested for the straw and oil pumpkin swath. However, the results were not shown in this thesis since they are thematically different. For the straw swath, image data was integrated into the existing grassland swath dataset. A strong generalization could be observed. Separate segmentation models were trained for oil pumpkins.

5.9.4 Pre-training for the Semantic Segmentation

The pre-training tests on the ERFNet showed that, in general, pre-training improves the accuracy of the model. The cover grass dataset generally led to better prediction accuracy compared to the sugar beet dataset. This can be attributed to the fact that the structures are very similar to grassland use cases. Pre-training utilizing the grass clover dataset was expected to improve the grass/legumes ratio estimation. However, the pre-training using the sugar beet dataset resulted in a better IoU.

Pre-training using the sugar beet dataset worsens the classification results for the cut segmentation. This might be attributed to ill-fitting features from the sugar beet dataset.

The linear and circular structures seem to be more similar to the clover grass dataset than to the sugar beet dataset for the swath segmentation and the cut segmentation. In general, both the swath segmentation and the cut segmentation resulted in higher accuracy and IoU. This can be explained by the fact that these applications are applied to the downscaled images and not to the image sections as with the grass/legumes segmentation. The annotations are not very accurate for the fine resolution images of the grass/legumes segmentation.

In general, all training image sets need to be extended to get more robust detections for the corner cases.

6 Conclusion and Outlook

Computer vision technologies offer a wide range of possibilities in agriculture. Farmers can gain many insights to make better decisions based on visual impressions. As agricultural machines become more and more automatized, they must be able to capture and interpret visual information. Key areas of importance include soil and plant parameters, soil cover, soil roughness, and yield estimation.

6.1 Summary

This thesis presents methods using a stereo camera system to measure soil cover and roughness, and to estimate the yield of grasslands. It demonstrates that these methods can be implemented in the field using a camera system with an integrated computing unit. The ERFNet CNN soil cover estimation method was more accurate in detecting dead organic matter than state of the art methods. An experiment with 10 people showed the large variance when manually labelling soil cover.

The method utilized to measure the soil roughness enables a tillage machine to be controlled in real time in order to produce a more homogeneous soil roughness during seeding on heterogeneous fields. A statistical comparison between the soil roughness and soil aggregate size showed that agronomic decisions can be inferred. As expected, the initial tests of plant emergence showed correlations with soil roughness.

The same camera systems can measure meadow height and estimate the grass/legumes ratio using a convolutional neural network. Additionally, a CNN based segmentation allows for the segmentation of grassland swaths and areas of cut grass. An investigation of the different parameters observed during mowing showed that the plant height and the plant composition serve as the basis to model dry matter yield in grasslands.

6.2 Outlook

In general, future research should focus on improving the real world datasets to gain insights into corner cases. This is true for all use cases, especially for those that use CNNs. Public datasets should also be adapted to represent the uncontrolled conditions of conventional fields like irregular lighting. Potential further steps are described in detail below.

6.2.1 Soil Cover Measurement

The soil cover measurement dataset could be expanded using automation. For this purpose, the NIR camera utilized in the grassland use case will be utilized to distinguish

automatically between the dead organic matter and the soil. This can reduce confusion between the dead organic matter and the soil, especially under direct sunlight.

In the future, the soil cover measurements could also be used to control a machine, as soil roughness based machine control on the power harrow. For example, information about soil cover could be used to control the ground pressure of tine harrows. It could also be investigated whether this method could be applied to an UAV. In particular, the UAV recording parameters like resolution and flight altitude must be investigated and the dataset must potentially be expanded.

Recent developments in the field of generative adversarial networks (GANs) have demonstrated their potential to augment images in agriculture [143]. All semantic segmentation applications in this thesis could benefit from this type of image augmentation.

6.2.2 Soil Roughness Measurement

In addition to measuring soil roughness, future research could also focus on measuring soil aggregates. This would enhance the direct agronomic benefit of the measurement as the aggregate size is a parameter more commonly used in agronomics. However, it is not entirely clear whether there is an actual benefit for the machine control system.

Further investigations could also examine the soil roughness and the plant emergence under different seeding rates. This line of research has hardly been investigated in the literature, but it certainly offers agronomic potential.

Finally, the algorithm could be integrated into cameras that do stereo matching independently. This will reduce hardware requirements by increasing the frame rate. It is expected that this will increase the agility and the precision of the power harrow control. The same is true, of course, for grassland methods that use stereo cameras.

6.2.3 Grassland Yield Measurement

Grassland composition estimation could be further improved by more extensive datasets, including new classes like rumex or thistle. This would not only increase the robustness, but also broaden the range of applications. When segmenting the cutting edge, an additional class lodging could be introduced. This would make the height measurement more robust. Additionally, future research should examine the density of grass swards because this may be a significant factor influencing the yield measurement.

Further research could also examine integrated applications besides yield mapping and grassland composition. For example, real time machine control could be used for swath forming, conditioner control or site-specific grassland reseeding.

This use case would especially benefit from an extended training set to improve the predictive quality for the grassland yields. Conditions vary depending on location, soil, weather, fertilization and species composition and can only be mapped through extensive field trials. Previous research has barely covered this heterogeneity in field trials.

Bibliography

- [1] ZAMG, *Starkniederschlag - Zukünftig mehr Starkregenereignisse?* <https://www.zamg.ac.at/cms/de/klima/informationsportal-klimawandel/standpunkt/klimazukunft/alpenraum/starkniederschlag>, Accessed: 2021-03-22.
- [2] H. Blanco-Canqui and R. Lal, „No-tillage and soil-profile carbon sequestration: An on-farm assessment,“ *Soil Science Society of America Journal*, vol. 72, no. 3, pp. 693–701, 2008.
- [3] „Economic and environmental evaluation of alternative pollution-reducing nitrogen management practices in central Illinois,“ *Agriculture, Ecosystems & Environment*, vol. 75, no. 1, pp. 41–53, 1999, ISSN: 0167-8809.
- [4] FAO, *The main principles of conservation agriculture*. <http://www.fao.org/conservation-agriculture/in-practice/soil-organic-cover>, Accessed: 2021-03-22, 2021.
- [5] M. Carter, „A review of conservation tillage strategies for humid temperate regions,“ *Soil and Tillage Research*, vol. 31, no. 4, pp. 289–301, 1994.
- [6] N. Fageria, V. Baligar, and B. Bailey, „Role of cover crops in improving soil and row crop productivity,“ *Communications in soil science and plant analysis*, vol. 36, no. 19-20, pp. 2733–2757, 2005.
- [7] K. M. Adam and D. C. Erbach, „Secondary Tillage Tool Effect on Soil Aggregation,“ *Transactions of the ASAE*, vol. 35, no. 6, pp. 1771–1776, 1992, ISSN: 0001-2351.
- [8] W. H. Wischmeier, D. D. Smith, *et al.*, „Predicting rainfall erosion losses-a guide to conservation planning,“ *Predicting rainfall erosion losses-a guide to conservation planning.*, 1978.
- [9] K. Köller, „Trends in seeding and mineral fertilizing,“ *LANDTECHNIK*, vol. 62, no. 6, pp. 382–383, 2007.
- [10] P. Strauss, D. Swoboda, and W. E. H. Blum, „How effective is mulching and minimum tillage to control runoff and soil loss? - a literature review,“ in *Proceedings of the International Conference*, vol. 25, 2003, pp. 545–550.
- [11] *OEPUL Programm*, <https://www.ama.at/Fachliche-Informationen/Oepul>, Accessed: 2021-04-26.
- [12] P. Riegler-Nurscher, J. Prankl, T. Bauer, P. Strauss, and H. Prankl, „A machine learning approach for pixel wise classification of residue and vegetation cover under field conditions,“ *Biosystems Engineering*, vol. 169, pp. 188 –198, 2018, ISSN: 1537-5110.

- [13] P. Riegler-Nurscher, J. Prankl, T. Bauer, P. Strauss, and H. Prankl, „An Integrated Image Analysis System for the Estimation of Soil Cover.“ in *International Conference on Agricultural Engineering CIGR - AgEng 2016*, Aarhus, Jun. 2016.
- [14] P. Riegler-Nurscher, J. Prankl, and M. Vincze, „Semantic Segmentation for the Estimation of Plant and Soil Parameters on Agricultural Machines.“ in *ICAS 2020, The Sixteenth International Conference on Autonomic and Autonomous Systems, Lisbon, Portugal, September 27 - October 1, 2020, Proceedings*, 2020, ISBN: 978-1-61208-787-0.
- [15] P. Riegler-Nurscher, J. Prankl, and M. Vincze, „Tillage Machine Control Based on a Vision System for Soil Roughness and Soil Cover Estimation.“ in *Computer Vision Systems, 12th International Conference, ICVS 2019, Thessaloniki, Greece, September 23-25, 2019, Proceedings*, D. Tzovaras, D. Giakoumis, M. Vincze, and A. A. Argyros, Eds., ser. Lecture Notes in Computer Science, vol. 11754, Springer, 2019, pp. 201–210, ISBN: 978-3-030-34994-3.
- [16] T. Bauer and P. Strauss, „A rule-based image analysis approach for calculating residues and vegetation cover under field conditions.“ *Catena*, vol. 113, pp. 363–369, 2014.
- [17] P. Riegler-Nurscher, G. Moitzi, J. Prankl, J. Huber, J. Karner, H. Wagentristl, and M. Vincze, „Machine vision for soil roughness measurement and control of tillage machines during seedbed preparation.“ *Soil and Tillage Research*, vol. 196, p. 104351, 2020, ISSN: 0167-1987.
- [18] P. Riegler-Nurscher, J. Prankl, T. Bauer, P. Strauss, and H. Prankl, „Robust online soil roughness measurement using stereo vision.“ in *EurAgEng, Vortrag: EurAgEng Conference 2018, Wageningen, Netherland; 2019-07-08 – 2019-07-12, 2018*.
- [19] P. Riegler-Nurscher, J. Karner, J. Huber, G. Moitzi, H. Wagentristl, M. Hofinger, and H. Prankl, „A system for online control of a rotary harrow using soil roughness detection based on stereo vision.“ in *Proceedings 75th Conference LAND. TECHNIK - AgEng 2017*, Hannover, Nov. 2017.
- [20] P. Riegler-Nurscher, J. Prankl, M. Hofinger, and M. Vincze, „Detektion eines Grünlandschwades mit Stereo RGB Kamera.“ in *40. GIL-Jahrestagung, Digitalisierung für Mensch, Umwelt und Tier.*, A. Meyer-Aurich, M. Gandorfer, and H. Floto, Eds., Bonn: Gesellschaft für Informatik e.V., 2020, pp. 265–270.
- [21] P. Riegler-Nurscher, J. Prankl, and J. Karner, „Beurteilung der Schnittqualität von Schneid- und Mähwerken mittels photogrammetrischer Methode.“ in *39. GIL-Jahrestagung, Digitalisierung für landwirtschaftliche Betriebe in kleinstrukturierten Regionen - ein Widerspruch in sich?*, Bonn: Gesellschaft für Informatik e.V., 2019, pp. 189–192.
- [22] P. Riegler-Nurscher, J. Prankl, M. Vincze, and H. Prankl, „A review on plant and soil properties estimation using ground-based computer vision techniques in the visible spectrum.“ in *Tagungsband LAND. TECHNIK 2020 VDI-Bericht 2374*, 2020, pp. 353–359, ISBN: 978-3-18-092374-1.

- [23] „Late fusion of multimodal deep neural networks for weeds classification,“ *Computers and Electronics in Agriculture*, vol. 175, p. 105–169, 2020, ISSN: 0168-1699.
- [24] T. Ruigrok, E. van Henten, J. Booij, K. van Boheemen, and G. Kootstra, „Application-Specific Evaluation of a Weed-Detection Algorithm for Plant-Specific Spraying,“ *Sensors*, vol. 20, no. 24, 2020, ISSN: 1424-8220.
- [25] J. Yu, Y. Li, C. Gao, H. Gao, G. Xia, Z. L. Yu, and Y. Li, „Exemplar-Based Recursive Instance Segmentation With Application to Plant Image Analysis,“ *IEEE Transactions on Image Processing*, vol. 29, pp. 389–404, 2020.
- [26] N. S. Monhollen, K. J. Shinnars, J. C. Friede, E. M. Rocha, and B. D. Luck, „In-field machine vision system for identifying corn kernel losses,“ *Computers and Electronics in Agriculture*, vol. 174, p. 105–169, 2020, ISSN: 0168-1699.
- [27] J. Ma, Y. Li, K. Du, F. Zheng, L. Zhang, Z. Gong, and W. Jiao, „Segmenting ears of winter wheat at flowering stage using digital images and deep learning,“ *Computers and Electronics in Agriculture*, vol. 168, p. 105–159, 2020, ISSN: 0168-1699.
- [28] C. Xu, H. Jiang, P. Yuen, K. Zaki Ahmad, and Y. Chen, „MHW-PD: A robust rice panicles counting algorithm based on deep learning and multi-scale hybrid window,“ *Computers and Electronics in Agriculture*, vol. 173, p. 105–175, 2020, ISSN: 0168-1699.
- [29] G. L. Andersen and T. Kuennen, „CULTI CAM HD - Efficient Weed Control in Row Crops using Active Implement Steering and Stereo Camera,“ in *VDI-Berichte Nr. 2361*, VDI, 2019, pp. 455–462.
- [30] Y. Li, H. Wang, L. M. Dang, A. Sadeghi-Niaraki, and H. Moon, „Crop pest recognition in natural scenes using convolutional neural networks,“ *Computers and Electronics in Agriculture*, vol. 169, p. 105–174, 2020, ISSN: 0168-1699.
- [31] A. N. Alves, W. S. Souza, and D. L. Borges, „Cotton pests classification in field-based images using deep residual networks,“ *Computers and Electronics in Agriculture*, vol. 174, p. 105–169, 2020, ISSN: 0168-1699.
- [32] H. Nazki, S. Yoon, A. Fuentes, and D. S. Park, „Unsupervised image translation using adversarial networks for improved plant disease recognition,“ *Computers and Electronics in Agriculture*, vol. 168, p. 105–117, 2020, ISSN: 0168-1699.
- [33] J. Lu, J. Hu, G. Zhao, F. Mei, and C. Zhang, „An In-field Automatic Wheat Disease Diagnosis System,“ *CoRR*, vol. abs/1710.08299, 2017. arXiv: 1710.08299.
- [34] L. Xu, C. Zheng, Z. Wang, and M. Nyongesah, „A digital camera as an alternative tool for estimating soil salinity and soil surface roughness,“ *Geoderma*, vol. 341, pp. 68–75, 2019, ISSN: 0016-7061.
- [35] A. Azizi, Y. A. Gilandeh, T. Mesri-Gundoshmian, A. A. Saleh-Bigdeli, and H. A. Moghaddam, „Classification of soil aggregates: A novel approach based on deep learning,“ *Soil and Tillage Research*, vol. 199, p. 104–169, 2020, ISSN: 0167-1987.

- [36] Q. Liu and V. Singh, „Effect of microtopography, slope length and gradient, and vegetative cover on overland flow through simulation,“ *Journal of Hydrologic Engineering*, vol. 9, no. 5, pp. 375–382, 2004.
- [37] A. Pandey, S. K. Himanshu, S. Mishra, and V. P. Singh, „Physically based soil erosion and sediment yield models revisited,“ *Catena*, vol. 147, pp. 595–620, 2016.
- [38] M. J. Marques, R. Bienes, L. Jiménez, and R. Pérez-Rodríguez, „Effect of vegetal cover on runoff and soil erosion under light intensity events. Rainfall simulation over USLE plots,“ *Science of the total environment*, vol. 378, no. 1-2, pp. 161–165, 2007.
- [39] A. G. Mohammad and M. A. Adam, „The impact of vegetative cover type on runoff and soil erosion under different land uses,“ *Catena*, vol. 81, no. 2, pp. 97–103, 2010.
- [40] S. Corak, T. Kaspar, and D. Meek, „Evaluating methods for measuring residue cover,“ *Journal of soil and water conservation*, vol. 48, no. 1, pp. 70–74, 1993.
- [41] R. Hartwig and J. Lafflen, „A meterstick method for measuring crop residue cover [in preventing soil erosion].,“ *Journal of Soil and Water Conservation*, vol. 33, pp. 90–91, 1978.
- [42] J. E. Morrison, C.-H. Huang, D. T. Lightle, and C. S. Daughtry, „Residue measurement techniques,“ *Journal of soil and water conservation*, vol. 48, no. 6, pp. 478–483, 1993.
- [43] H.-H. Voßhenrich, J. Brunotte, and B. Ortmeier, „Methoden zur Bewertung der Strohverteilung und Einarbeitung,“ *LANDTECHNIK–Agricultural Engineering*, vol. 58, no. 2, pp. 92–93, 2003.
- [44] A. Laliberte, D. Browning, J. Herrick, and P. Gronemeyer, „Hierarchical object-based classification of ultra-high-resolution digital mapping camera (DMC) imagery for rangeland mapping and assessment,“ *Journal of Spatial Science*, vol. 55, no. 1, pp. 101–115, 2010.
- [45] O. V. de Paul, „Remote sensing, surface residue cover and tillage practice,“ *Journal of Environmental Protection*, vol. 3, no. 02, p. 211, 2012.
- [46] F. Pforte, B. Wilhelm, and O. Hensel, „Evaluation of an online approach for determination of percentage residue cover,“ *Biosystems engineering*, vol. 112, no. 2, pp. 121–129, 2012.
- [47] D. Zhou, M. Li, Y. Li, J. Qi, K. Liu, X. Cong, and X. Tian, „Detection of ground straw coverage under conservation tillage based on deep learning,“ *Computers and Electronics in Agriculture*, vol. 172, p. 105369, 2020, ISSN: 0168-1699.
- [48] T Behrens and W Diepenbrock, „Using digital image analysis to describe canopies of winter oilseed rape (*Brassica napus* L.) during vegetative developmental stages,“ *Journal of agronomy and crop science*, vol. 192, no. 4, pp. 295–302, 2006.

- [49] L. T. Bennett, T. S. Judd, and M. A. Adams, „Close-range vertical photography for measuring cover changes in perennial grasslands,“ *Journal of Range Management*, pp. 634–641, 2000.
- [50] D Terrance Booth, S. E. Cox, C. Fifield, M. Phillips, and N. Williamson, „Image analysis compared with other methods for measuring ground cover,“ *Arid Land Research and Management*, vol. 19, no. 2, pp. 91–100, 2005.
- [51] C Campillo, M. Prieto, C Daza, M. Monino, and M. Garcia, „Using digital images to characterize canopy coverage and light interception in a processing tomato crop,“ *Hortscience*, vol. 43, no. 6, pp. 1780–1786, 2008.
- [52] L. C. Purcell, „Soybean canopy coverage and light interception measurements using digital imagery,“ *Crop Science*, vol. 40, no. 3, pp. 834–837, 2000.
- [53] M. Kırıcı, E. O. Güneş, Y. Çakır, and S. Şentiirk, „Vegetation measurement using image processing methods,“ in *Agro-geoinformatics (Agro-geoinformatics 2014), Third International Conference on, IEEE, 2014*, pp. 1–5.
- [54] J. Hu, M. Dai, and S. Peng, „An automated (novel) algorithm for estimating green vegetation cover fraction from digital image: UIP-MGMEP,“ *Environmental monitoring and assessment*, vol. 190, no. 11, p. 687, 2018.
- [55] G García-Mateos, J. Hernández-Hernández, D Escarabajal-Henarejos, S Jaen-Terrones, and J. Molina-Martínez, „Study and comparision of color models for automatic image analysis in irrigation management applications,“ *Agricultural water management*, vol. 151, pp. 158–166, 2015.
- [56] J. D. Luscier, W. L. Thompson, J. M. Wilson, B. E. Gorham, and L. D. Dragut, „Using digital photographs and object-based image analysis to estimate percent ground cover in vegetation plots,“ *Frontiers in Ecology and the Environment*, vol. 4, no. 8, pp. 408–413, 2006.
- [57] F. Pérez-Cabello, A. Cerdá, J De La Riva, M. Echeverría, A. García-Martín, P Ibarra, T Lasanta, R Montorio, and V. Palacios, „Micro-scale post-fire surface cover changes monitored using high spatial resolution photography in a semiarid environment: A useful tool in the study of post-fire soil erosion processes,“ *Journal of Arid Environments*, vol. 76, pp. 88–96, 2012.
- [58] 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*, vol. 125, pp. 184–199, 2016.
- [59] X. Bai, Z. Cao, Y. Wang, Z. Yu, Z. Hu, X. Zhang, and C. Li, „Vegetation segmentation robust to illumination variations based on clustering and morphology modelling,“ *Biosystems engineering*, vol. 125, pp. 80–97, 2014.
- [60] 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*, vol. 39, no. 12, pp. 11 149–11 155, 2012.

- [61] A. K. Mortensen, M. Dyrmann, H. Karstoft, R. N. Jørgensen, R. Gislum, *et al.*, „Semantic segmentation of mixed crops using deep convolutional neural network.“ in *CIGR-AgEng Conference, 26-29 June 2016, Aarhus, Denmark. Abstracts and Full papers*, Organising Committee, CIGR 2016, 2016, pp. 1–6.
- [62] M. Braunack and A. Dexter, „Soil aggregation in the seedbed: A review. I. Properties of aggregates and beds of aggregates.“ *Soil and Tillage Research*, vol. 14, no. 3, pp. 259–279, 1989, ISSN: 0167-1987.
- [63] J Lipiec, R Walczak, B Witkowska-Walczak, A Nosalewicz, A Słowińska-Jurkiewicz, and C Sławiński, „The effect of aggregate size on water retention and pore structure of two silt loam soils of different genesis.“ *Soil and Tillage Research*, vol. 97, no. 2, pp. 239–246, 2007, ISSN: 0167-1987.
- [64] M. Braunack, „Effect of aggregate size and soil water content on emergence of soybean (*Glycine max*, L. Merr.) and maize (*Zea mays*, L.),“ *Soil and Tillage Research*, vol. 33, no. 3, pp. 149–161, 1995, ISSN: 0167-1987.
- [65] J Guérif, G Richard, C Dürr, J. Machet, S Recous, and J Roger-Estrade, „A review of tillage effects on crop residue management, seedbed conditions and seedling establishment.“ *Soil and Tillage Research*, vol. 61, no. 1, pp. 13–32, 2001, XVth ISTRO Conference on Tillage at the Threshold of the 21st Century: Looking Ahead, ISSN: 0167-1987.
- [66] I. Håkansson, Åsa Myrbeck, and A. Etana, „A review of research on seedbed preparation for small grains in Sweden.“ *Soil and Tillage Research*, vol. 64, no. 1, pp. 23–40, 2002, Soil Fragmentation and Seedbed Characterization, ISSN: 0167-1987.
- [67] M Cadena-Zapata, W. Hoogmoed, and U. Perdok, „Field studies to assess the workable range of soils in the tropical zone of Veracruz, Mexico.“ *Soil and Tillage Research*, vol. 68, no. 2, pp. 83–92, 2002, ISSN: 0167-1987.
- [68] M. Braunack and J. McPhee, „The effect of initial soil water content and tillage implement on seedbed formation.“ *Soil and Tillage Research*, vol. 20, no. 1, pp. 5–17, 1991, ISSN: 0167-1987.
- [69] H. Nasr and F. Selles, „Seedling emergence as influenced by aggregate size, bulk density, and penetration resistance of the seedbed.“ *Soil and Tillage Research*, vol. 34, no. 1, pp. 61–76, 1995, ISSN: 0167-1987.
- [70] L. Ye, W. Tan, L. Fang, L. Ji, and H. Deng, „Spatial analysis of soil aggregate stability in a small catchment of the Loess Plateau, China: I. Spatial variability.“ *Soil and Tillage Research*, vol. 179, pp. 71–81, 2018, ISSN: 0167-1987.
- [71] K. Juhos, S. Czigány, B. Madarász, and M. Ladányi, „Interpretation of soil quality indicators for land suitability assessment ? A multivariate approach for Central European arable soils.“ *Ecological Indicators*, vol. 99, pp. 261–272, 2019, ISSN: 1470-160X.
- [72] R. R. Allmaras, R. E. Burwell, W. E. Larson, R. F. Holt, and W. W. Nelson, „Total Porosity and random roughness of the interrow zone.“ pp. 1–22, U.S. Department of Agriculture Conservation Research Report No. 7 1966.

- [73] R. Turner, R. Panciera, M. A. Tanase, K. Lowell, J. M. Hacker, and J. P. Walker, „Estimation of soil surface roughness of agricultural soils using airborne LiDAR,“ *Remote Sensing of Environment*, vol. 140, pp. 107–117, 2014, ISSN: 0034-4257.
- [74] M. Aguilar, F. Aguilar, and J. Negreiros, „Off-the-shelf laser scanning and close-range digital photogrammetry for measuring agricultural soils microrelief,“ *Biosystems Engineering*, vol. 103, no. 4, pp. 504–517, 2009, ISSN: 1537-5110.
- [75] L. Zhixiong, C. Nan, U. Perdok, and W. Hoogmoed, „Characterisation of Soil Profile Roughness,“ *Biosystems Engineering*, vol. 91, no. 3, pp. 369–377, 2005, ISSN: 1537-5110.
- [76] R. Sandri, T. Anken, T. Hilfiker, L. Sartori, and H. Bollhalder, „Comparison of methods for determining cloddiness in seedbed preparation,“ *Soil and Tillage Research*, vol. 45, no. 1, pp. 75–90, 1998, ISSN: 0167-1987.
- [77] F. Marinello, A. Pezzuolo, F. Gasparini, J. Arvidsson, and L. Sartori, „Application of the Kinect sensor for dynamic soil surface characterization,“ *Precision Agriculture*, vol. 16, no. 6, pp. 601–612, 2015, ISSN: 1573-1618.
- [78] J. Gilliot, E. Vaudour, and J. Michelin, „Soil surface roughness measurement: A new fully automatic photogrammetric approach applied to agricultural bare fields,“ *Computers and Electronics in Agriculture*, vol. 134, pp. 63–78, 2017, ISSN: 0168-1699.
- [79] T. Riegler, C. Rechberger, F. Handler, and H. Prankl, „Image processing system for evaluation of tillage quality,“ *Landtechnik*, vol. 69, no. 3, pp. 125–131, 2014.
- [80] O. Taconet and V. Ciarletti, „Estimating soil roughness indices on a ridge-and-furrow surface using stereo photogrammetry,“ *Soil and Tillage Research*, vol. 93, no. 1, pp. 64–76, 2007, ISSN: 0167-1987.
- [81] J. Shan and C. K. Toth, *Topographic laser ranging and scanning: principles and processing*. CRC press, 2018.
- [82] T. Bauer, P. Strauss, M. Grims, E. Kamptner, R. Mansberger, and H. Spiegel, „Long-term agricultural management effects on surface roughness and consolidation of soils,“ *Soil and Tillage Research*, vol. 151, pp. 28–38, 2015, ISSN: 0167-1987.
- [83] G. L. Andersen, „CULTI CAM HD: Efficient Weed Control in Row Crops using Active Implement Steering and Stereo Camera,“ in *Tagungsband LAND. TECHNIK 2019 VDI-Bericht 2361*, 2019, pp. 455–461, ISBN: 978-3-18-092361-1.
- [84] O. Chimi-Chiadjeu, S. L. Hégarat-Masclé, E. Vannier, O. Taconet, and R. Dusséaux, „Automatic clod detection and boundary estimation from Digital Elevation Model images using different approaches,“ *CATENA*, vol. 118, pp. 73–83, 2014, ISSN: 0341-8162.
- [85] M. Demmel, T. Schwenke, H. Heuwinkel, F. Locher, and J. Rottmeier, „Ertragsermittlung von Grünland—erste Ergebnisse,“ *LANDTECHNIK—Agricultural Engineering*, vol. 57, no. 3, pp. 146–147, 2002.

- [86] K. Wild, S. Ruhland, and S. Haedicke, „Local Yield Detection of Grass in a Mower Conditioner,“ in *2005 ASAE Annual Meeting*, American Society of Agricultural and Biological Engineers, 2005, p. 1.
- [87] P Miklenda, F Kumhála, and V Prošek, „Feed rate measurement technique and yield maps creating in fodder plant harvesting,“ *Agr. Eng. Res.*, vol. 52, no. 4, pp. 123–128, 2006.
- [88] F. Kumhála, M. Kroulík, and V. Prošek, „Development and evaluation of forage yield measure sensors in a mowing-conditioning machine,“ *Computers and electronics in agriculture*, vol. 58, no. 2, pp. 154–163, 2007.
- [89] V. C. Hammen and D. Ehlert, „Online-Ertragsmessung in Klee gras mit dem Pendulum Meter,“ *LANDTECHNIK–Agricultural Engineering*, vol. 54, no. 3, pp. 156–163, 1999.
- [90] D. Ehlert, U. Völker, and K. Dammer, „Pendulum meter in practical operation,“ *Landtechnik*, vol. 58, no. 1, pp. 16–17, 2003.
- [91] D Ehlert, V Hammen, and R Adamek, „Online sensor pendulum-meter for determination of plant mass,“ *Precision Agriculture*, vol. 4, no. 2, pp. 139–148, 2003.
- [92] Y. KAIZU, J.-M. CHOI, and T.-H. KANG, „Grass height and yield estimation using a three-dimensional laser scanner,“ *Environmental Control in Biology*, vol. 50, no. 1, pp. 41–51, 2012.
- [93] H. Safari, T. Fricke, B. Reddersen, T. Möckel, and M. Wachendorf, „Comparing mobile and static assessment of biomass in heterogeneous grassland with a multi-sensor system,“ *Journal of Sensors and Sensor Systems*, vol. 5, no. 2, pp. 301–312, 2016.
- [94] T. Fricke, F. Richter, and M. Wachendorf, „Assessment of forage mass from grassland swards by height measurement using an ultrasonic sensor,“ *Computers and electronics in agriculture*, vol. 79, no. 2, pp. 142–152, 2011.
- [95] B. Reddersen, T. Fricke, and M. Wachendorf, „A multi-sensor approach for predicting biomass of extensively managed grassland,“ *Computers and electronics in agriculture*, vol. 109, pp. 247–260, 2014.
- [96] S. Biewer, T. Fricke, and M. Wachendorf, „Determination of dry matter yield from legume–grass swards by field spectroscopy,“ *Crop science*, vol. 49, no. 5, pp. 1927–1936, 2009.
- [97] A. Schaumberger, J. Schellberg, and J. Hollberg, „Spektrometrie–moderne, nicht-invasive Erfassungstechnik von Pflanzenbeständen,“ *Bericht zum*, vol. 20, p. 23, 2015.
- [98] J. M. Serrano, S. Shahidian, and J. R. M. Da Silva, „Monitoring pasture variability: optical OptRx® crop sensor versus Grassmaster II capacitance probe,“ *Environmental monitoring and assessment*, vol. 188, no. 2, p. 117, 2016.

- [99] T. Fricke and M. Wachendorf, „Combining ultrasonic sward height and spectral signatures to assess the biomass of legume–grass swards,“ *Computers and electronics in agriculture*, vol. 99, pp. 236–247, 2013.
- [100] Wachendorf, Ingwersen, and Taube, „Prediction of the clover content of red clover and white clover-grass mixtures by near-infrared reflectance spectroscopy,“ *Grass and Forage Science*, vol. 54, no. 1, pp. 87–90, 1999.
- [101] S. Skovsen, M. Dyrmann, A. Mortensen, K. Steen, O. Green, J. Eriksen, R. Gislum, R. Jørgensen, and H. Karstoft, „Estimation of the Botanical Composition of Clover-Grass Leys from RGB Images Using Data Simulation and Fully Convolutional Neural Networks,“ *Sensors*, vol. 17, no. 12, p. 2930, 2017, ISSN: 1424-8220.
- [102] J. Schellberg, M. J. Hill, R. Gerhards, M. Rothmund, and M. Braun, „Precision agriculture on grassland: Applications, perspectives and constraints,“ *European Journal of Agronomy*, vol. 29, no. 2-3, pp. 59–71, 2008.
- [103] R. R. Pullanagari, R. A. Dynes, W. M. King, I. J. Yule, S. Thulin, N. M. Knox, A. Ramoelo, D Michalk, G Millar, and W Badgery, „Remote sensing of pasture quality,“ in *Proc. 22nd International Grasslands Congress 2013*, 2013, p. 6.
- [104] Sick AG, *WGS - Fahrerassistenz zur Erkennung und Vermessung einer Schwade*. https://cdn.sick.com/media/docs/8/18/818/Product_information_WGS_de_IM0068818.PDF, Accessed: 2021-03-22.
- [105] M. R. Blas and M. Blanke, „Stereo vision with texture learning for fault-tolerant automatic baling,“ *Computers and Electronics in Agriculture*, vol. 75, no. 1, pp. 159–168, 2011, ISSN: 0168-1699.
- [106] Y. Yang, S. Hallman, D. Ramanan, and C. C. Fowlkes, „Layered Object Models for Image Segmentation,“ *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1731–1743, 2012.
- [107] A. Montillo, J. Shotton, J. Winn, J. E. Iglesias, D. Metaxas, and A. Criminisi, „Entangled decision forests and their application for semantic segmentation of CT images,“ in *Biennial International Conference on Information Processing in Medical Imaging*, Springer, 2011, pp. 184–196.
- [108] V. Badrinarayanan, A. Kendall, and R. Cipolla, *SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation*, 2016. arXiv: 1511.00561 [cs.CV].
- [109] I. Ulku and E. Akagunduz, „A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D images,“ *arXiv preprint arXiv:1912.10230*, 2019.
- [110] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*, 2017. arXiv: 1606.00915 [cs.CV].
- [111] F. Lateef and Y. Ruichek, „Survey on semantic segmentation using deep learning techniques,“ *Neurocomputing*, vol. 338, pp. 321–348, 2019, ISSN: 0925-2312.

- [112] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, „Erfnet: Efficient residual factorized convnet for real-time semantic segmentation,“ *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 263–272, 2017.
- [113] S. Mehta, M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi, *ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation*, 2018. arXiv: 1803.06815 [cs.CV].
- [114] F. Handler, C. Rechberger, M. Nadlinger, P. Riegler-Nurscher, and L. Weninger, „Grubberwalzen: Wie groß sind die Unterschiede?,“ vol. 2/2019, pp. 68–73, 2020.
- [115] F. Handler and P. Riegler-Nurscher, „Wie wirkt die Bodenbearbeitung auf den Erosionsschutz?,“ vol. 2/2020, pp. 52–53, 2020.
- [116] A. Milioto and C. Stachniss, „Bonnet: An Open-Source Training and Deployment Framework for Semantic Segmentation in Robotics using CNNs,“ in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 7094–7100.
- [117] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, *et al.*, „SLIC superpixels compared to state-of-the-art superpixel methods,“ *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [118] N. Chebrolu, P. Lottes, A. Schaefer, W. Winterhalter, W. Burgard, and C. Stachniss, „Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields,“ *The International Journal of Robotics Research*, 2017.
- [119] A. Criminisi and J. Shotton, *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media, 2013.
- [120] J. Shotton, J. Winn, C. Rother, and A. Criminisi, „Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context,“ *International Journal of Computer Vision*, vol. 81, no. 1, pp. 2–23, 2009.
- [121] J. Long, E. Shelhamer, and T. Darrell, „Fully convolutional networks for semantic segmentation,“ in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.
- [122] K. He, X. Zhang, S. Ren, and J. Sun, „Deep Residual Learning for Image Recognition,“ *CoRR*, vol. abs/1512.03385, 2015. arXiv: 1512.03385.
- [123] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, 2015. arXiv: 1505.04597 [cs.CV].
- [124] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, *ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation*, 2016. arXiv: 1606.02147 [cs.CV].
- [125] C. Shorten and T. M. Khoshgoftaar, „A survey on Image Data Augmentation for Deep Learning,“ *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019, ISSN: 2196-1115.

- [126] P. M. Blok, F. K. van Evert, A. P. M. Tielen, E. J. van Henten, and G. Kootstra, „The effect of data augmentation and network simplification on the image-based detection of broccoli heads with Mask R-CNN,“ *Journal of Field Robotics*, vol. 38, no. 1, pp. 85–104, 2021.
- [127] J. Lafflen, M. Amemiya, and E. Hintz, „Measuring crop residue cover,“ *Journal of soil and Water Conservation*, vol. 36, no. 6, pp. 341–343, 1981.
- [128] J. Schönhammer, „Der Arbeitseffekt zapfwellengetriebener Bodenbearbeitungsgeräte,“ PhD thesis, Technische Universität München, 1982.
- [129] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press., 2003.
- [130] *OpenCV library*, <http://opencv.org/>, Accessed: 2021-03-22.
- [131] J. Eitzinger, V. Daneu, G. Bodner, G. Kubu, W. Loiskandl, P. Macaigne, S. Thaler, A. Schaumberger, C. Wittmann, F. Weidle, A. Kann, E. Murer, C. Krammer, M. Trnka, and M. Hayes, *Drought monitoring system for Austrian agriculture - AgroDroughtAustria. Final Project Report*. 2016. [Online]. Available: https://meteo.boku.ac.at/report/BOKU-Met_Report_25_online.pdf.
- [132] J. Schönhammer, „Der Arbeitseffekt zapfwellengetriebener Bodenbearbeitungsgeräte,“ PhD thesis, Technische Universität München, 1982.
- [133] J. Kouwenhoven, „Effect of mouldboard and chisel ploughing and their timing on seedbed quality,“ *Soil and Tillage Research*, vol. 7, no. 3, pp. 205–219, 1986, ISSN: 0167-1987.
- [134] H. Kirchmeier, R. Geischeder, and M. Demmel, „Bearbeitungseffekt und Leistungsbedarf von Kreiseleggen mit unterschiedlichen Kreiselgeometrien,“ *LANDTECHNIK–Agricultural Engineering*, vol. 60, no. 4, pp. 196–197, 2005.
- [135] *MSO RaDist Multi-Target sensor*, <https://www.mso-technik.de/mso-produkte/abstandsmessung/radist.html>, Accessed: 2021-03-22.
- [136] S. Skovsen, M. Dyrmann, A. K. Mortensen, M. S. Laursen, R. Gislum, J. Erikson, S. Farkhani, H. Karstoft, and R. N. Jorgensen, „The GrassClover Image Dataset for Semantic and Hierarchical Species Understanding in Agriculture,“ in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.
- [137] *scikit-learn library*, <https://scikit-learn.org/>, Accessed: 2021-03-22.
- [138] D. Kingma and J. Ba, „Adam: A Method for Stochastic Optimization,“ *International Conference on Learning Representations*, Dec. 2014.
- [139] Open Perception Foundation, *PCL - Point Cloud Library*, <http://www.pointclouds.org/>, Accessed: 2021-03-22.
- [140] *Boost libraries*, <http://www.boost.org/>, Accessed: 2021-03-22.
- [141] K. Kawamura, Y. Sakuno, Y. Tanaka, H.-J. Lee, J. Lim, Y. Kurokawa, and N. Watanabe, „Mapping herbage biomass and nitrogen status in an Italian ryegrass (*Lolium multiflorum* L.) field using a digital video camera with balloon system,“ *Journal of Applied Remote Sensing*, vol. 5, pp. 3562–, Nov. 2011.

- [142] F. J. Ancin-Murguzur, G. Taff, C. Davids, H. Tømmervik, J. Mølmann, and M. Jørgensen, „Yield Estimates by a Two-Step Approach Using Hyperspectral Methods in Grasslands at High Latitudes,“ *Remote Sensing*, vol. 11, no. 4, 2019, ISSN: 2072-4292.
- [143] H. Nazki, S. Yoon, A. Fuentes, and D. S. Park, „Unsupervised image translation using adversarial networks for improved plant disease recognition,“ *Computers and Electronics in Agriculture*, vol. 168, p. 105 117, 2020.