Deep learning and hyperspectral imaging for unmanned aerial vehicles
Dijkstra, Klaas

DOI:
10.33612/diss.131754011

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Recently there has been much interest in hyperspectral imaging research and applications. Hyperspectral cameras collect image data from across the electromagnetic spectrum. These cameras aim to capture a spectrogram for each pixel to form a hyperspectral cube. The application area for these types of cameras is broad and varies from vegetation inspection to chemical finger printing.

Small battery-powered multi-copters called Unmanned Aerial Vehicles (UAVs) have recently shown great potential for a large number of applications. A camera mounted on a UAV is used for the inspection of large objects or large areas. For example, in the field of precision agriculture they are used for measuring Chlorophyll to determine crop health. Furthermore, UAVs are used for the inspection of wind turbines, to take images of cracks, pinholes and other defects.

UAV applications could benefit more from hyperspectral imaging technology, but these devices have intrinsic limitations that makes using them in conjunction with each other challenging. This is mainly caused by the sensitivity to movement or low spatial resolution of the hyperspectral devices combined with the limited payload capabilities of UAVs.
Deep learning or, more specifically, Convolutional Neural Networks (CNNs) have shown to obtain state-of-the-art performance in a multitude of research fields. Many applications already benefit from deep learning. Usually these deep learning models are trained directly from data in an end-to-end fashion. This dissertation revolves around the question if algorithms from the field of deep learning can mitigate the difficulties that are caused by the limitations encountered in combining hyperspectral imaging and UAVs. This trinity of technologies: deep learning, hyperspectral imaging and UAVs, serves as a framework within which this research is defined.

In Chapter 1 an introduction into the topics that are addressed in this dissertation is given. That chapter gives an overview and analysis of the research questions and proposes several areas within the trinity of technologies on which this research is performed.

In Chapter 2 a hyperspectral Liquid Crystal Tunable Filter (LCFT) system of 28 bands is used to distinguish between Alternaria infections and Ozone (O₃) damage on Potato-plant leaves. If a small subset of spectral bands can be found that is still able to distinguish between diseases it is more feasible to develop a multi-camera system that can be used on a UAV. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) feature extraction and features selection techniques are used to select a subset of spectral bands. This is combined with classification experiments to investigate the performance differences between Multi Layer Perceptron (MLP), Support Vector Machine (SVM), k-Nearest Neighbor (kNN) and Gaussian-model classifiers. It was found that using a per-pixel classification, the disease on the leaf could be identified with good performance that depends on the type of feature extraction, feature selection and the type of classifier.

In Chapter 3 several novel architectures for crosstalk correction and demosaicking of hyperspectral cubes are discussed. A sensor with a Multispectral Color Filter Array (MCFA) contains a mosaic of small optical filters to produce a hyperspectral image. Crosstalk between spectral bands and a low resolution are unwanted side effects for these types of sensors. However, other properties like weight and size make cameras that contain these sensors ideal for utilization on a UAV. That chapter shows the quantitative and qualitative results for CNN architectures that scale a hyperspectral cube up to 16 times the original
SUMMARY

resolution. The experiments show that the CNNs designed for this purpose are suitable for correcting crosstalk and simultaneously scaling up the hyperspectral cube, while maintaining a spatial and spectral reconstruction with a Structural Similarity (SSIM) index of 0.89. It was also found that the seemingly unwanted effect of sensor crosstalk actually helps upscaling. This shows that deep learning can mitigate several limitations of these types of hyperspectral cameras when used on a UAV.

In Chapter 4 CentroidNet is introduced. This novel hybrid deep learning architecture can be used to detect centroids of objects. It uses a U-Net CNN as a backbone. This CNN is trained to produce a field of voting vectors that point to the nearest centroid of objects located in the image. A computer vision algorithm inspired on a Hough transform is used to aggregate all vectors into a landscape of votes. Positions in the image that receive a large amount of votes are regarded as centroids. One of the strong properties of CentroidNet is that the model can be trained on patches of images and inference can be done seamlessly on full-resolution images. Experiments show that CentroidNet outperforms, based on F1-score, the You Only Look Once Version 2 (YOLOv2) and RetinaNet models for localization and counting potato crops in aerial images.

In Chapter 5 a redesign of the CentroidNet algorithm is proposed that also produces voting vectors for the outline of objects so that shape and size of the objects as well as their centroids can be estimated. Several ablation studies are performed with varying backbones: U-Net and three varieties of Deeplab V3+ using a ResNet50, ResNet101 and Xception backbone. This study also investigates the use of, combinations of, Mean Squared Error (MSE) loss, Vector Loss (VL), Cross-entropy loss and Intersection over Union (IoU) loss. Three image datasets were used in the experiments: aerial images of potato crops, microscopic cell nuclei and bacterial colonies in Petri-dishes. CentroidNetV2 is compared to You Only Look Once Version 3 (YOLOv3) and Mask Recurrent Convolutional Neural Network (MRCNN). On all three data sets CentroidNetV2 achieves the highest recall and on two of the three datasets CentroidNetV2 achieves the highest F1 score. The best segmentation mask and precision is produced by MRCNN, but it detects less small objects than CentroidNetV2.

In Chapter 6 the results are discussed and conclusions are given. This research was framed within the trinity of technologies: deep learning,
hyperspectral imaging and UAVs. During this research it was found that developing solutions with a strong emphasis on the applications itself has been beneficial for the outcome. CNN models that are specifically designed for the optical properties of the hyperspectral camera sensor show excellent results for crosstalk correction and upscaling of hyperspectral cubes. It has also been shown that when a CNN is used to preprocess images and a specially designed computer vision algorithm is then used to detect objects, this combination performs better than models based solely on deep learning. This shows that traditional computer vision paradigms are useful in combination with deep learning to increase performance. The idea is that prior knowledge about the application is implicitly or explicitly added to the model, which causes the model to be less general but more accurate.