A COMBINED METHOD FOR VEGETATION CLASSIFICATION BASED ON VISIBLE BANDS FROM UAV IMAGES: A CASE STUDY FOR INVASIVE WILD PARSNIP PLANTS

by

Jingyi Liu

A thesis submitted to the School of Environmental Studies

In conformity with the requirements for

Degree of Master of Environmental Studies

Queen's University

Kingston, Ontario, Canada

(February, 2018)

Abstract

Wild parsnip is an invasive plant that has serious health risks to humans due to the toxin in its sap. Monitoring its presence and spread has been challenging for conservation authorities due to its small size and irregular shape. On traditional remotely-sensed images acquired by satellites or airplanes, wild parsnip cannot be distinguished from other vegetation due to low spatial resolution of the imagery. Unmanned Aerial Vehicles (UAV) can obtain ultra-high resolution (UHR) imagery and have been used for vegetation-related monitoring in both environmental and agricultural applications in recent years. In this study, UAV images captured at Lemoine Point Conservation Area in Kingston, Ontario, are used to test a methodology for distinguishing wild parsnip. The objective of this study is to develop an efficient invasive wild parsnip classification workflow based on UHR digital UAV imagery. The challenge is that all processing options are based only on visible bands information (RGB) of the digital cameras on UAV, which means there is no infrared spectral information to use. The UAV image processing flow included image orientation, digital elevation model (DEM) and digital surface model (DSM) extractions, individual orthophoto production, and orthomosaic generation processing using Simactive's software. Three vegetation indices and three texture features are calculated and the first two significant features are added to the mosaicked images. The Random Forest algorithm with relevant variables is used as the classifier to distinguish wild parsnip plants from other vegetation types. The optimal image resolution in parsnip analyses are undertaken by comparing accuracy assessments. The results provide an executable workflow to distinguish wild parsnip and show that UAV images, with a simple digital-light camera, are an appropriate and economic resource for small and irregular vegetation detection. The combined method yields reliable and valid outcomes in detecting wild parsnip plants and shows good performance in mapping vegetation.

Acknowledgements

First and foremost, I am deeply thankful to my supervisor, Dr. Dongmei Chen, who has provided guidance and insights during the whole two-year studies. Thank you for helping me choose the amazing idea of processing UAV images to study invasive species in this thesis. Without your patient instructions, I would not feel confident on my master's study. Also, many thanks to you encourage me to attend the conference and assist me to receive the award. I always remember the words that you told us in the regular lab meeting, and in private, to help me realize what I really like in my life. No matter wherever I go in the future, I will always bring all memories you gave me.

I would also like to thank the committee members, Dr. Graham Whitelaw, Dr. Paul Treitz, Dr. Ryan Danby, Dr. Diane Orihel for your insightful comments and suggestions on my thesis. Special thanks to all members of our fantastic Laboratory of Geographic Information and Spatial Analysis (LaGISA) lab at Queen's University who have accompanied and helped me during my master studies. Thanks for all your supports.

I really appreciate the help of Master of Environmental Studies staff: Paul Bass and Karen Topping. Thank you for always being there and assisting me whenever I needed help.

Thanks to Travis York at Cataraqui Region Conservation Authority and Derek Cooper at Kingston Aerials for providing the UAV data to support my graduate research.

During my two-year studies, financial assistance was provided by Queen's University, the School of Environmental Studies, and grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Ontario Centre of Excellence awarded to Dr. Dongmei Chen. I feel grateful to all assistance.

Special thanks to Minhao Gao who supported me at the end of my master's studies. Thanks to Wenqing Ye and Nanfeng Liu who have offered help when I met any academic problems. Also thanks to my best friends in Canada, Jasmine Wang, and Zoe Guo, I am so lucky to meet these awesome girls.

In the end I would like to thank my wonderful parents: Dianwei Liu and Li Zhu. You gave me the

greatest encouragement from my home country China during the whole two years. I know you are always on my side and support me all the time.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CHM	Canopy Height Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
ENP	Everglades National Park
EXG	Excess Green
FNEA	Fractal Net Evolution Approach
LAI	Leaf Area Index
LAN	Local Area Network
GPS	Global Positioning System
GLCM	Gray-level Co-occurrence Matrix
GLDV	Gray-level Difference Vector
IHS	Intensity Hue-Saturation
IMU	Inertial Measuring Unit
ML	Maximum Likelihood
MSS	Multispectral Scanner
NGBDI	Normalized Green-Red Difference Index
PPD	Phytophotodermatitis
PRCM	Predictive Road Condition Modeling
PRI	Photochemical Reflectance Index
RF	Random forest
SVM	Support Vector Machine
TCARI	Transformed Chlorophyll Absorption in Reflectance Index
TM	Thematic Mapper

UAS	Unmanned Aircraft System
UAV	Unmanned Aircraft Vehicles
UAVRS	UAV-based Remote Sensing of Road Condition
VEG	Vegetation
VF	Vegetation Fraction
VI	Vegetation Indices
WAN	Wide Area Network

Chapter 1

Introduction

1.1 Background

Invasive plants are non-native plants to a specific region, and show a tendency of spreading and can cause damage to the wellbeing of an ecosystem by affecting native plants, and potentially risking human and animal health (Boersma, Reichard, & Van Buren, 2006). Some invasive plants can cause severe economic, environmental, or ecological changes (Hooper et al., 2005). Invasive plants particularly affect biodiversity by reducing the number of species in an area thus reducing the gene pool of an area, and affecting native plants in their community. Several studies have compared the difference between the presence of plants in some areas before and after the impact of invasive species (Davis, 2003; Rpsenzweig, 2001; Sax et al., 2002; Slobodkin, 2001). The most significant impact is habitat destruction, resulting in removal of a large number of local species, changing the local habitat, affecting human health, and generating enormous costs (Dewey, 1991; Pimentel, 2005; Lounibos, 2002). Therefore, scientists and managers have been seeking effective solutions to remove invasive plants for a long time.

Wild parsnip is an invasive plant that grows up to 1.5m high with yellow flowers in umbrella shapes with diameters reaching 15 centimeters. It can be found in large patches or as a single plant. It was brought to North America by European immigrants for its edible root (Zohary et al., 2012). Wild parsnip is considered toxic since it causes *phytophotodermatitis* (PPD), which burns the skin of both humans and animals through contact with a chemical in the plant's sap and when the person or animal is subsequently exposed to UV radiation (Averill & DiTommaso, 2007). This toxic chemical in the sap, called *furanocoumarins*, can cause intermediate phototoxic reactions, and the affected area can remain discoloured for up to two years (Carlsen & Weismann, 2007). Therefore, some vegetation management offices, such as Ontario Invasive Plant Council, have put greater importance on monitoring and controlling wild parsnip (Danielle & Kellie, 2014). Currently, the practice used to control the spread of wild parsnip is by eradication at the appropriate time (when the flower buds are starting to grow). An incorrect mowing time will result in an even worse spreading problem (Averill & DiTommaso, 2007). However, identifying wild parsnip and monitoring its presence and spread in a large region has been challenging for conservation agencies due to its small size and irregular shape, especially for single scattered plants.

Remote sensing images acquired by satellites or airplanes, whether they are low resolution or ultra-high resolution (Pande-Chhetri et al., 2017) have been applied for monitoring vegetation in different research areas for many years (Melesse et al., 2007). However, traditional satellite or aircraft images cannot capture plant-level details, especially for scattered and small plants like wild parsnip. Ultra-high resolution images such as UAV images have been widely applied in vegetation analysis due to the advantages of high resolution, high frequency, and easy operation (Salam íet al., 2014).

UAV is becoming more popular in recent years because of its advantages as an economic and repeatable method with easily-controlled equipment. UAV can capture ultra-high resolution images that provide detailed features suitable for estimating the location of individual plants within several hectares (Colomina & Molina, 2014). As the flying altitude of UAV can be adjusted, the resolution of the acquired images ranges from kilometers to centimeters. With this level in resolution, UAV images can satisfy the requirements of small vegetation analysis, particularly in a relatively small study area.

UAV used for remote sensing and surveying usually involves a camera for high-resolution image acquisition. These UAV cameras can be inexpensive similar to digital cameras or very expensive similar to LiDAR equipment (Feng, Liu & Gong, 2015). Commonly, cameras used in UAV's contain three visible bands (RGB bands) only, and have a limited spectral resolution. The limited spectral bands, especially lack of infrared bands, are usually not enough to identify and classify vegetation.

Vegetation indices and texture features calculated based on RGB bands can be treated as additional bands for vegetation classification. Vegetation indices, which are spectral calculations emphasizing vegetation properties (Bannari et al., 1995), have demonstrated to be effective in land cover classification since 1972 (e.g. Rouse, 1972). Vegetation indices have successfully improved the classification accuracy results with different airborne and satellite platforms (Bendig et al., 2015; Price et al., 2002; Viña et al., 2011). Texture is another feature that analyzes the homogeneity of images based on scale, uniformity, regularity, etc. (Haralick, 1979). Texture feature analysis resulting in extracted spatial changes of brightness occupies an important position in remote sensing (Kitada & Fukuyama, 2012; Racoviteanu & Williams, 2012). In recent years, texture features have been implemented in grassland classification, wetland species classification, and forest classification, and that improved the classification accuracy (Dawkins & Esiobu, 2016; Puissant et al, 2005; Szantoi et al., 2013).

Common vegetation classification methods can be classified as pixel-based and object-based. Pixel-based classification is based on individual pixels and assign each pixel to a class; whereas object-based classification groups pixels into features with shapes and sizes (Tarabalka et al., 2009). Object-based image classification has shown better accuracy in many previous studies since it combines spectral and spatial features into the process (Myint et al., 2011; Tarabalka et al., 2009). However, this classification method presents different image segmentation choices, which require multiple input parameters that are not easy to process (Yu et al., 2006).

The primary gap in the current research is the lack of efficient methods to distinguish single small and irregularly shaped species, such as wild parsnip. At present, the common way of identifying wild parsnip is still through observing and collecting data from field surveys in person. Although human visual identification through fieldwork may have a higher accuracy of locating invasive plants, the high field cost, the high time demand, and low efficiency make field surveys impractical for monitoring. The existing workflows of classifying vegetation from satellite-based images require knowledge of the optimal ground sample distance (GSD) of images, and additional field input to be able to distinguish the invasive plants from other vegetation species, since some invasive plants may appear scattered or in patches. Furthermore, the ability of UAV images to distinguish small vegetation with only visible bands (RGB) is questionable. Generally, satellite or aircraft-based images have at least the infrared band to help classify vegetation. There is a need for more tests to see whether ultra-high resolution RGB UAV images can compensate for the lack of infrared or thermal bands in classifying small scattered plants.

1.2 Research Objectives and Questions

The primary objective of this study is to develop an optimal workflow for classifying wild parsnip based on RGB visible bands of ultra-high resolution UAV imagery. The main question that this research is intended to address is whether wild parsnip can be distinguished from other vegetation using ultra-high resolution UAV images

More specifically, three research questions are listed to address this objective:

a. What is the optimal resolution of UAV images to distinguish wild parsnip?
b. What are the appropriate vegetation indices and texture features that can be added to the RGB bands to help distinguish wild parsnip?
c. What is the most effective algorithm for classifying small and irregular vegetation?

After the testing, a workflow with pixel-based random forest (RF) classification method is proposed to map wild parsnip based on visible bands with additionally suitable vegetation indices and texture features (section 4 of the thesis).

1.3 Thesis Overview

This thesis contains 5 chapters and is organized as the follows:

- Chapter 2 is the literature review of key definitions, important methods, and successful studies in vegetation classification.
- Chapter 3 is the methodology used for this research on exploring the optimal workflow to distinguish wild parsnip from RGB visible bands of ultra-high resolution UAV imagery.
- Chapter 4 is the results and discussions for the optimal workflow selection combining RGB bands, texture features, and vegetation indices with random forest classifier.
- Chapter 5 concludes the thesis by summarizing key results, as well as the limitations and suggestions for future research work.

Chapter 2

Literature Review

2.1 Parsnip Overview

Given the increasing impacts of human activities, it is becoming more important to recognize and understand how ecosystems and biodiversity are being changed (Tylianakis et al, 2008). One significant cause of ecosystem alteration is invasive alien plant species, whose impacts have been widely discussed through an extensive literature. For instance, invasive plants have been demonstrated to affect the natural growth of local vegetation (Hejda et al., 2009), reduce crop yields and threaten food security (Cook et al., 2011), and cause harm to human health (Kopelson, 2014).

Wild parsnip is an invasive plant that can grow up to 1.5m high with umbrella-shaped yellow flowers approximately a 15 centimeters diameter. Wild parsnip, a biennial plant, is found either in large patches or as single plants (Figure 2.1). It was brought to North America by European immigrants in Virginia in 1609 for its edible root (Zohary et al., 2012). Since the 1800s, its taproot had become common in North American cuisine. However, today the plant has emerged as an invasive species located around roadsides and waste places across North America (Berenbaum et al., 1984). The parsnip is not only a harmful invasive species, but it can also cause injury to human skin (Figure 2.2) if there has been a contact with the plant's sap under sunlight. The species has been seen to also affect native plants, reduce the quality and marketability of crops, and lower weight and fertility in livestock (Berenbaum & Zangerl, 1997).

(a)

(b)

(c)

Figure 2.1 Wild Parsnip (a) Single Parsnip Plant in Summer (OFAH/OMNR Invading Species Awareness Program, 2012) (b) Single Parsnip Plant in Fall (c) Clustered Parsnip Plants in Summer

Figure 2.2 A Boy with Phototoxic Reactions after Coming into Contact with Wild Parsnip (Carlsen & Weismann, 2007)

Averill & DiTommaso (2007) presented information on parsnip including its history, distribution, habitat, biology, and values. Parsnip was initially distributed in Europe and temperate Asia, and now is found in Germany, Australia, Canada, and the United States among other places (USDA ARS, 2006). In Canada, the cultivated parsnip was first established in 1612, and wild parsnip was first found around Montreal and Quebec City (Cain et al., 2009). In Canada, the Northwest Territories and Nunavut are the only regions where parsnip is not found (Darbyshire, 2003). In Ontario, parsnip is considered a toxic plant since it can cause Phytophotodermatitis (PPD) as described below (Mulligan, 1987).

Wild parsnip is a monocarpic biennials plant and grows to a typical size (root crown diameters at least 0.5cm) in that fall season in order to flower the following summer (Averill & DiTommaso, 2007). Wild parsnip reproduces only by seeds, and spreads by wind, water and during mowing. Usually, plants grow from spring to fall, and in the cold winter, it is senescent (Zohary et al., 2012). Wild parsnip needs at least two years to mature, at which time it flowers and then dies. There is no evidence in the literature that wild parsnip hybridizes with other plants.

Although cultivated parsnip and wild parsnip belong to the same family, the components inside the two parsnips are different (Cain et al., 2009). Cultivated parsnips produce bigger and tastier roots than wild parsnips. Cultivated parsnip is edible with less toxic properties, and wild parsnip grows as a weed and has higher toxicity (Berenbaum et al., 1984). The toxicity is due to the chemical compounds (furanocoumarins) inside wild parsnip which can lead to abrasions and cell damage in humans, livestock, and plants when in contact with the chemical followed by UV exposure (Zangerl et al., 1989; Rietschel et al., 1995). The effects of the toxicity of furanocoumarins may influence DNA transcription and replication (Scott et al., 1976).

The sap inside the parsnip can cause PPD in humans and livestocks, which is a chemical burn lasting a few days. The PPD symptoms include redness, burning, and blisters, and the area of the inflammation will remain discoloured for up to two years (Averill & DiTommaso, 2007; Burrows & Tyrl, 2001; Cain et al., 2009). Its toxicity increases during the plant's flowering season. In Canada, flowering usually happens between late May and October, and the peak flowering period is between mid-June and the end of July (Alex & Switzer, 1992).

Serious reactions in humans to wild parsnip have been reported from hospitals in Copenhagen, Denmark by Carlsen and Weismann (2007). Fifteen children around the age of 5 had irregularly shaped erythematous areas on their faces and legs after a picnic day in the forest. Doctors were able to diagnose the skin reactions and demonstrate that it was caused by PPD from wild parsnip in the forest. Another 4-year-old girl was suffering from inflamed and scorching eczema on her arm and face. She got these symptoms after picking flowers from a yellow and green plant, which was recognized as wild parsnip. It is more common with children since they are unfamiliar with wild parsnip (Coffman et al., 1985); nevertheless, it can also happen with adults who are hiking or gardening outdoors (Cain et al., 2009).

Skin contact with wild parsnip causes the same reaction in livestock and pets. Dogs (Figure 2.3) exposed to mashed parsnip leaves experience severe sunburn and secondary skin infections (Hayden, 2000; Lea, 2015). Farmers also reported that four sows suffered serious consequences after walking a field where wild parsnip was present (“Parsnip Poison Pigs”, 2014). Therefore, animals can be hurt if both the sap of wild parsnip and UV light reach the skin.

Figure 2.3 Photograph of a Dog’s back after walking and being exposed to Wild Parsnips (Lea, 2015)

Given the serious health impacts of the species, wild parsnip is regarded as a prohibited noxious weed by local weed control organizations in Ontario (Cain et al., 2009). Therefore, the Ontario Invasive Plant Council has been considering methods to control wild parsnip (Danielle & Kellie, 2014). Mowing is a practical way to control wild parsnip; however, parsnip mowing should be done at a specific time (when the flower buds are starting to grow) to avoid regrowth by the seedlings (Baskin & Baskin, 1979). Biological methods of control have also been considered, but agents for controlling wild parsnip are not easily applied (Søborg et al., 1996). Public awareness campaigns should be conducted to warn people of the health impacts associated with contacting wild parsnips, since they are common in some public parks and roadways. After contact, people should wash their skin immediately, and avoid sunlight in the following 2 to 3 days (Burrows & Tyrl, 2001).

Monitoring and identifying wild parsnip before it spreads widely will reduce the harmful influence on human beings, livestock, and agriculture by limiting exposure to the plant. Wild parsnip can grow scattered or in patches in grassland, gardens or agricultural fields. Currently, field surveys are still the main approach to detect wild parsnip; however, this is a time-consuming process and with the high cost of human labor, observing and collecting wild parsnip in the field may become inefficient. Therefore, developing and testing a state-of-art monitoring technology to identify wild parsnip is important for conservation efforts.

2.2 Remote Sensing for Vegetation Identification and Monitoring

Remote sensing involves obtaining data from aircraft and satellites using various types of sensors to measure the objects on the earth (Schowengerdt, 2011). Vegetation is regarded as an essential component of the ecosystem, and vegetation classification based on quantitative methods is necessary to help identify and monitor plant resources (Xiao et al., 2004). Xie et al., (2008) presented a comprehensive review on vegetation classification analysis using remote sensing technology. Vegetation classification is applied on different imagery, whether global or local in scales, at a specific time or for longer periods (Langley et al., 2001; Nordberg & Evertson, 2003). Vegetation classification includes steps such as acquiring data from different sensors, applying image classification methods, and assessing the accuracy of classification. Usually, remote sensing sensors acquire images from various sensors mounted in satellites and aircraft such as MODIS, AVHRR, Landsat, IKONOS, and QuickBird with different spectral, spatial and temporal resolutions (e.g. Torres-Sánchez et al., 2013). Table 2-1 is a summary of recent studies that have used different remote sensing inputs to map invasive plants. These studies successfully used both high spectral and spatial resolutions from different sensors to classify invasive species.

In remote sensing, vegetation identification and monitoring procedures include steps of pre-processing images to increase the quality of initial images, and developing classification methods (classifier) to identify types of vegetation either based on pixels or objects (Sluiter, 2005). The final step is an accuracy assessment of the classification, the output is compared with the actual data gathered through field work (Xie et al., 2008). Based on the results of the assessment, the accuracy could be improved by using other more advanced image classification methods. Choosing the most appropriate image classifier will always be a challenge, there isn't a classification method that fits all types of research, and each study will require a specific classifier.

Table 2-1 Summary of recent studies that have classified invasive species using remote sensing data. AVIRIS = Airborne Visible/Infrared Imaging Spectrometer, MSS = Multispectral Scanner, TM = Thematic Mapper, ETM = Enhanced Thematic Mapper, AISA = Advanced Imaging Spectrometer for Applications (Chance, 2014)

Invasive Species	Study Area	Type	Remote Sensing Sensors	Spatial Resolution	Spectral Resolution	Sources
Hoary Cress	Idaho	0.5m to 1m tall rhizomatis perennial	HyMap (hyperspectral)	3m	450-2500nm, 126 channels at 15-20nm each	Mundt et al. (2005)
Iceplant, jubata grass	California	Iceplant - mat-forming succulent, jubata grass – grass	AVIRIS (hyperspectral)	4.5 m	174 channels 374-394nm, 1133-1464nm, 1773-2051nm	Underwood et al. (2003)
Cheatgrass (Bromus tectorum)	North Dakota	Grass	Landsat MSS, TM, ETM	30 m	1 thermal band at 10400 nm, 3-7 spectral bands from 400 nm to 2550 nm,	Bradley & Mustard (2006)
Water caltrop, Phragmites australis, and Purple loosestrife	New York	water caltrop - submerged stem, leaves floating on the surface, Phragmites - large perennial grass, purple loosestrife - herbaceous	Quickbird, photo interpretation of aerial photos (multispectral)	Quickbird = 2.4m	450-890nm (non-continuous), 4 channels	Laba et al. (2008)
Late goldenrod	Japan	Clonal growth, seed dispersal, highly competitive	AISA Eagle (hyperspectral)	1.5m	67 channels, 397nm to 983nm, 8.9nm each	Ishii & Washitani (2013)
Chinese privet	North Carolina	Understory shrub	Optech ALTM Gemini 3100 LiDAR	1 point/m2	N/A	Singh et al. (2015)
Leafy spurge, and spotted knapweed	Montana	Leafy spurge - herbaceous perennial, spotted knapweed – herb	Probe-1 (hyperspectral)	3 m	128 channels, 440nm to 2507nm	Lawrence et al. (2006)

2.2.1 Vegetation Indices from Remote Sensing

The first remote sensing carried out to find a relationship between spectral reflectance and vegetation occurred in 1972, after the launch of a NASA satellite (Rouse 1972). The earliest study that established definitions of a vegetation index (VI) and conducted specific VI studies on vegetation was in 1973 by Rouse. The objective of using VIs in remote sensing is to enhance identification of vegetation types, amounts, and conditions in the scene, while minimizing solar radiation and soil influences (Moulin & Guerif, 1999). The concept of using VIs as a quantitative measurement is combining mathematical operations of different channels (Bannari et al., 1995). The common channels in VIs are the visible and near-infrared paties of the spectrum. Through various combinations of these spectral values researchers/managers are able to discriminate vegetation from soils and other land covers (Tucker, 1979). Table **2-2** presents the common VI applications in remote sensing studies. VI applications have been regarded as a straightforward and convenient method in remote sensing applications (Leprieur et al., 1994; Govaerts et al., 1999; Vertraete & Pinty, 1996).

Table 2-2 Common Vegetation Indices Applications in Remote Sensing (Bannari, Morin, Bonn & Huete, 1995)

Vegetation Index	Abbreviation	Formula	Source
Atmospherically Resistant Vegetation Index	ARVI	(NIT-RB)/(NIR+RB)	Kaufman & Tanre, 1992
Differenced Vegetation Index	DVI	NIR-R	Clevers, 1986
Modified SAVI	MSAVI	$[2NIR+1-\sqrt{(2NIR+1)^2-8(NIR-R)}]/2$	Qi et al., 1994
Normalized Difference Index	NDI	(NIR-MIR)/(NIR+MIR)	McNairn & Protz, 1993
Normalized Difference Vegetation Index	NDVI	(NIR-R)/(NIR+R)	Rouse et al., 1974
Perpendicular Vegetation Index	PVI	$(NIR - aR-b)/(\sqrt{a^2+1})$	Jackson et al., 1980
Ratio Vegetation Index	RVI	R/NIR	Pearson & Miller, 1972
Soil Adjusted Vegetation Index	SAVI	(NIR-R)(1+L)/(NIR+R+L)	Huete, 1989
Transformed SAVI	TSAVI	[a(NIR-aR-b)]/(R+aNIR-ab)	Baret et al., 1989
Normalized Difference Greenness Index	NDGI	(G-R)/(G+R)	Chamard et al., 1991
Normalized Green-Blue Difference Index	NGBDI	(G-B)/(G+B)	Du & Noguchi, 2017 Sanjerehei, 2014 Wang et al., 2015
Excess Green	EXG	2G-R-B	Wang et al., 2015
Vegetation	VEG	$G/(R^aB^{(1-a)})$	Guijarro et al., 2011

The application of VIs can range from objects to the entire ecosystem (Boegh et al., 2002; Curran, 1983; Viña et al., 2011; Xiao et al., 2002). For this thesis only studies that applied VIs to specific species were reviewed because the focus was on distinguishing wild parsnip. Viña et al., (2011) conducted a study to assess the ability of several VIs to estimate the green leaf area index (Green LAI) of two crops. Eight VIs were assessed, and the results showed that the CIRed-edge was the most appropriate index, which can be applied as a precise and affordable method for estimating Green LAI at different scales by using satellite sensors data.

VIs have been applied widely in vegetation classification. For example, Price et al., (2002) conducted another study to compare the performance of the Thematic Mapper (TM) band combinations and VIs in distinguishing six grassland types in eastern Kansas. The VIs were calculated based on RGB and near-infrared bands, and the results showed that the VIs performed better than only the raw TM bands from grassland classification. Another study that used VIs in vegetation analysis is Bendig et al. (2015). They illustrated that the GnyLi index with plant height was suitable for biomass analysis, and GRVI, MARVI, and RGBVI with plant height may be useful for further biomass monitoring studies.

Along with VIs that were used with moderate remote sensing image data, VIs were also applied in ultra-high resolution image applications. Berni et al., (2009) showed the successful application of using VIs to monitor vegetation through multiple band UAV images. However, some of the UAV images only contained RGB bands and that resulted in limited VI choices. RGB-based VIs such as, NGBDI (Normalized Green-Blue Difference Index), EXG (Excess Green), and VEG (Vegetation) were discussed in numerous studies. For example, Du & Noguchi (2017) conducted a study by using NGBDI, EXG and other VIs to map wheat yields; Sanjerehei (2014) assessed the ability of several VIs including NGBDI, to estimate vegetation cover in shrublands; Wang et al (2015) separated vegetated areas from non-vegetated areas by using VIs including EXG and NGBDI; and Guijarro et al (2011) applied EXG, VEG and other VIs in mapping agricultural images. These studies demonstrated that NGBDI, EXG, and VEG contribute to the mapping of land cover, but the limitation of only using VIs in land cover classification were also presented. To seek other additional useful variables, texture features (described below) could be a choice to contribute to and enhance vegetation studies. David & Ballado (2016) conducted a study to map land cover in vegetated areas using RGB UAV images. Texture features and nine VIs were used as supplemental data. The results showed both texture features and VIs improved the classification accuracy from weeds to eggplants. This study demonstrated the effectiveness of adding texture features to VIs in vegetation classification when there are insufficient spectral bands.

2.2.2 Texture Features from Remote Sensing

Texture features are multiple visual shapes of objects, and include characteristics such as lightness, smoothness, fineness and size (Materka & Strzelecki, 1998). The four main processes in texture analysis are (1) feature extraction, (2) texture discrimination, (3) texture classification, and (4) shape from texture (Levine, 1985). Feature extraction is a computing procedure that describes the property representation of each image. Texture discrimination segments the textural image into sub-regions based on physical differences. Texture classification is a transformation from the unknown classes of the image to some defined classes. Shape from texture is a progress that reconstruct the image from 2D to 3D (Wang et al., 2016). To calculate texture features there are four approaches, which are (1) structural, (2) statistical, (3) model-based, and (4) transform (Haralick, 1979; Levine, 1985). In recent years, the statistical method has been commonly used, which includes first-order and second-order textural analysis (Haralick, 1979). First order textural analysis is calculated based on individual pixels and includes mean, variance, skewness, kurtosis, energy and entropy. Second-order textural analysis is calculated based on pairs of pixels, which are derived from the joint probability distributions (Alparone 1990). The algorithm is computed from the gray-level co-occurrence matrix (GLCM), which explains the spatial relationships, and the accuracy of color texture segmentation (Valkealathi, 1998). It includes angular second moment, correlation, inertia, absolute value, inverse difference, entropy, and maximum probability (Alparone 1990).

Texture features are used in vegetation classification based on past successful studies. Szantoi et al., (2013) developed an effective workflow to map wetland vegetation based on high-resolution images with texture features and VIs as supplementary data. The results showed that texture features effectively classified the wetland vegetation, and assisted with monitoring the plant communities. Murrary et al., (2010) proposed the inclusion of texture features to vegetation classification by using high-resolution IKONOS images. The results showed that the inclusion of texture and multispectral features improved the classification accuracy. Coburn & Roberts (2004) employed both first and second-order texture features in forest classification by using high-resolution images. The outputs showed that supervised classification combined with texture features achieved the best results in forest species classifications. Studies have also been carried out in urban land cover classification that used texture features to improve overall accuracy. For example, Puissant et al., (2005) applied texture features to improve the accuracy of pixel-based classification on high spatial resolution images.

2.2.3 Image Classification Methods in Remote Sensing

Land cover image classifications can be divided into two methods (1) pixel-based, and (2) object-based. Both methods work in classifying units to classes through multiple algorithms (Duro et al., 2012). Pixel-based image classifications have been the main method for classifying land covers for a long time, whereas object-based image classifications have become popular more recently (Blaschke, 2010). Numerous studies have compared the accuracy of the two methods in mapping land cover (Flanders et al., 2003; Thomas et al., 2003; Yu et al., 2006). For example, Flanders et al., (2003) used object-based image classifiers to map forest classes with ETM+ image data, and the results showed 36% higher accuracy than the pixel-based ML method. However, the object-based methods for small and irregular objects still need more research to improve accuracy. Previous studies show that pixel-based methods have successfully distinguished irregular shapes in land cover (e.g. Huang et al., 2002; Gislason et al., 2006).

Scientists and researchers have made great progress in inventing and improving advanced classification algorithms to achieve higher classification accuracy. The classification algorithms include maximum likelihood (ML), support vector machine (SVM), random forest (RF) and heterogeneous clustering approaches. ML is one of the most commonly used supervised classification methods in remote sensing. ML based on pixels assumes the data within each input feature is normally distributed, and each pixel assigns to the class that achieved the highest probability value (Otukei & Blaschke, 2010). SVM is another supervised classification method that assigns each training sample into one of the two pre-defined classes (Vapnik, 1999). It can be applied in either pixel-based or object-based classifications. RF is a new supervised classification method in remote sensing and has become popular in recent years. The output of RF is based on selecting the majority vote of the classification decision trees. It can be used based on either pixel-based or object-based classifications (Feng et al., 2015).

The comparisons between these three algorithms or two of the three algorithms were conducted in several studies, and the majority of them indicated that RF achieved higher accuracy results. For example, Ok et al. (2012) found that RF performed about 8% better than ML in agricultural crop classification, and Kulkarni & Lowe (2016) found that RF was better than ML and SVM in land cover classification. However, some studies also came to opposite conclusions, for example, Lee et al. (2013) found that ML performed around 10% better than RF in land cover classification, and Nitze et al. (2012) found that SVM achieved higher accuracy than ML and RF in crop classification. Since ML and SVM have been applied in remote sensing classification for many years, and RF is a relatively new approach, more research on RF is needed to determine its accuracy. Therefore, this study mainly focused on classification based on RF after the comparison with ML and SVM.

RF is a supervised classification method using classifiers like trees, where the input is combining a number of decision trees in training, and the output is obtained from the classification results of the individual trees (Feng et al., 2015). The diagram in Figure 2.4 visually explains the RF operating principles. In the training stage, RF uses a number of Classification and Regression Trees (CART), which is made of decision trees, each decision tree is trained based on a bootstrapped from the original data by choosing the input variables randomly to define a split (Breiman et al., 1984; Gislason, Benediktsson, & Sveinsson, 2006). In the classification stage, each decision tree votes to the class of input, and the output is defined by the majority votes. The parameters of the algorithm are defined by the user, and RF is not sensitive to the parameters (Dietterich, 2000). The advantages of RF compared to other classification algorithms is that it can rank variables' importance and has the flexibility of performing multiple types of data analyses, and it is suitable for beginners (Cutler et al., 2007).

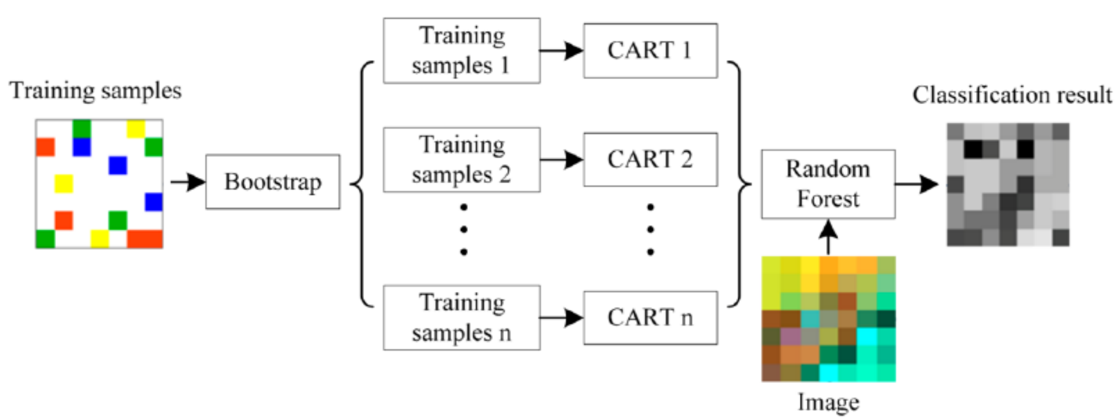


Figure 2.4 Illustration of the Random Forest Algorithm (Feng et al., 2015).

The RF is a robust machine learning classification algorithm and has been applied to map complex land covers (e.g. Rodriguez-Galiano et al., 2012; Pal, 2005). In Rodriguez-Galiano et al studies, RF was shown to decrease the interference from data reduction and noise, and it could also rank the importance of input variables which brought benefits for further studies. Pal (2005) tried to explore the performance of RF and to compare it with the SVM classifier. Results illustrated that both classifiers showed similar results, but RF was easier to operate with less parameter settings than SVM. Cutler et al. (2007) conducted a study to test the performance of RF in mapping species in an ecological area, and to compare it with four other statistical classifiers. The results showed that RF performed better than the other classifiers for all species in the study area. In addition, RF has also been applied in urban scene classifications. Guo et al., (2011) applied RF to map urban land covers, and the results demonstrated that RF could increase the overall accuracy to 72%, which was much better than other classifiers. One of the unique features of RF is ranking the importance of input variables. Gislason et al., (2006) found out that the importance ranking of variables was useful in improving image classification accuracy.

2.3 Unmanned Aircraft Vehicles Applications

2.3.1 General UAV Applications

The definition of an unmanned aerial vehicle (UAV) is a general aircraft without a human pilot onboard and is a component of an unmanned aircraft system (UAS). The flight of UAVs can be operated by a human operator, or by onboard computers (Unmanned Aircraft System (UAS), 2010). UAVs are regarded as a low-cost platform for acquiring digital images; therefore, it has become a valuable source of data for investigation, surveillance, mapping, and 3D modeling (Nex & Remondino, 2013). For geomatics applications, UAVs became a standard source for data acquisition only recently, though its first application was thirty years ago (Pajares, 2015).

Traditional airborne remote sensing is still important for large area coverage. UAV advantages include its high resolution, low-cost and time effectiveness, and it is especially useful for rapid response applications (Nex & Remondino, 2013). UAV images have been employed in many areas, including agriculture, forestry, architecture, environment, emergency management, and traffic monitoring (Niethammer et al., 2010; Zhang, 2008; Coppa et al., 2009).

Remote sensing satellites have been used in agricultural management for a long time, however, some conventional remote sensing systems cannot meet the needs of some applications (e.g., Lu et al., 1997). For example, crop controlling and damage estimation applications require continuous data, and high-resolution images (Moran et al., 1997). An ideal remote sensing application in agriculture should include the ability to capture high-resolution images and to be able to fly at numerous times (Berni et al., 2009). Therefore, UAVs have been applied in agriculture fields for its portable size, low weight, adjusted flight speed, low altitude and extreme capability (Lu et al., 1997).

Herwitz et al. (2004) used UAVs for local agricultural resource monitoring in a study area in Hawaii. This study illustrated the effectiveness of a slow-flying UAV, equipped with a high-resolution camera and a multispectral camera, to monitor a local agricultural region for an extended period. UAVs' imagery has been used for environment regulation, such as road condition assessment (e.g. Zhang, 2008). UAV was selected as the data acquiring platform because of its ability to fly close and obtain sufficient road details (Zhang, 2008). Niethammer et al. (2010) utilized UAVs with digital cameras to monitor and analyze active landslides in the Southern French Alps. The Niethammer et al. (2010) study proposed methods to deal with initial UAV images efficiently, and conducted a digital surface model that was more accurate than some traditional remote sensing data. UAV technology has been also used in emergency rescue, such as the MORAKOT typhoon case in Taiwan (Chou et al., 2010). Thus, UAV platform could conquer transport barriers and obtain real-time information from damaged regions.

2.3.2 UAV Image Applications in Vegetation Classification

This thesis uses UAV images as a data source. The following is a review of UAV images applied in vegetation classifications. UAV can be used as a remote sensing input since its advantages include acquiring high-resolution images, cost-effective, and flexible flying altitudes at an acceptable time (Bryson et al., 2009). Bryson et al., (2010) presented a study for classifying several vegetation species using UAV information from global positioning system (GPS), an inertial measuring unit (IMU), and a digital camera. The results demonstrated that UAV images were an appropriate approach for differentiating between invasive weeds and native vegetation.

UAVs can work with either visible bands camera or multispectral bands camera, or both (Torres-Sánchez et al., 2014). The study conducted by Yu et al., (2006) provided an object-based classification to classify vegetation types by using UAV multi-band images. However, not all UAVs contain both digital and spectral cameras; some studies found solutions for vegetation classification with only RGB bands (Torres-Sánchez, 2015; Laliberte & Rango, 2009; Torres-Sánchez, 2013; Laliberte, 2010). For example, Laliberte & Rango (2009) acquired RGB bands from UAV images to identify rangeland vegetation, and captured texture and scale features as supplementary data. The study showed that UAV images with only visual bands could be complemented by texture features.

Accordingly, the main challenge of applying UAV images in remote sensing is the limited spectral resolution. Due to this limitation, UAV often carries an off-the-shelf digital camera which provides high spatial resolution images in spectral resolution only (three visible bands). These three bands alone may be inadequate; therefore, this research is seeking solutions and alternatives for this drawback.

Chapter 3

Methodology

The workflow of this study is shown in Figure 3.1. It can be divided into three parts: data preprocessing, image classification, output assessment and accuracy evaluation. Data preprocessing included image orthorectification, Digital Surface Model (DSM) and Digital Elevation Model (DEM) extraction, and image mosaicking from Simactive software. Image classification was based on RF classifier and includes masking RGB images, calculating texture features and deriving vegetation indices as input variables. RGB masked images used Canopy Height Model (CHM) to remove trees and tall bushes, and vegetation indices to remove the non-vegetation land cover. Output assessment and accuracy evaluation were analyzed based on confusion matrix in parsnip classification to define the optimal image resolution and the optimal input variables.

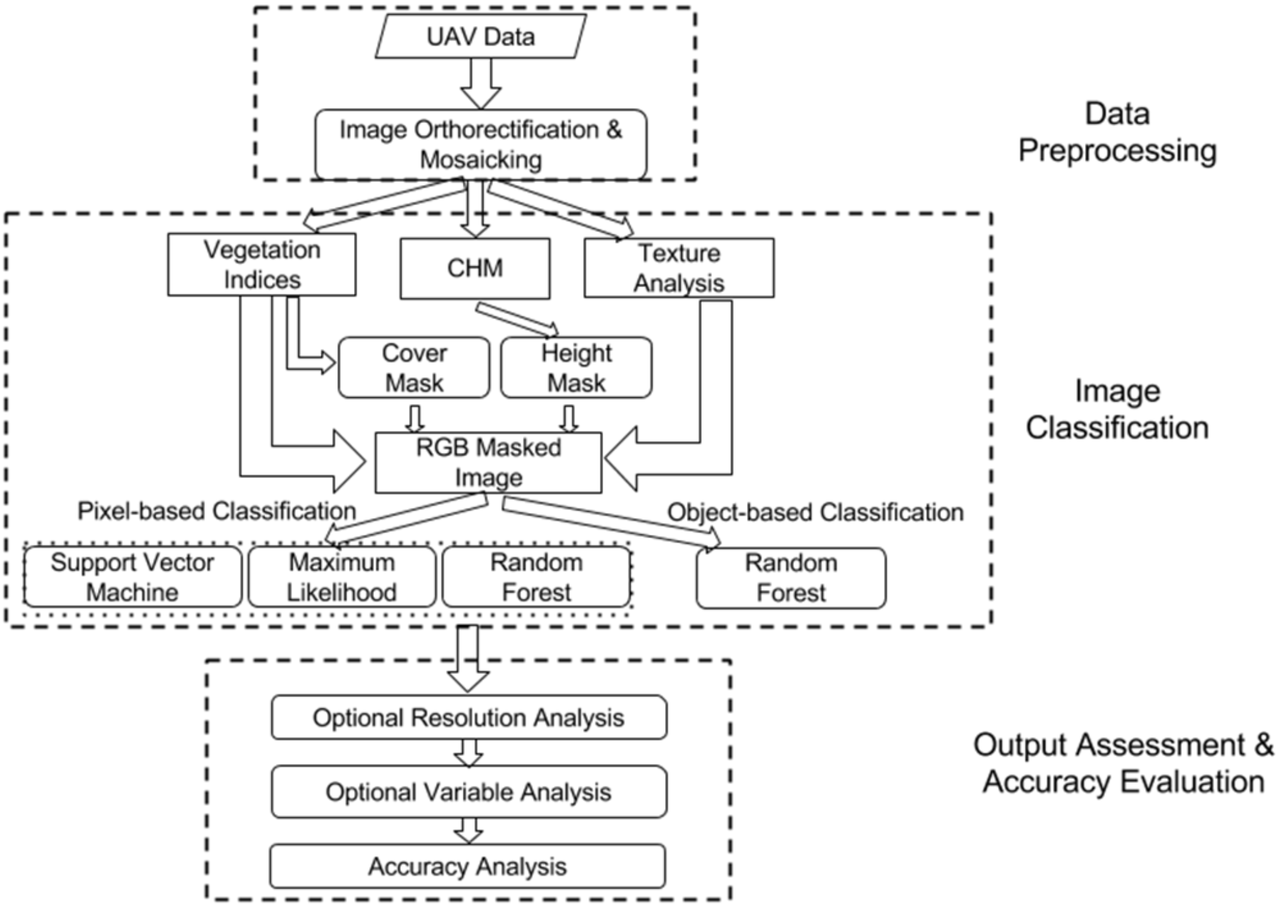


Figure 3.1 Workflow of distinguishing wild parsnip

3.1 Data & Study Area

The study area is Lemoine Point Conservation Area, located at 44.23 °N, 76.61 °W, in Kingston, Ontario, Canada and bordered by Lake Ontario and Collins Bay. Lemoine Point Conservation Area is a heavily-used conservation area with both recreation and natural resources and offers hiking, picnicking, cycling, and a series of outdoor activities. Wild parsnip has been found increasingly in the conservation area and is becoming a growing concern for the Cataraqui Region Conservation Authority. Some places with large wild parsnip infestations were secured from access to prevent people and pets from being hurt (Cataraqui Region Conservation Authority, 2017).

The Cataraqui Region Conservation Authority would like to explore the use of UAV technology to detect and locate wild parsnip, and seek solutions to remove them effectively. Digital photographs were collected by a UAV operated by Kingston Aerials on July 27, 2016 for the study area (Lemoine Point Conservation area). This time period coincides with the end of the flowering season of the wild parsnip. Three-band (red, green, and blue) images were acquired with a commercial digital camera Sony A7R (Zeiss batis 25mm) at approximately 152m above terrain. The individual raw image size was 7360 x 4912 with a spatial resolution (GSD) of 0.94 cm per pixel.

DSM is an elevation model with the tops of surface objects such as buildings, vegetation, and other objects, whereas DEM represents the elevation of bare earth without any objects (Li 1990). DSM and DEM have been generated in the software Correlator3D in this thesis. The processing started with inputted individual UAV images and their GPS info, camera parameters and the exterior image orientation. In the next step, key points were extracted based on the optimized the SIFT algorithm (Simactive User Manual, 2016). If two key points are found to be the same on two different images, the two points are called one pair of tie points. The tie points in each pair represents the same location in adjacent images, and are used to link two images (Kennedy & Cohen, 2003). The software was automatically finding millions of tie points on images, and each pair of the tie point generated one 3D point. Figure 3.2 shown how the height was generated from tie points on two overlapped images. Millions of 3D points with known positions were generated into the 3D point cloud. The top points of the point cloud were used to generate the DSM, and the ground points were used to develop the DEM. Figure 3.3 illustrated an example about a transform from DSM to DEM in Correlator3D.

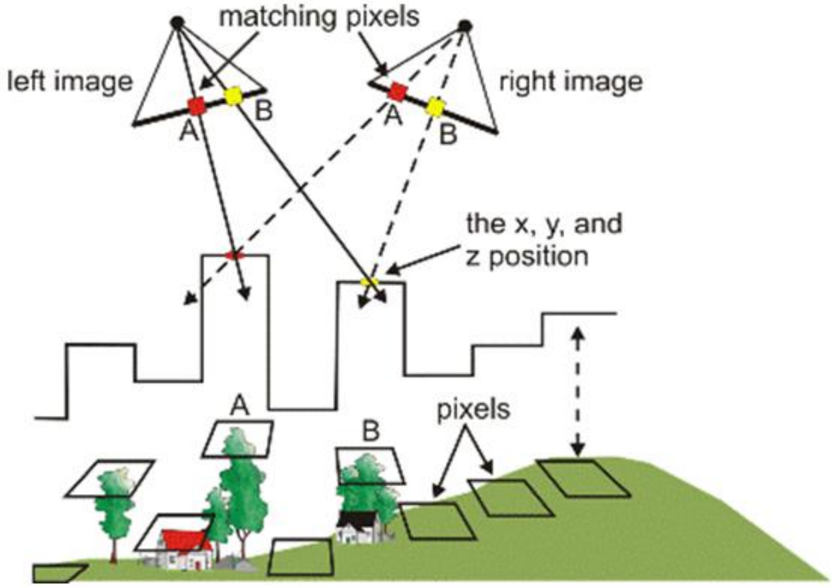


Figure 3.2 An Illustration of the Height Calculated on Two Overlapped Images (PCI Geomatics, 2015)

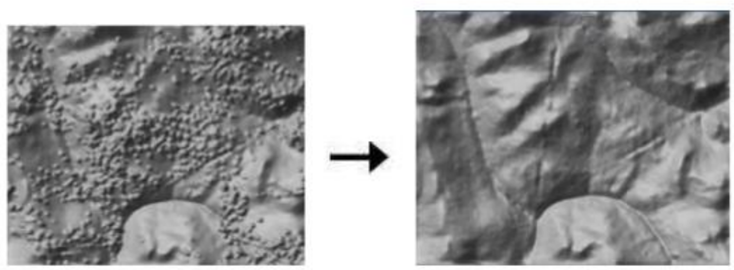


Figure 3.3 An Example of Correlator3D Generated DSM to DEM (Simactive User Manual, 2016)

Correlator3D provides completely automated merging of individual orthorectified images, and seamless transitions between adjacent images, based on a series of input UAV images. Correlator3D automatically extracts key points from individual images, builds tie points between images, and estimates the location of each feature, based on the UAV GPS information and camera parameters provided. The process matches images with each individual latitude, longitude, and altitude (XYZ) values, which are recorded from the carried GPS. Correlator3D mosaics the photos from different positions and angles, and stitches the photos into a seamless image through the algorithms selected from the software. The procedures used to mosaic the photos are based on previous tie points that are constructed between photos. According to the tie points, one of the two photos is rotated to match the other, and then the photos are stitched together. After that, the image uses the same method to stitch each subsequent photo, thereby stitching all the photos into one seamless image. The spectral values of the mosaicked image are derived from the overlapping portions of the photos by taking the average of overlapping areas.

Due to the consideration for illustration purposes and given the processing time limitations of the classification software, a subset of the mosaicked image with different land covers including wild parsnip, trees, roads, grass, and shrubs was selected and used in the following tests and analyses. The area of the selected subset is 52 meters multiple 70 meters. Figure 3.4 shows the final image mosaic for the study area and selected subset. Areas identified as a, b, c, d, and e in Figure 3.4 correspond to the enlarged images of Figure 3.5.

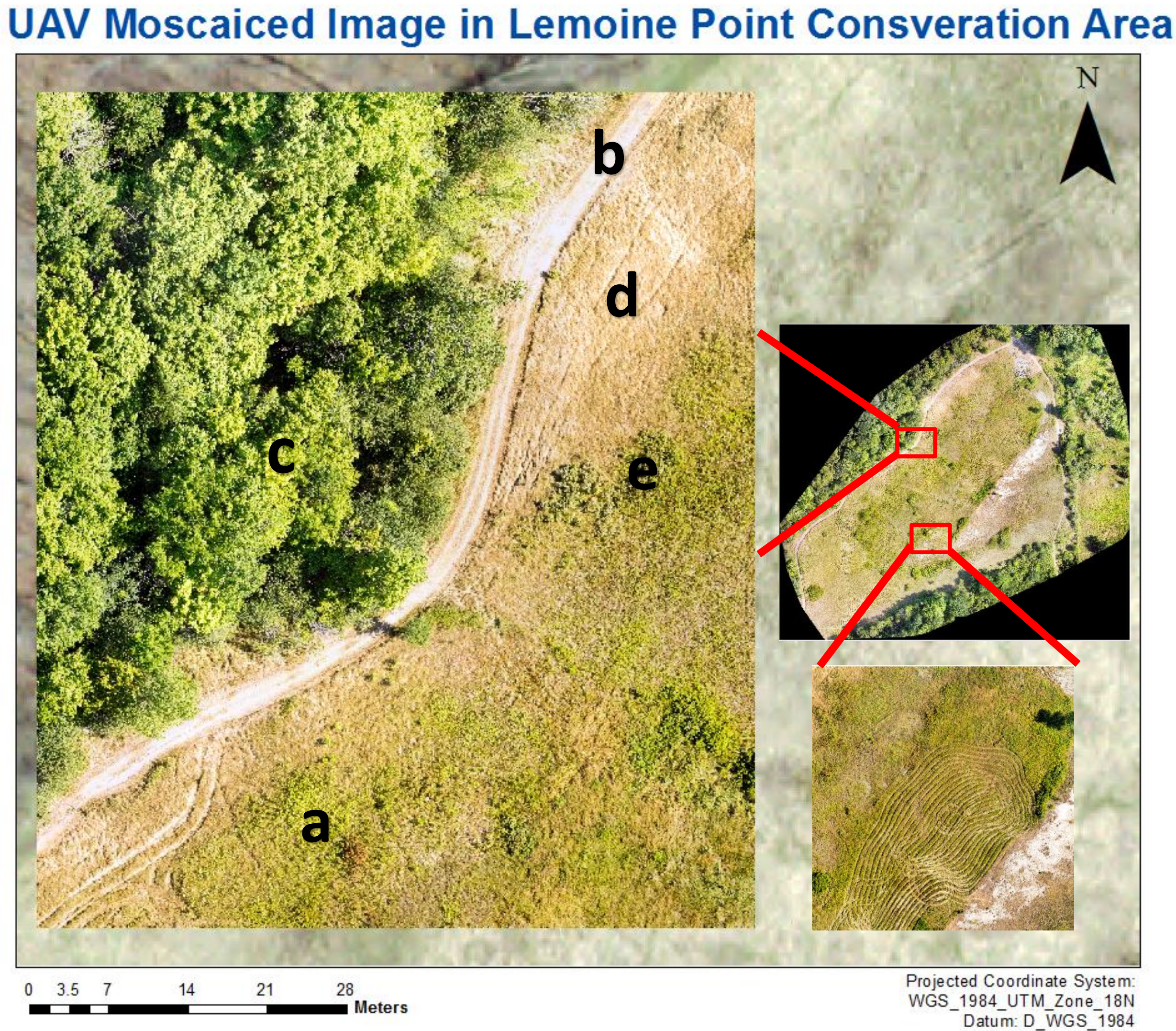
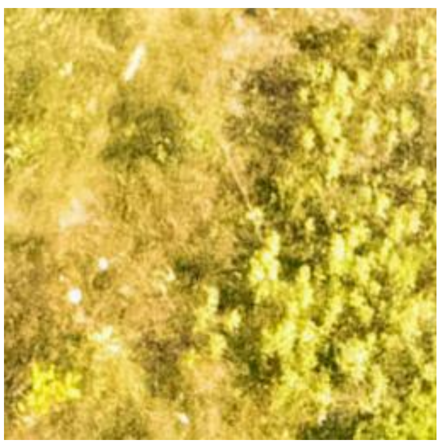


Figure 3.4 The UAV mosaicked image in study area (abcde are enlarged in Figure 3.5)

Field data were collected on 5th September 2016 to provide training samples for the classification processes. Five different land cover types were identified in the study area including wild parsnip, trees, roads, shrubs, and grass. Some flowering and other vascular plants, such as Ragweed and Milkweed, growing in the study area were classified into shrubs because they were not the interest of this study. Some easily recognized polygons of each land cover type were first drawn in the UAV mosaicked image by selecting the region of interest in ENVI 5.2. I went to the study area on 5th September 2016, for the first time to investigate each polygon and record their land cover types based on their coordinate locations with GPS. On 25th September 2016, I went to the study area to re-check all selected polygons and land cover types. After that, the polygons with land cover types were finally confirmed on UAV images. Since wild parsnip was the main study object, the number of the polygons covering wild parsnip were larger than other land cover types. 33 polygon samples for training were collected for wild parsnip, and 10, 3, 19, and 24 polygon samples were collected for trees, roads, shrubs, and grass, respectively. Figure 3.5 shows the detailed image examples of each land cover type.



(a)

(b)

(c)

(d)

(e)

Figure 3.5 An illustration of different land cover type examples (a) Parsnip; (b) Road; (c) Trees; (d) Grass; (e) Shrubs

The spatial resolution of raw mosaicked images is 0.94cm. One of the objectives of this study was to identify a suitable image resolution range to distinguish parsnip. The higher image resolution provides more detailed land cover features, but some redundant detailed information would result in more complicated processing procedures, which is unnecessary. On the other hand, the coarser image resolution would decrease the ability to distinguish different land covers. Therefore, the impact of image resolution on distinguishing parsnip was explored in order to identify the appropriate pixel size (or ground sampling distance (GSD)) of images that can provide enough information for parsnip distinction, and reduce unnecessary processing time. ENVI 5.2 provided the tool that could easily adjust the image resolution defned by the user. The testing images in this study were resampled into 0.01m, 0.02m, 0.03m, 0.05m and 0.1m separately. The resampling of the images was processing in ENVI by the nearest neighbor method.

3.2 Feature Extraction

3.2.1 Vegetation Indices

Three standard mathematical VIs based on visible bands were tested for separating non-vegetated and vegetated areas. Generally, the majority of VIs commonly used in remote sensing are calculated based on visible and near-infrared bands, such as NDVI (normalized difference vegetation index). The number of VIs only based on visible bands are limited. Three VIs including NGBDI (Normalized Green-Blue Difference Index), EXG (Excess Green), and VEG (Vegetation) are calculated using ENVI 5.1 software (Gitelson et al., 2002; Woebbecke et al., 1995; Hague et al., 2006). The equations are listed as below:

$$NGBDI = \frac{\mathrm{G-B}}{\mathrm{G+B}} \quad (1)$$

$$EXG = 2G - R - B \quad (2)$$

$$VEG = \frac{\mathrm{G}}{R^{\mathrm{a}}B^{(1-\mathrm{a})}} \quad \text{with a equals 0.667} \quad (3)$$

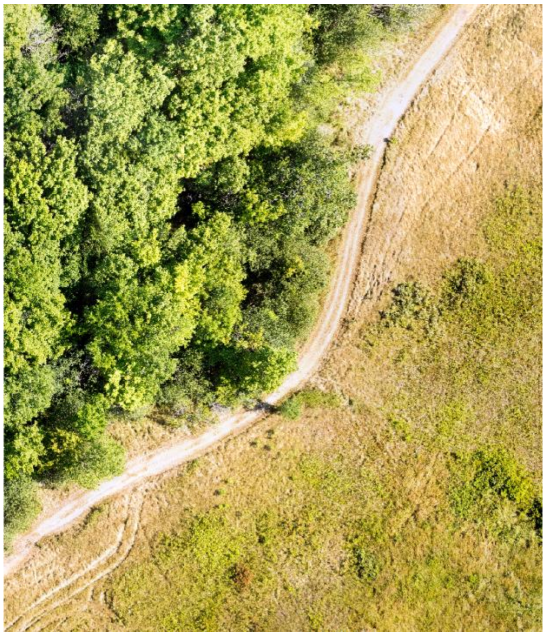
Each calculation applied on UAV images transforms a combined three-band image to a single greyscale band. In greyscale images, the values of pixels in vegetated areas shows greater intensity levels than non-vegetation areas. Therefore, a threshold could be applied on vegetation indices to separate vegetated and non-vegetated area. These three vegetation indices were applied as input features for further classifying procedures to test the capability of wild parsnip classification.

3.2.2 Masked Image

Canopy Height Model (CHM) represents the tree over-story height above the topographic surface, and values equal the Digital Elevation Model (DEM) subtracted from the Digital Surface

Model (DSM) (Balzter, Rowland & Saich, 2007). Correlator3D automatically generates dense DSM and point clouds from input imagery, and filters non-ground pixels through a module to extract a DEM. Due to the height differences between wild parsnip, and trees and part of high bushes, only pixels with the chance to grow wild parsnip remained. Based on the field survey and initial tests, pixels with CHM higher than 1.5m were excluded in this procedure.

Since wild parsnip only grows in vegetated area, non-vegetated areas could be excluded from further study. For each image, a threshold of particular VI could be set to distinguish vegetation and non-vegetation. If the VI value of a pixel was lower than the threshold, then it was placed to the non-vegetated category. Since the NGBDI band showed the biggest differences between vegetation and non-vegetation areas among these three VIs images, a threshold of 0.2 was applied in NGBDI image after initial testing and checking. If NGBDI is equal to or lower than 0.2, the land cover areas were treated as non-vegetation. On the other hand, if NGBDI is greater than 0.2, the land cover areas were treated as vegetation. Therefore, the cover mask with NGBDI larger than 0.2 was applied for further image processing. The new masked RGB image without high vegetation and non-vegetation area is illustrated in Figure 3.6.



(a) (b)

Figure 3.6 (a) original RGB image; (b) masked RGB image

3.2.3 Texture Features

Texture feature applications have been applied in remote sensing for classifying vegetation types, especially used as additional information to compensate the lack of near-infrared band in UAV image applications (Feng, Liu & Gong, 2015). Second-order texture metrics based on pairs of pixels obtained from gray-level co-occurrence matrix (GLCM) are one of the commonly used metrics in features analysis (Haralick & Shanmugam, 1979; Materka & Strzelecki, 1998). Three least correlated texture features include mean, variance, and entropy. These were chosen in this study. The calculations all used the green band to reduce redundancy since RGB bands were highly correlated. The equations were applied as below.

$$Entropy = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i,j) \times \log_2[p(i,j)] \quad (1)$$

$$Mean = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} i \times p(i,j) \tag{2}$$

$$Variance = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - Mean\ i)^2 \times p(i,j) \tag{3}$$

The mean represents the average intensity of the texture within the image; the variance explains the variation of intensity around the mean; and the entropy describes the measure of histogram consistency (Materka & Strzelecki, 1998). G is the number of grey levels, p (i,j) is the matrix estimates of the joint probability between pairs of pixels. The calculation procedure for texture features is based on a moving window from pixel to pixel. Usually, the larger window size, the coarser the information obtained. The accuracy of the texture feature calculation is related to the appropriate scale that collects pixels within the same class and separates pixels from different classes. Since wild parsnip is an irregular and small object, only smaller moving window sizes (3*3, 5*5 and 7*7) were selected to test. The most appropriate 3*3 moving window size was chosen due to the texture feature results comparisons between three different window sizes. Texture features based on a 3*3 window size were added to further classification procedure as additional ancillary bands.

3.3 Image Classification

3.3.1 Random Forest

Random forest as a machine learning method has advantages of running quickly on large data sets, handling abundant input variables, estimating the importance of ranking input variables, and measuring the prediction error (Rodriguez-Galiano et al., 2012). The procedure combines a set of uncorrelated Classification and Regression Trees (CART) and distributes all involving decision tree results to classify each training sample. Two main steps are involved in RF classifications: bootstrap strategy and decision tree node splitting positions. Bootstrap strategy is repeatedly selecting random samples with replacement of the training set. The number of repetition is defined by the user (Dietterich, 2000). Decision tree node splitting positions are also defined by the user. The positions depend on the number of predictors (Dietterich, 2000). Two parameters need to be defined in the RF procedure: ntree and mty (Breiman, 2001). The ntree is the number of trees grown in the calculation (Breiman, 2001). Usually, the larger ntree determined, the lower out-of-bag (OOB) error generated. This algorithm provides a function to plot OOB error and ntree to confirm the ntree selection. Figure 3.7 shows the plot between OOB error rates and the number of trees. OOB error decreased from 10% to 6.5% as ntree increased from 1 to 200. The OOB error rate then became quite stable after ntree reached 1300. Ntree was set to 1300 instead of 200, because it only took approximately one additional minute by adding 800 trees with more accurate and reliable results. The mty is the number of randomly selected predictor variables (Breiman, 2001). The method to calculate mty in this algorithm is the square root of the total input variables numbers. In this study, if the input variables were only three RGB bands, mty equals $\sqrt{3}$; if the input variables were RGB plus six testing features, mty equals $\sqrt{9}$; if the input variables were RGB plus two most significant features, mty equals $\sqrt{5}$ (See Appendix ABC for python code random forest).

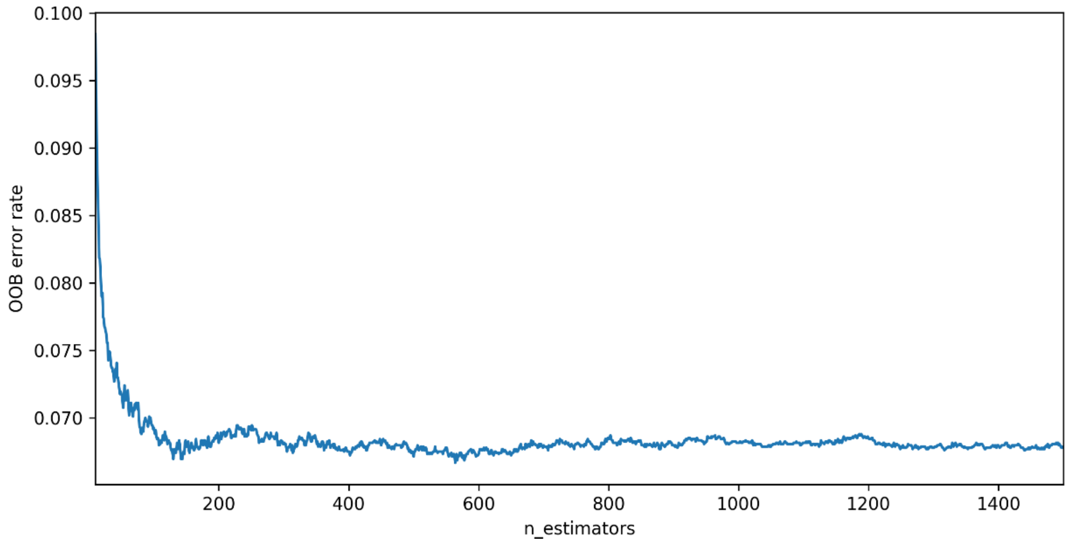


Figure 3.7 The plot of OOB error rate VS n_estimators

3.3.2 Maximum Likelihood & Support Vector Machine

Since RF is a new supervised learning algorithm applied in remote sensing, the traditional remote sensing methods like maximum likelihood (ML) and support vector machine (SVM) can be a benchmark to help assess the ability of RF. The maximum likelihood classification algorithm is one of the most common classifiers for supervised classification. ML algorithm uses Bayes' theory to calculate the probability allocations for each class, typically assigning each pixel into a class (Erdas, 1999). It is a parametric classifier which classifies each pixel into a class based on a Gaussian probability density function (pdf) model (Paola & Schowengerdt, 1995). Each pixel is assigned to the class with the highest probability value (Otukei & Blaschke, 2010). The advantages of the ML are that it calculates variance-covariance during the classification for normally distributed data; however, the results may not be satisfied for non-normal distributed data (Erdas, 1999; Otukei & Blaschke, 2010).

SVM is another supervised algorithm for classification and regression. Unlike ML classifiers, the SVM is a non-parametric classifier. Vapnik & Chervonenkis (1971) first proposed this theory. SVM gives each training sample to one or the other class, and builds a model to assign a new training sample to one or the other class, which makes it a non-probabilistic binary linear classifier (Vapnik, 1999). The objective of this classifier is to determine a class for each new training sample. SVM is a complex method compared to ML and RF, since the kernel, and the parameters for the kernel, and the method to generate SVM are all chosen by users. The successful studies of applying SVM in remote sensing have been illustrated in many studies (e.g. Huang et al., 2002). In this study, both ML and SVM were processed in ENVI Classic 5.2.

3.3.3 Object-based Image Classification

Since all the methods described above were pixel-based, this thesis also compared the pixel-based method with the-state-of-art object-based classification method. Pixel-based image classification is based on the information on each single pixel, but object-based image classification is based on the information from some similar pixels named objects. Each object contains some similar pixels that measured texture features, sizes, shapes, and spectral information (Burnett & Blaschke, 2003). Object-based image classification also needed to choose the classifier, and RF was selected because this study mainly wanted to test the ability of RF in wild parsnip distinction. The object-based RF classification procedure was run in python, but only a smaller sized image was tested because the capability of my python code was limited. The raw image size was 1312 * 679 with the image resolution of 0.03m.

The object-based RF classification procedure could be divided into four parts: image segmentation, feature selection, RF classification, and accuracy assessment. In this study, the first step adopted Felzenszwalb segmentation to divide the image into segments (Boykov & Funka-Lea, 2006). The second step calculated features including min, max, mean, variance, skewness, and kurtosis. The third step applied RF based on the training samples as the classifier to distribute different objects. The last step applied the same validation samples as pixels to create a confusion matrix to acquire accuracy results. The python code is listed in Appendix D.

3.4 Output Assessment & Accuracy Evaluations

The step of validation provided the necessary assurance for the assessment of classification results. Validation data were selected using randomly sampling method. The advantage of random sampling is that each class has an equal chance to be selected. Random point selections guaranteed the samples were chosen unbiasedly and were representative of the class proportion (Chen and Wei 2009). Each selected point was visually inspected and labelled to a class based on its color and texture on original image. Most points can be easily distinguished visually with the help of wild parsnip data provided from the Conservation Authority. Points on the edge of different land cover that cannot be clearly identified were discarded. In total, 4120 randomly selected pixels were used as validation (reference data) in the accuracy assessment, among which 2510, 790, and 820 samples were identified as grass, parsnip and shrub on the ground, respectively. A confusion matrix was created for each classification to illustrate the classification accuracy through Kappa index, overall accuracy (OA), producer accuracy (PA) and user accuracy (UA) based on validation samples. The optimal image resolution and variables selection were assessed to determine the most suitable parsnip classification workflow.

After the optimal parsnip classification workflow was determined, the selected workflow was applied to another mosaicked image subset (Figure 3.8) located in different area to test whether the optimal workflow was able to distinguish wild parsnip.

Figure 3.8 UAV Mosaicked Image for Validation in Lemoine Point Conservation Area

Chapter 4

Results and Discussions

4.1 Comparison of Three Pixel-based Classifiers

Random forest (RF), maximum likelihood (ML), and support vector machine (SVM) are three commonly used supervised learning algorithms in remote sensing. This study applied these three algorithms to identify and distinguish wild parsnip. To statistical analyze the classification results between different image resolutions under different input variables selected, a confusion matrix was generated during the validation procedure to assess the accuracy of classification results. The reference data samples were selected randomly, where the number of points in each class depended on the proportion of each class on images.

The testing procedures used RGB-only, RGB+2 (RGB plus two features), and RGB+6 (RGB plus six features) images as input, under five different image resolutions (0.01m, 0.02m, 0.03m, 0.05m, and 0.1m), with three different pixel-based classifiers (RF, ML, and SVM). It was not possible to show all 45 maps in this thesis, classification results for the 0.03m RGB+2 image (higher accuracy) under three classifiers are presented in Figure **4.1**. To show the differences between raw image and classification result in detail, one small area of the raw UAV image and three classification results are shown in Figure **4.1**. The yellow parts in Figure **4.1** (b), (c), (d) were classified into wild parsnip. As we can see, the wild parsnips were distinguished well from other land covers. RF in particular provided more smooth results. In Figure **4.1** (b) and (c), some grasses and shrubs were misclassified into wild parsnip, whereas in Figure **4.1** (d), the misclassifications of wild parsnips were significantly decreased.

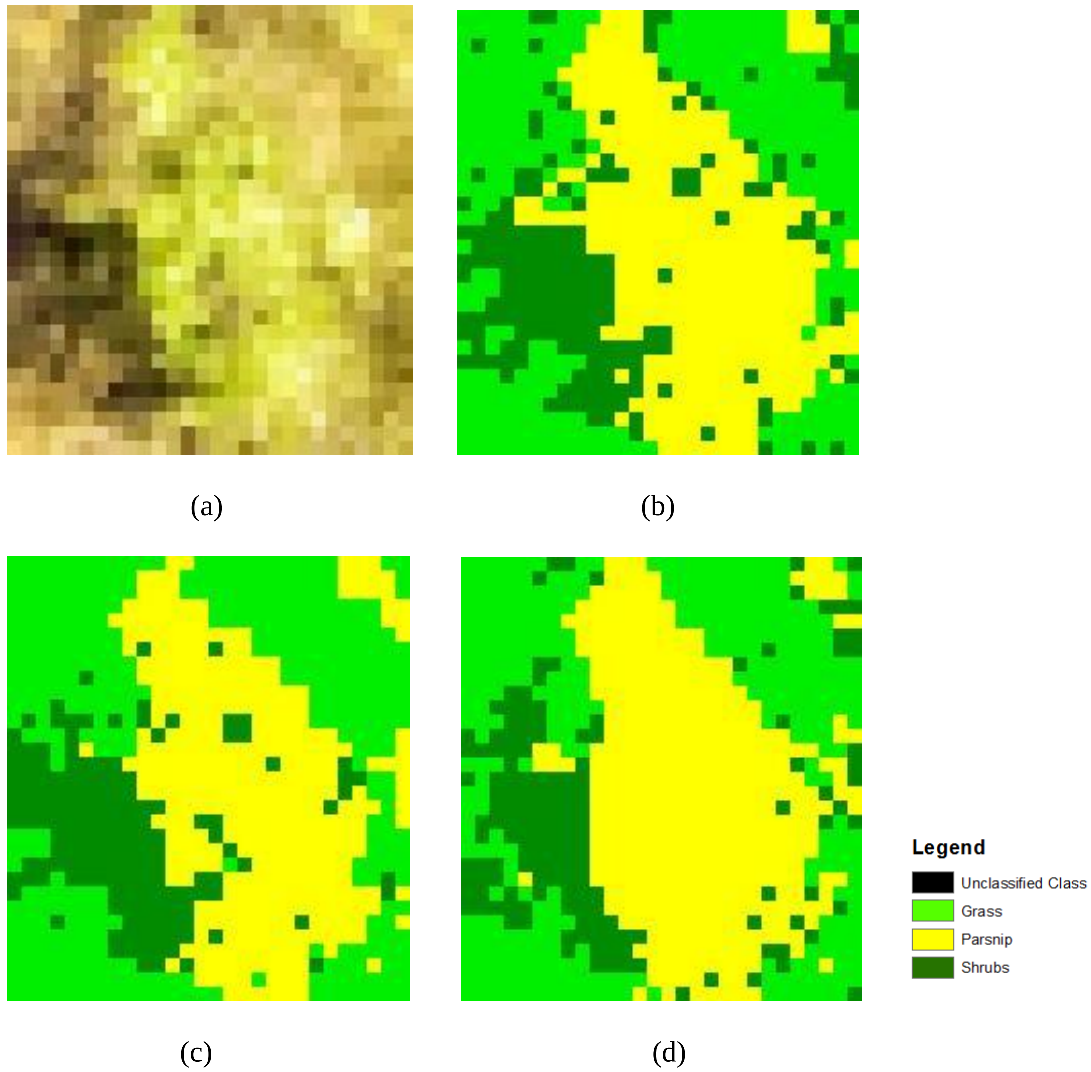
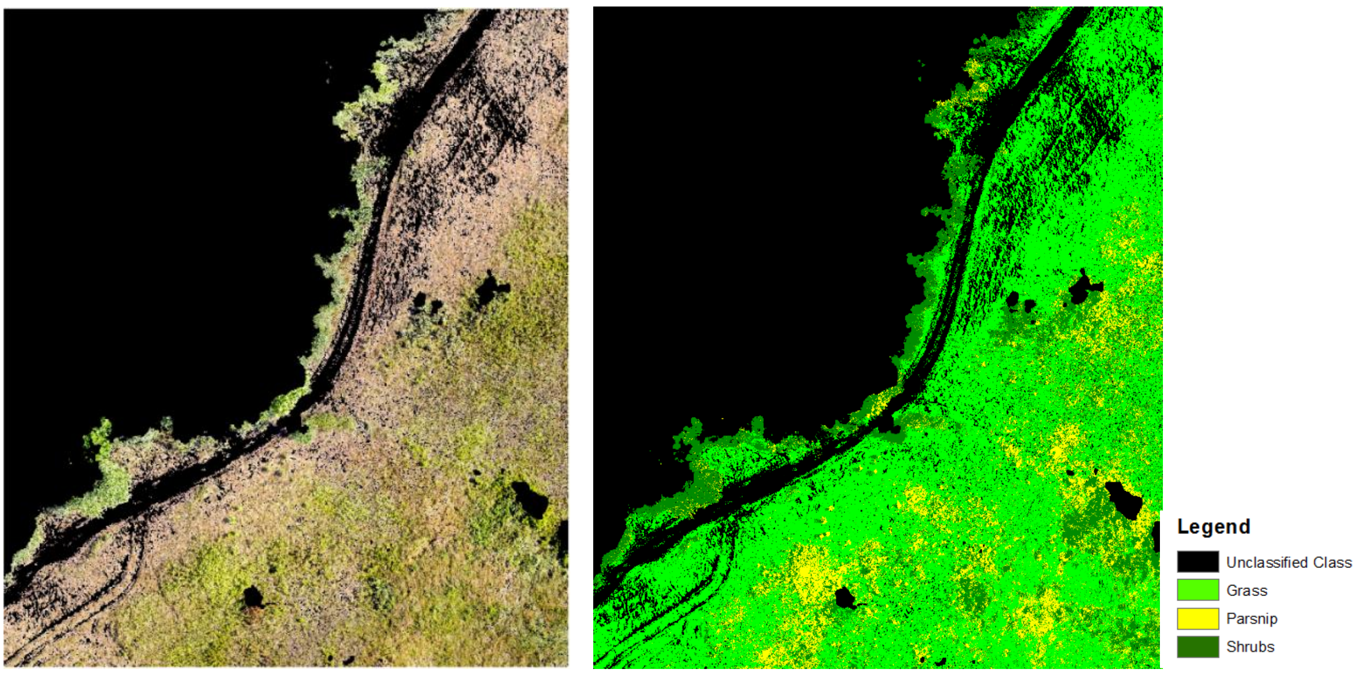


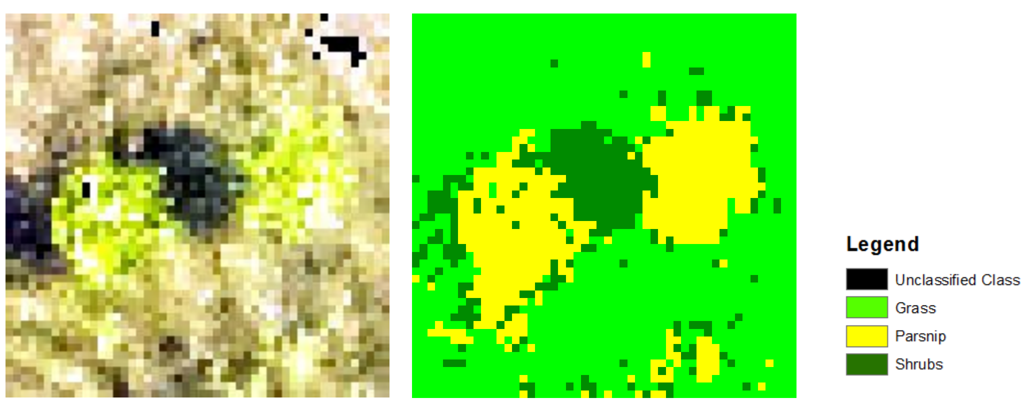
Figure 4.1 Detailed 0.03m RGB+2 Image Classification Results (a) Raw UAV Image (b) ML Classification Result (c) SVM Classification Result (d) RF Classification Result

Since the number of wild parsnips were not the majority in the image, the visual differences on wild parsnips in the whole image were hard to see between three classification results. RF classification results produced the most accurate results; therefore, the whole UAV masked image and the RF classification result of 0.03m RGB+2 image are shown in Figure 4.2. The masked image was the raw UAV mosaic image excluding roads, trees, and high shrubs. To show the differences between UAV masked image and classification results in detail, one small area shows the RF classification results in detail in Figure 4.3. As Figure 4.3 shows, the wild parsnips, even grown as single plants, were distinguished very well from other land covers. Some part of shadows in the image might be misclassified into shrubs or grass, but it did not affect distinguishing wild parsnip from other land covers.



(a) (b)

Figure 4.2 The Comparison between RGB Masked Image and 0.03m RGB+2 RF Classification Image (a) UAV Masked Image (b) RF Classification Results



(a) (b)

Figure 4.3 Detailed 0.03m RGB+2 Image Classification Results (a) UAV Masked Image (b) RF Classification Result

The overall accuracy (OA) under different image resolutions with different input variables by different pixel-based classifiers are shown in Table 4-1. As the table shows, RF achieved higher accuracy results than other classifiers under the same input situation. No matter how many variables were inputted, RF outperformed ML and SVM under the same resolution image. No matter how many resolution images were selected, RF outperformed ML and SVM under the same input variables. Especially, the highest OA was the 0.03m RGB+6 image using RF as the classifier, and the second highest OA was the 0.03m RGB+2 image using RF as the classifier. It was not quite meaningful to test RGB+6 images with ML as classifiers, because RF performed much better than ML based on the tests. In addition, SVM was a time-consuming classifier, and the OA was obviously lower than RF, so SVM under RGB+6 images were also not meaningful to test.

Table 4-1 Overall Accuracy (OA) between Three Classifiers under Different Image Resolutions with Different Input Variables

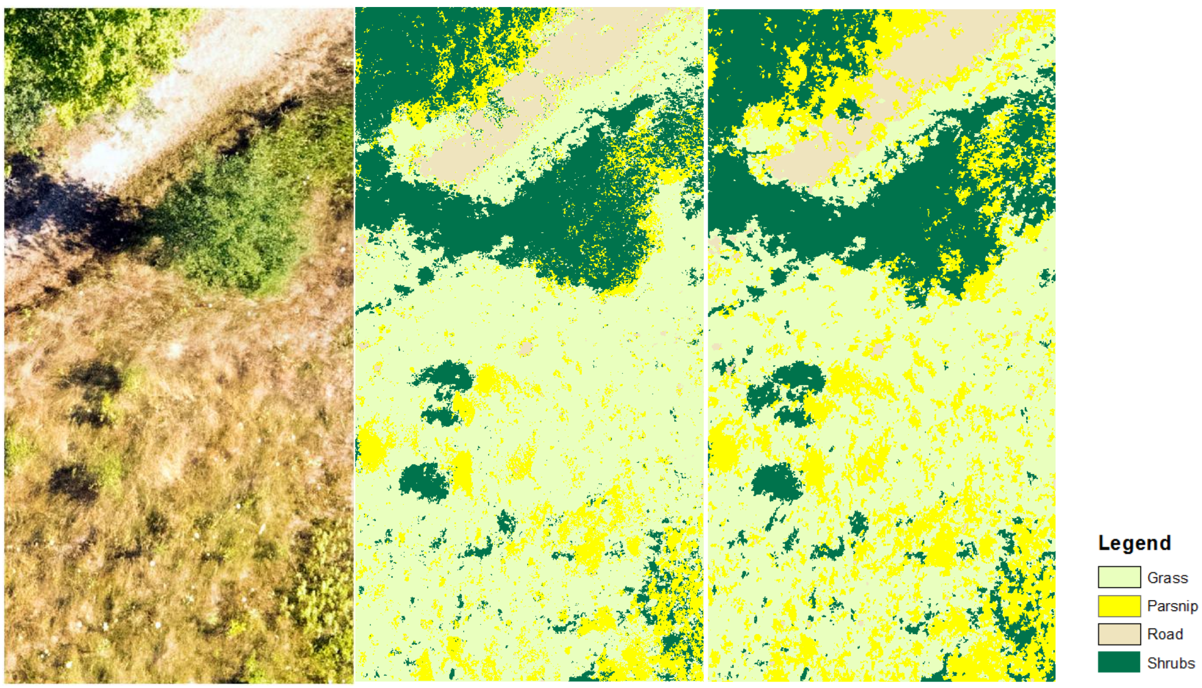
OA (%)		RGB-only	RGB+2	RGB+6
ML	0.01m	65.48	70.20	N/A
	0.02m	72.35	76.79	N/A
	0.03m	66.50	72.33	N/A
	0.05m	65.50	71.67	N/A
	0.1m	60.89	68.23	N/A
RF	0.01m	74.35	73.54	79.26
	0.02m	78.22	78.64	83.12
	0.03m	81.11	**84.70**	**86.45**
	0.05m	77.68	80.58	80.62
	0.1m	70.22	74.76	73.69
SVM	0.01m	68.23	75.63	N/A
	0.02m	74.33	78.51	N/A
	0.03m	70.86	75.24	N/A
	0.05m	69.32	73.68	N/A
	0.1m	61.09	70.09	N/A

Table 4-1 illustrates that RF classifier performed better than traditional ML and SVM classifiers. Overall accuracy was increased by 12.37% and 9.46% over ML and SVM under 0.03m RGB+2 images, separately. Since RF is the new supervised learning algorithm applied in remote sensing, the traditional remote sensing methods such as ML and SVM can be a benchmark to help assess the ability of RF. ML used Baye's theory that assumes each input variable is normally distributed; however, this assumption is rarely met when analyzing real data (Vapnik, 1979). The RF algorithm outputted the results with the highest number of votes from decision trees, which eliminated the bias in classifying procedure. SVM was more complicated than RF because of the parameter settings, and SVM was a binary classifier, which was more suitable for pairwise classifications (Mountrakis et al., 2011). Also, RF only needed several minutes to process, but SVM needed one hour per processing time.

4.2 Comparison of Pixel-based VS Object-based RF Classifications

Previous analysis was all based on pixels on the image, but the state-of-art method in remote sensing was also common to use classification based on objects. Object-based classification works better than pixel-based classification in regular or large patches of land cover type under different image resolutions (Robertson & King 2011); however, the capability of object-based image classification in irregular and small land cover type still needs to be demonstrated, like parsnip. Based on previous tests, RF was the most appropriate classifier in distinguishing wild parsnip. Therefore, the comparison between pixel-based and object-based by using RF classifier with same input images, same training samples, and validation samples were analyzed in this study.

Figure 4.4 shows the detailed comparison results between pixel-based image classification and object-based image classification. Inaccurate classification results for parsnip are obvious in the object-based image classification, simply based on a visual inspection. The object-based image classification did not apply a quantitative approach to assess the result, because the misclassification can be clearly seen on the image between pixel-based results and object-based results. After visual inspection of the classification results of several land cover types between object-based and pixel-based, we determined that many grass covered areas was falsely classified as parsnip in the object-based image classification, indicating it was not suitable for irregular and small size land cover types. The object-based classification results for road surfaces demonstrates better performance than pixel-based, which means object-based was more appropriate for large continuous land cover types. On the other hand, pixel-based did not need image segmentation, which was simpler and saved processing time. Therefore, pixel-based image classification was shown to have better performance in small and single plants, whereas, object-based image classification was shown to have better achievements in large land covers.



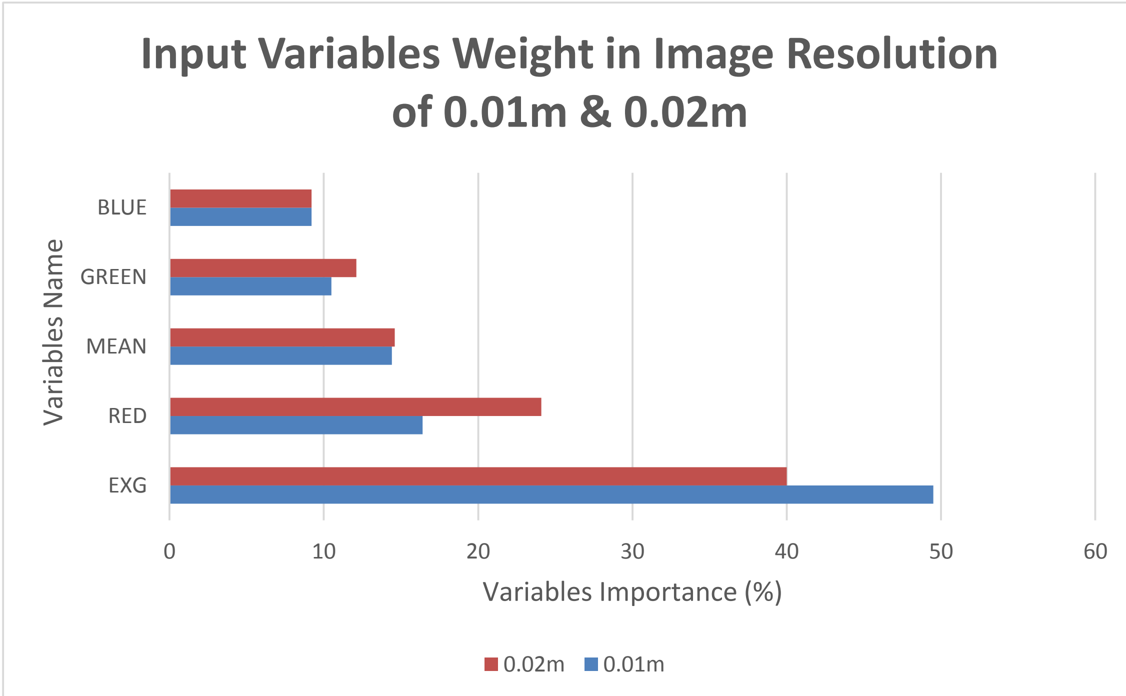
(a) (b) (c)

Figure 4.4 Classification Results Comparison between Pixel-based and Object-based RF(a) UAV Raw Image (b) Pixel-based RF Classification Results (c) Object-based RF Classification Results

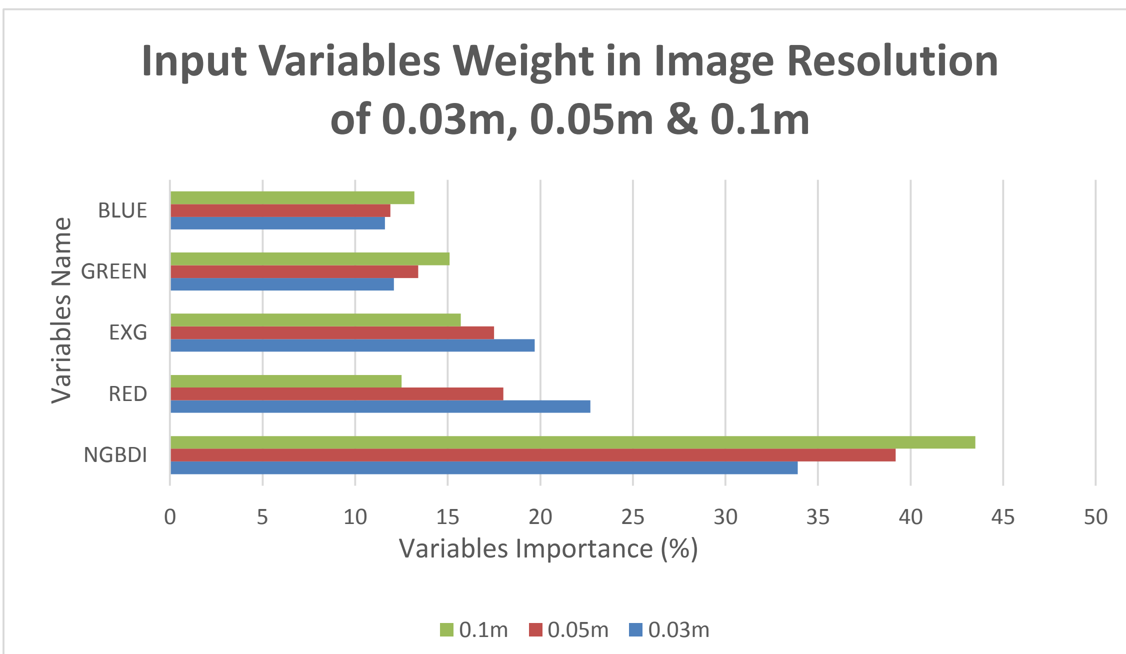
4.3 The Impact of Input Variables

After selecting the most suitable classifier, the significance of input variables is also evaluated. RF provides a function of input variables important rankings, which can be used to evaluate the contribution of different input variables to classification. The input variables included three original visible bands (red, green and blue), three vegetation indices (NGBDI, EXG, and VEG), and three texture features (mean, variance, and entropy). Since one of the objectives of this study was to find the simplest and optimal solution to distinguish wild parsnip, nine variables were identified too complicated to process in practical applications. In order to simplify the procedure, we choose the five most important variables. According to the RF variable importance rankings, RGB were always ranked in the top five, and the other two important variables varied based on spatial resolution.

Based on Figure 4.5(a), if image resolution was lower than 0.03m, the five most important variables were Red band, Green band, Blue band, EXG and Mean; if image resolution was equal or greater than 0.03m (Figure 4.5(b)), the five most important variables were Red band, Green band, Blue band, NGBDI, and EXG. When the image resolution was lower than 0.03m, EXG had a greater importance than other variables. Mean was also important since the mean value could reduce the variance in the extremely high-resolution image. When the image resolution was equal or bigger than 0.03m, texture features were not among the five most important variables. NGBDI was the most important variable because the blue band was not as necessary in a coarser image. Therefore, besides the three original visible bands, EXG and Mean features were added as additional bands when image resolution was lower than 0.03m; NGBDI and EXG were added as additional bands when image resolution was equal or larger than 0.03m.



(a)



(b)

Figure 4.5 Input Variables Weight Comparison (a) Image Resolution of 0.01m & 0.02m (b) Image Resolution of 0.03m, 0.05m & 0.1m

The OA comparison between RGB-only and RGB+2, before and after including additional variables, under pixel-based RF, is shown in Table 4.1, and also emphasized in Figure 4.6. Under each different image resolutions, the average OA increased 2.9% from RGB-only to RGB+2. In image resolution 0.03m, the significant increment of 4.2% from RGB-only to RGB+2, was verified that the inclusion of additional variables could improve classification accuracy. Meanwhile, kappa value was also increased by 0.1 after adding additional variables under image resolution 0.03m (Figure 4.7).

Figure 4.6 OA Comparison between with (RGB+2) and without (RGB only) Additional Variables

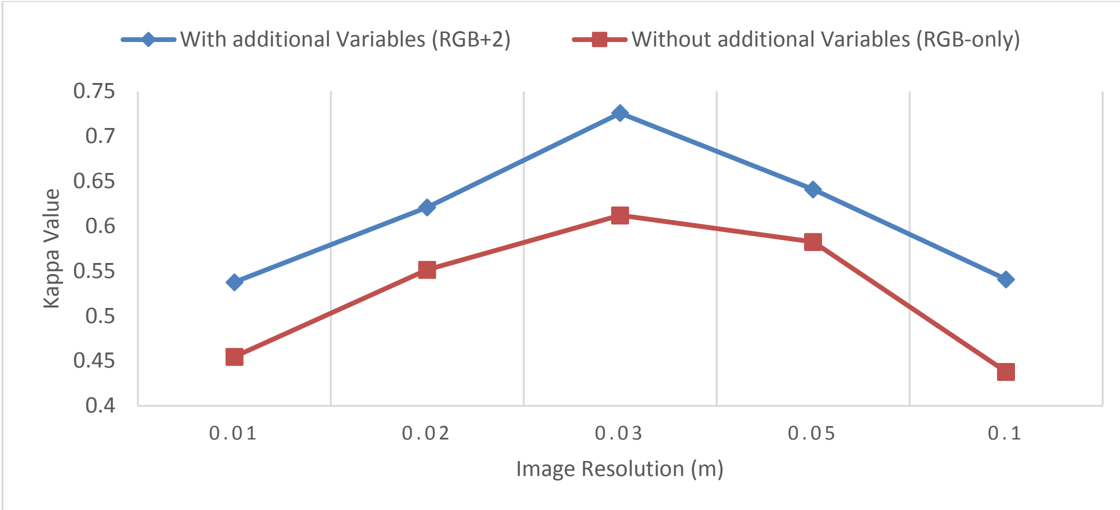


Figure 4.7 Kappa Value Comparison between with (RGB+2) and without (RGB only) Additional Variables

4.4 The Impact of Spatial Resolution

The impact of image resolution on classification was examined. One advantage of UAV images is their very high spatial resolution; however, not the higher resolution image obtained, the better classification result could achieve. The higher resolution image needs more preprocessing and processing time because of the larger file size; therefore, the classification accuracy at different image resolutions was compared. As shown in

Figure 4.8 & Figure 4.9, with the five important input variables by using pixel-based RF classifier, the OA reached the highest (84.70%), as well as Kappa that reached the highest (as 0.7258) at the image resolution of 0.03m. Meanwhile, the OA was increasing from 73.54% to 84.70% in image resolution from 0.01m to 0.03m, and decreasing from 84.70% to 74.76% in image resolution from 0.03m to 0.1m. The kappa was increasing from 0.5375 to 0.7258 in image resolution from 0.01m to 0.03m, and decreasing from 0.7258 to 0.5407 from 0.03m to 0.1m.

Therefore, the optimal image resolution for distinguishing parsnip was around 0.03m. Since wild parsnip was growing in irregular and small shapes, 0.03m was an intermediate value for both wild parsnip growing status in large patches and single plants.

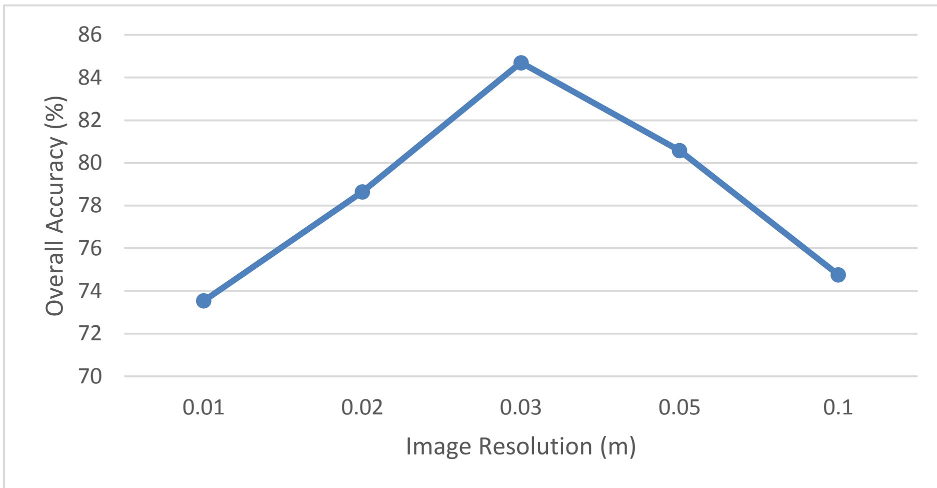


Figure 4.8 OA under Different Image Resolutions with RGB+2 Variables using Pixel-based RF Classifier

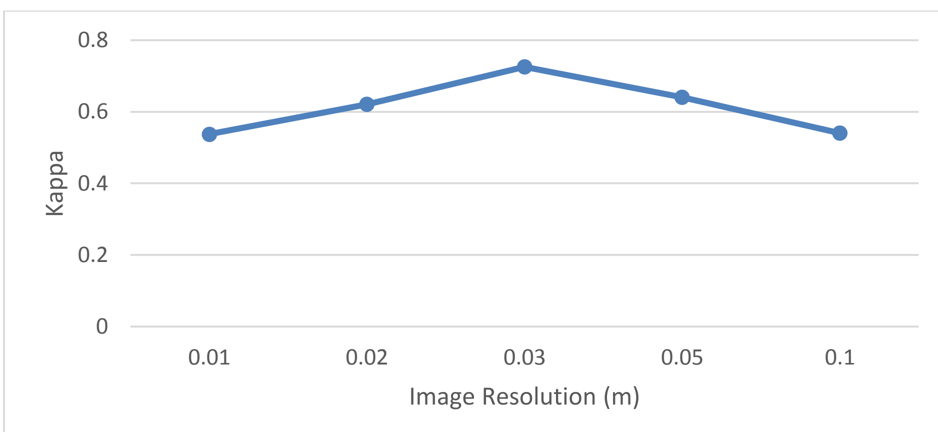


Figure 4.9 Kappa Value under Different Image Resolutions with RGB+2 Variables using Pixel-based RF Classifier

Besides OA and kappa value, the other important information in the confusion matrix were accuracy (UA), producer accuracy (PA), Omission Errors (OE), and Commission Error (CE) of each land cover class, which are shown in Table 4-2, Table 4-3, Table 4-4, Table 4-5 and Table 4-6. If only the parsnip is considered, the classification image of 0.03m resolution has the highest accuracy for both UA and PA. UA increased from 55.74% to 74.42% when the image resolution changed from 0.01m to 0.03m, and decreased from 74.42% to 51.22% at image resolution from 0.03m to 0.1m. PA increased from 43.04% to 81.01% when the image resolution changed from 0.01m to 0.03m, and decreased from 81.01% to 53.16% at image resolution from 0.03m to 0.1m.

Table 4-2 Error matrix for accuracy assessment at Image Resolution 0.01m

	Grass	Parsnip	Shrub	Total	User's Accracy	Commission Error
Grass	2120	30	160	2310	91.77	8.23
Parsnip	180	340	90	610	55.74	44.26
Shrub	210	420	570	1200	47.5	52.5
Total	2510	790	820	4120		
Producer's Accuracy	84.46	43.04	69.51		Overall Accuracy 73.54%	
Omission Error	15.54	56.96	30.49		Kappa 0.538	

Table 4-3 Error Matrix for Accuracy Assessment at Image Resolution 0.02m

	Grass	Parsnip	Shrub	Total	User's Accracy	Commission Error
Grass	2200	130	70	2400	91.67	8.33
Parsnip	240	460	170	870	52.87	47.13
Shrub	70	200	580	850	68.24	31.76
Total	2510	790	820	4120		
Producer's Accuracy	87.65	58.23	70.73		Overall Accuracy 78.64%	

Omission Error	12.35	41.77	29.27		Kappa 0.621

Table 4-4 Error Matrix for Accuracy Assessment at Image Resolution 0.03m

	Grass	Parsnip	Shrub	Total	User's Accracy	Commission Error
Grass	2270	80	110	2460	92.28	7.22
Parsnip	90	640	130	860	74.42	25.58
Shrub	150	70	580	800	72.5	27.5
Total	2510	790	820	4120		
Producer's Accuracy	90.44	81.01	70.73		Overall Accuracy 84.7%	
Omission Error	9.56	18.99	29.27		Kappa 0.726	

Table 4-5 Error Matrix for Accuracy Assessment at Image Resolution 0.05m

	Grass	Parsnip	Shrub	Total	User's Accracy	Commission Error
Grass	2220	150	260	2630	84.41	15.59
Parsnip	200	570	30	800	71.25	28.75
Shrub	90	70	530	690	76.81	23.19
Total	2510	790	820	4120		
Producer's Accuracy	88.45	72.15	64.63		Overall Accuracy 80.58%	
Omission Error	11.55	27.85	35.37		Kappa 0.641	

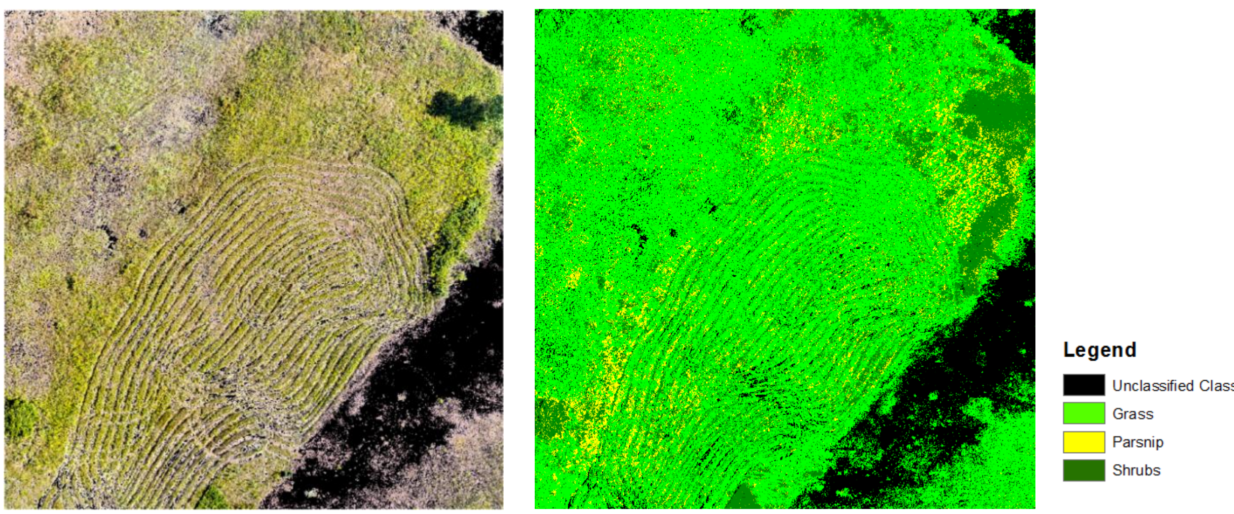
Table 4-6 Error Matrix for Accuracy Assessment at Image Resolution 0.1m

	Grass	Parsnip	Shrub	Total	User's Accracy	Commission Error
Grass	2070	310	160	2630	81.50	18.50
Parsnip	330	420	70	800	51.22	48.78
Shrub	110	60	590	690	77.63	22.37
Total	2510	790	820	4120		
Producer's Accuracy	82.47	53.16	71.95		Overall Accuracy 74.76%	
Omission Error	17.53	46.84	28.05		Kappa 0.541	

4.5 Application of the Optimal Workflow for Another Image

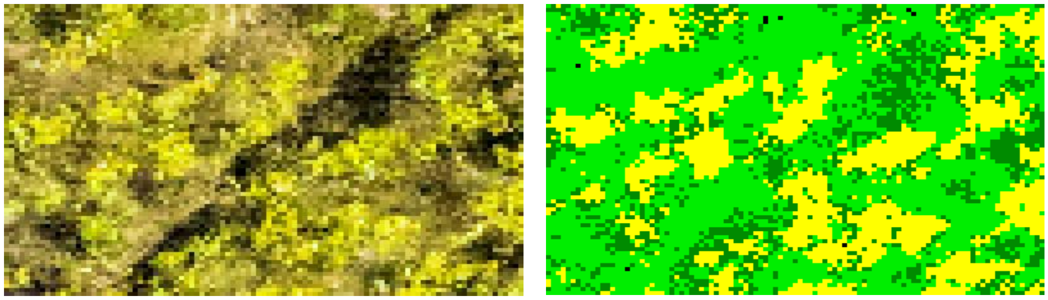
The optimal workflow to distinguish wild parsnip from other land covers was explored, by applying RF as the classifier with 0.03m resolution image as input, which contained Red, Green, Blue, NGBDI and EXG features, and excluded non-vegetation areas and trees. Another mosaicked image from different regions of the study area was tested to evaluate the performance of the proposed workflow (Figure **4**.**10**).The training samples were collected on 5th September 2016 and 25th September 2016, which were the same day as previous field work. The image processing was followed the optimal workflow, and the validating samples was using the same way as previous adopted. The whole and part of classification results are shown in Figure 4.11 and Figure 4.12, which obviously distinguished parsnip from other land cover types. The overall accuracy achieved 83.22%, and the kappa value was 0.6587. The user accuracy and producer accuracy of parsnip were 78.57% and 73.33%, respectively. This tested image contained a relatively smaller numbers of wild parsnip, but the accuracy of wild parsnip was still high enough. Hence, this tested image has demonstrated that the created workflow was practical, effective, and accurate.

Figure 4.10 RGB Raw Image



(a) (b)

Figure 4.11 (a) RGB Masked Image (b) Classification Image



(a) (b)

Figure 4.12 Detailed Images (a) RGB Masked Image (b) Optimal Classification Results

Chapter 5

Conclusion and Summary

5.1 Conclusion Overview

Since invasive species may bring harm to urban ecological systems, and affect people, it has become essential to monitor their existence and spread. Wild parsnips are growing in scattered or as single plants in small sizes and irregular shapes; therefore, low-resolution remote sensing images are too coarse for wild parsnip research. UAVs can obtain UHR imagery and have been successfully monitoring both environmental and agricultural areas. Despite UAV's success in mapping invasive plants, mapping wild parsnip in a specific area is largely unexplored. In this study, UAV images of Lemoine Point Conservation Area in Kingston, Ontario, were used to test the methodology of identifying wild parsnip. The primary objective of this study was to develop an efficient and accurate classification workflow for invasive wild parsnip using UHR digital UAV imagery. The main challenge was that all processing options were based only on the visible bands (RGB) from the digital cameras of UAV, which is insufficient in detecting small and irregular vegetation.

This study proposed an optimal and simplified hybrid method with higher accuracy by combining random forest, texture analysis, and vegetation indices to distinguish parsnip in vegetation fields based on UAV remote sensing. The optimal workflow for distinguishing wild parsnip from other land covers was (1) by excluding non-vegetation area and trees, and (2) then applying random forest as the classifier with 0.03m resolution image as input, which contains Red, Green, Blue, NGBDI and EXG features. If the image resolution was lower than 0.03m, the five most important variables were Red, Green, Blue, EXG and Mean; if the image resolution was equal or greater than 0.03m, the five most important variables were Red, Green, Blue, NGBDI, and EXG. The overall accuracy increased from 73.54% to 84.70% under image resolution from 0.01m to 0.03m, and decreased from 84.70% to 74.76% under image resolution from 0.03m to 0.1m. In the 0.03m resolution image, the significant increment of 4.2% verified that the inclusion of additional variables improves the classification accuracy. Moreover, when using Random Forest instead of ML or SVM, wild parsnip classification accuracy increased by about 13% and 8% in average, respectively, which indicates that Random forest outperformed ML and SVM in wild parsnip analyses. When comparing pixel-based to object-based image classification, pixel-based performed better than object-based for small and irregular vegetation. This study demonstrates that UAV is an outstanding platform for vegetation monitoring, and that combined methods can provide accurate classification results and reduce the limitations of off-the-shelf digital cameras.

UAV image preprocessing flow includes image orientation, DEM and DSM extraction, individual orthophoto production, and orthomosaic generation, and these processes were done on Simactive's software. Three vegetation indices and three texture features were calculated and added the first two significant variables as supplementary data to the mosaicked images. The random forest algorithm with relevant variable choices was selected as the classifier to differentiate wild parsnip plants from other vegetation types. The optimal image resolution in parsnip analysis was also discussed by comparing accuracy assessments. The results provided an optimal and executable workflow to distinguish wild parsnip and demonstrated that UAV images are an appropriate and economic resource for small and irregular vegetation types, even if it's only equipped with a simple digital-light camera. The combined method brought reliable and valid outcomes in detecting invasive wild parsnip as well as demonstrated good performance in mapping vegetation.

This research has provided Cataraqui Conservation Authority with an effective workflow to identify and locate wild parsnip with UAV instead of human field surveying. By implementing this workflow, the Conservation Authority can save time and money to quickly identify the locations of wild parsnips and develop strategies for eliminating or controlling its spread.

5.2 Limitations

Even though the workflow performed well, the approach may still have three limitations. The first limitation is related to the spectral resolution of the data. As the data of this research is all from UAV images, there is no infrared bands information. If infrared bands are included, more vegetation indices could be tested. Based on the literature reviews, infrared bands are significant and necessary in vegetation studies. The solution for this limitation is either to use a camera that can capture infrared bands, or apply other high-resolution data with more spectral bands.

Secondly, land cover types in this UAV acquired area did not contain other yellow flowering types of vegetation. This workflow may not be suitable for distinguishing wild parsnip from other similar sized plants with yellow flowers. The resolution of this limitation is to test the workflow identified through this research with images that contain another similar sized yellow flowering plant.

Lastly, since this study did not use any techniques to remove shadows, some shadows are misclassified into shrubs. The confusion between shadows and shrubs would affect the overall accuracy of classification. Although shadows might not bring too much influences on distinguishing wild parsnip in this study, algorithms or techniques for removing or considering shadows should be adopted in the image preprocessing procedure in the future.

5.3 Recommendations for Future Work

Benefits of using UAV images in classifying the invasive plant wild parsnip, are clearly presented in this thesis, and there is potential to build more on this research.

In order to locate wild parsnip more accurately, future data could be collected from the study area that includes other similar yellow flowering plants. Different spatial and spectral information could be calculated on different plants, providing a more precise representation of where detected plants occur on the landscape. To map wild parsnip more accurately, researchers can require additional data on invasive plants from local environmental departments, to cover larger spatial and spectral information.

The UAV images in this research were captured in the summer, in the future, it is recommended that researchers collect plant species data during the early flowering season. This may help identify the number of plants of each species early, and benefit monitoring and management.

Providing accurate maps of wild parsnip locations will reduce and prevent human and livestock injury; therefore, it is also necessary to locate wild parsnip on private lands. Since private lands may include different vegetation, researchers could collect images from private lands to test the efficiency of workflow.

References

Ahmed, A., Nagai, M., Tianen, C., & Shibasaki, R. (2008). UAV based monitoring system and object detection technique development for a disaster area. *International Archives of Photogrammetry, Remote Sensing and Spatial information Sciences*, *37*, 373-377. Alex, J. F., & Switzer, C. M. (1992). Ontario weeds. *Descriptions*, *Illustrations and Keys to their Identification, 2nd edition*. Ontario Ministry of Agriculture and Food Toronto, ON, Canada.

Alparone, L., Argenti, F., & Benelli, G. (1990). Fast calculation of co-occurrence matrix parameters for image segmentation. *Electronics Letters*,*26*(1), 23-24. doi:10.1049/el:19900015

Averill, K. M., & Ditommaso, A. (2007). Wild Parsnip (Pastinaca sativa): A Troublesome Species of Increasing Concern. *Weed Technology*, 21(01), 279-287. doi:10.1614/wt-05-186.1

Balzter, H., Rowland, C., & Saich, P. (2007). Forest canopy height and carbon estimation at Monks Wood National Nature Reserve, UK, using dual-wavelength SAR interferometry. *Remote Sensing of Environment*,*108*(3), 224-239. doi:10.1016/j.rse.2006.11.014

Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, *13*(1-2), 95-120.

Baret, F., Guyot, G., & Major, D. J. (1989). TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation. In *Geoscience and Remote Sensing Symposium, 1989. IGARSS'89. 12th Canadian Symposium on Remote Sensing., 1989 International* (Vol. 3, pp. 1355-1358). IEEE.

Baskin, J. M., & Baskin, C. M. (1979). Studies on the autecology and population biology of the weedy monocarpic perennial, Pastinaca sativa. *The Journal of Ecology*, 601-610.

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., ... & Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation, 39*, 79-87.

Berenbaum, M. R., & Zangerl, A. R. (1997). Phytochemical diversity and redundancy in ecological interactions. *Biochimie,79*(5), 309. doi:10.1016/s0300-9084(97)83519-5

Berenbaum, M. R., Zangerl, A. R., & Nitao, J. K. (1984). Furanocoumarins in seeds of wild and cultivated parsnip. *Phytochemistry,23*(8), 1809-1810. doi:10.1016/s0031-9422(00)83503-7

Berni, J. A., Zarco-Tejada, P. J., Suárez, L., & Fereres, E. (2009). Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(3), 722-738.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2-16.

Boegh, E., Soegaard, H., Broge, N., Hasager, C. B., Jensen, N. O., Schelde, K., & Thomsen, A. (2002). Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sensing of Environment*, *81*(2), 179-193.

Boersma, P. D., Reichard, S. H., & Van Buren, A. N. (2006). *Invasive species in the Pacific Northwest*. University of Washington Press.

Boykov, Y., & Funka-Lea, G. (2006). Graph cuts and efficient ND image segmentation. *International Journal of Computer Vision, 70*(2), 109-131.

Bradley, B. A., & Mustard, J. F. (2006). Characterizing the landscape dynamics of an invasive plant and risk of invasion using remote sensing. *Ecological Applications*, *16*(3), 1132-1147.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees. Wadsworth, Belmont

Bryson, M., Johnson-Roberson, M., & Sukkarieh, S. (2009, May). Airborne smoothing and mapping using vision and inertial sensors. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on* (pp. 2037-2042). IEEE.

Bryson, M., Reid, A., Ramos, F., & Sukkarieh, S. (2010). Airborne vision-based mapping and classification of large farmland environments. *Journal of Field Robotics*, *27*(5), 632-655.

Burnett, C., & Blaschke, T. (2003). A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological modelling*, *168*(3), 233-249.

Burrows, G. E., & Tyrl, R. J. (2001). *Toxic Plants of North America* (No. 581.659097 B8T6).

Cain, N., Darbyshire, S. J., Francis, A., Nurse, R. E., & Simard, M. (2010). The Biology of Canadian Weeds. 144. Pastinaca sativa L. *Canadian Journal of Plant Science, 90*(2), 217-240. Doi: 10.4141/cjps09110

Carlsen, K., & Weismann, K. (2007). Phytophotodermatitis in 19 children admitted to hospital and their differential diagnoses: Child abuse and herpes simplex virus infection. *Journal of the American Academy of Dermatology*, *57*(5).

Cataraqui Region Conservation Authority (2017). *Lemoine Point Conservation Area.* Retrieved July 17, 2017, from https://crca.ca/conservation-lands/conservation-areas/lemoine-point-conservation-area/

Chance, C. M. (2016). *Mapping the distributions of two invasive plant species in urban areas with advanced remote sensing data.* University of British Columbia. Retrieved from https://open.library.ubc.ca/cIRcle/collections/24/items/1.0305687

Chen, D. and H. Wei. (2009). *"The effect of spatial autocorrelation and class proportion on the accuracy measures from different sampling designs".* ISPRS Journal of Photogrammetry and Remote Sensing*64(2):264-282*

Chou, T. Y., Yeh, M. L., Chen, Y. C., & Chen, Y. H. (2010). *Disaster monitoring and management by the unmanned aerial vehicle technology*. IAPRS, Vol. XXXVIII, Part 7B

Clevers, J. G. P. W. (1986). The application of a vegetation index in correcting the infrared reflectance for soil background. In *remote sensing for ressources development and environmental management. International symposium. 7* (pp. 221-226).

Coburn, C. A., & Roberts, A. C. (2004). A multiscale texture analysis procedure for improved forest stand classification. *International journal of remote sensing*, *25*(20), 4287-4308.

Coffman, K., Boyce, W. T., & Hansen, R. C. (1985). Phytophotodermatitis simulating child abuse. *American Journal of Diseases of Children*, *139*(3), 239-240.

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 92,* 79-97. doi:10.1016/j.isprsjprs.2014.02.013

Cook, D. C., Fraser, R. W., Paini, D. R., Warden, A. C., Lonsdale, W. M., & De Barro, P. J. (2011). Biosecurity and yield improvement technologies are strategic complements in the fight against food insecurity. *PLoS One*, *6*(10), e26084.

Coppa, U., Guarnieri, A., Pirotti, F., & Vettore, A. (2009). Accuracy enhancement of unmanned helicopter positioning with low-cost system. *Applied Geomatics*, *1*(3), 85-95.

Courel, M. F., Chamard, P. C., Guenegou, M. C., Lerhun, J., Levasseur, J., & Togola, M. (1991, October). Utilisation des bandes spectrales du vert et du rouge pour une meilleure évaluation des formations végétales actives. In *Congrès AUPELF-UREF*.

Cowling, S. A., & Field, C. B. (2003). Environmental control of leaf area production: implications for vegetation and land-surface modeling. *Global Biogeochemical Cycles*, *17*(1).

Curran, P. (1983). Estimating green LAI from multispectral aerial photography. *Photogrammetric Engineering and Remote Sensing*, *49*, 1709-1720.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). *Random Forests for Classification in Ecology*. *Ecology*, *88*(11), 2783-2792. doi:10.1890/07-0539.1

Danielle, T., & Kellie, S. (2014). *Invasive Wild Parsnip (Pastinaca sativa) Best Management Practices in Ontario.* Ontario Invasive Plant Council. Retrieved from http://www.ontarioinvasiveplants.ca/wp-content/uploads/2016/07/OIPC_BMP_WildParsnip_Feb182014_FINAL2.pdf

Darbyshire, S. J. (2003). Inventory of Canadian agricultural weeds. Agriculture and Agri-Food Canada. *Research Branch, Ottawa, Ontario*.

David, L. C. G., & Ballado, A. H. (2016, November). Vegetation indices and textures in object-based weed detection from UAV imagery. *In Control System, Computing and Engineering (ICCSCE), 2016 6th IEEE International Conference on* (pp. 273-278). IEEE.

Davis, M. A. (2003). Biotic globalization: does competition from introduced species threaten biodiversity? *AIBS Bulletin*, *53*(5), 481-489.

Dawkins, K., & Esiobu, N. (2016). Emerging insights on brazilian pepper tree (Schinus terebinthifolius) invasion: the potential role of soil microorganisms. *Frontiers in plant science*, *7*.

Dewey, S. A. (1991). *Weedy thistles of the western USA. Westview Special Studies in Agricultural Science and Policy Noxious Range Weeds National Noxious Range Weed Conference on a Forum for Continuing Cooperation*. Westview Press, Boulder, CO.

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, *40*(2), 139-157.

Du, M., & Noguchi, N. (2017). Monitoring of Wheat Growth Status and Mapping of Wheat Yield's within-Field Spatial Variations Using Color Images Acquired from UAV-camera System. *Remote Sensing*, *9*(3), 289.

Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, *118*, 259-272.

Erdas Inc., 1999. Erdas Field Guide. *Erdas Inc.*, Atlanta, Georgia.

Feng, Q., Liu, J., & Gong, J. (2015). UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. *Remote Sensing, 7*(1), 1074-1094. Doi:10.3390/rs70101074

Flanders, D., Hall-Beyer, M., & Pereverzoff, J. (2003). Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. *Canadian Journal of Remote Sensing*, *29*(4), 441-452.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Sciencedirect*, *27*, 294-300. doi:10.1016/j.patrec.2005.08.011

Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ. 80*, 76–87

Govaerts, Y. M., Verstraete, M. M., Pinty, B., & Gobron, N. (1999). Designing optimal spectral indices: A feasibility and proof of concept study. *International journal of remote sensing*, *20*(9), 1853-1873.

Guijarro, M., Pajares, G., Riomoros, I., Herrera, P. J., Burgos-Artizzu, X. P., & Ribeiro, A. (2011). Automatic segmentation of relevant textures in agricultural images. *Computers and Electronics in Agriculture*, *75*(1), 75-83.

Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(1), 56-66. doi:10.1016/j.isprsjprs.2010.08.007

Hague, T., Tillett, N.D., Wheeler, H., 2006. Automated crop and weed monitoring in widely spaced cereals. *Precis. Agric. 7*, 21–32.

Haralick, R. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE, 67*(5), 786-804. doi:10.1109/proc.1979.11328

Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics, (6),* 610-621.

Hawthorne, T. L., Elmore, V., Strong, A., Bennett-Martin, P., Finnie, J., Parkman, J., & Reed, J. (2015). Mapping non-native invasive species and accessibility in an urban forest: a case study of participatory mapping and citizen science in Atlanta, Georgia. *Applied Geography*, *56*, 187-198.

Hayden, J. (2000). Parsnip Phyto-Photosensitivity. *PIG JOURNAL, 46*, 116-117.

Hejda, M., Pyšek, P., & Jarošík, V. (2009). Impact of invasive plants on the species richness, diversity and composition of invaded communities. *Journal of Ecology*, *97*(3), 393-403.

Herwitz, S., Johnson, L., Dunagan, S., Higgins, R., Sullivan, D., Zheng, J., Brass, J. (2004). Imaging from an unmanned aerial vehicle: agricultural surveillance and decision support. *Computers and Electronics in Agriculture*, *44*(1), 49-61. doi:10.1016/j.compag.2004.02.006

Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E., Hungate, B. A., Matulich, K. L., & O'Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, *486*(7401), 105-108.

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, *23*(4), 725-749.

Huete, A. R. (1989). Soil influences in remotely sensed vegetation-canopy spectra, Theory and Applications of Optical Remote Sensing G. Asrar, 107–141.

Ishii, J., & Washitani, I. (2013). Early detection of the invasive alien plant Solidago altissima in moist tall grassland using hyperspectral imagery. *International journal of remote sensing*, *34*(16), 5926-5936.

Jackson, R. D., Pinter Jr, P. J., Reginato, R. J., & Idso, S. B. (1980). Hand-held radiometry: A set of notes developed for use at the Workshop of Hand-held radiometry.

Kaufman, Y. J., & Tanre, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE transactions on Geoscience and Remote Sensing*, *30*(2), 261-270.

Kennedy, R. E., & Cohen, W. B. (2003). Automated designation of tie-points for image-to-image coregistration. *International Journal of Remote Sensing*, *24*(17), 3467-3490.

Kitada, K., & Fukuyama, K. (2012). Land-use and land-cover mapping using a gradable classification method. *Remote Sensing*, *4*(6), 1544-1558.

Kopelson, H. (2014). The Impacts of Invasive Plant Species on Human Health. *City Wild. Paper 6.* Retrieved from http://digitalcommons.esf.edu/citywild/6/

Kulkarni, A. D., & Lowe, B. (2016). Random forest algorithm for land cover classification. *Computer Science Faculty Publications and Presentations. Paper*, *1*.

Laba, M., Downs, R., Smith, S., Welsh, S., Neider, C., White, S., & Baveye, P. (2008). Mapping invasive wetland plants in the Hudson River National Estuarine Research Reserve using quickbird satellite imagery. *Remote Sensing of Environment*, *112*(1), 286-300.

Laliberte, A. S., & Rango, A. (2009). Texture and scale in object-based analysis of subdecimeter resolution unmanned aerial vehicle (UAV) imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(3), 761-770.

Laliberte, A. S., Herrick, J. E., Rango, A., & Winters, C. (2010). Acquisition, orthorectification, and object-based classification of unmanned aerial vehicle (UAV) imagery for rangeland monitoring. *Photogrammetric Engineering & Remote Sensing*, *76*(6), 661-672.

Langley, S. K., Cheshire, H. M., & Humes, K. S. (2001). A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland. *Journal of Arid Environments*, *49*(2), 401-411.

Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment*, *100*(3), 356-362.

Lea, M. (2015, July 10). Wild parsnip blamed for dog's burns. Retrieved October 18, 2017, from http://www.thewhig.com/2015/07/10/wild-parsnip-blamed-for-dogs-burnsn-noxious-weed-man-believes-his-six-year-old-boxer-jackson-came-in-contact-with-wild-parsnip-sap-during-brief-time-in-weedsoneplan

Lee, R. Y., Ou, D. Y., Shiu, Y. S., & Lei, T. C. Comparisons of Using Random Forest And Maximum Likelihood Classifiers With Worldview-2 Imagery For Classifying Crop Types.

Leprieur, C., Verstraete, M. M., & Pinty, B. (1994). Evaluation of the performance of various vegetation indices to retrieve vegetation cover from AVHRR data. *Remote Sensing Reviews*, *10*(4), 265-284.

Levine, M. D. (1985). *Vision in man and machine*. McGraw-Hill College.

Li, Z., Zhu, C., & Gold, C. (2004). *Digital terrain modeling: principles and methodology*. CRC press.

Lombaert, G. A., Siemens, K. H., Pellaers, P., Mankotia, M., & Ng, W. (2001). Furanocoumarins in celery and parsnips: method and multiyear Canadian survey. *Journal of AOAC International*, *84*(4), 1135-1143.Materka, A., & Strzelecki, M. (1998). Texture analysis methods–a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, 9-11.

Lounibos, L. P. (2002). Invasions by insect vectors of human disease. *Annual review of entomology*, *47*(1), 233-266.

Lu, Y. C., Daughtry, C., Hart, G., & Watkins, B. (1997). The current state of precision farming. *Food reviews international*, *13*(2), 141-162.

Materka, A., & Strzelecki, M. (1998). Texture Analysis Methods – A Review, Technical University of Lodz, Institute of Electronics, *COST B11 report, Brussels*, 9-11.

McNairn, H., & Protz, R. (1993). Mapping corn residue cover on agricultural fields in Oxford County, Ontario, using Thematic Mapper. *Canadian Journal of Remote Sensing*, *19*(2), 152-159.

Melesse, A. M., Weng, Q., Thenkabail, P. S., & Senay, G. B. (2007). Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. *Sensors,7*(12), 3209-3241. Doi:10.3390/s7123209

Moran, M. S., Inoue, Y., & Barnes, E. M. (1997). Opportunities and limitations for image-based remote sensing in precision crop management. *Remote sensing of Environment*, *61*(3), 319-346.

Moulin, S. (1999). Impacts of model parameter uncertainties on crop reflectance estimates: a regional case study on wheat. *International Journal of Remote Sensing*, *20*(1), 213-218.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS *Journal of Photogrammetry and Remote Sensing*, *66*(3), 247-259. doi:10.1016/j.isprsjprs.2010.11.001

Mulligan, G. A. (1987). Common weeds of Canada. Field guide.

Mundt, J. T., Glenn, N. F., Weber, K. T., Prather, T. S., Lass, L. W., & Pettingill, J. (2005). Discrimination of hoary cress and determination of its detection limits via hyperspectral image processing and accuracy assessment techniques. *Remote Sensing of Environment*, *96*(3), 509-517.

Murray, H., Lucieer, A., & Williams, R. (2010). Texture-based classification of sub-Antarctic vegetation communities on Heard Island. *International Journal of Applied Earth Observation and Geoinformation*, *12*(3), 138-149.

Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment*, *115*(5), 1145-1161.

Nex, F., & Remondino, F. (2013). UAV for 3D mapping applications: a review. *Applied Geomatics, 6*(1), 1-15. Doi: 10.1007/s12518-013-0120-x

Niethammer, U., Rothmund, S., James, M. R., Travelletti, J., & Joswig, M. (2010). UAV-based remote sensing of landslides. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 38*(Part 5), 496-501.

Nitze, I., Schulthess, U., & Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proc. of the 4th GEOBIA*, 7-9.

Nordberg, M. L., & Evertson, J. (2005). Vegetation index differencing and linear regression for change detection in a Swedish mountain range using Landsat TM® and ETM+® imagery. *Land Degradation & Development*, *16*(2), 139-149.

OFAH/OMNR Invading Species Awareness Program. (2012). *Wild Parsnip*. Retrieved from: http://www.invadingspecies.com

Ok, A. O., Akar, O., & Gungor, O. (2012). Evaluation of random forest method for agricultural crop classification. *European Journal of Remote Sensing*, *45*(1), 421-432.

Otukei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, *12*, S27-S31.

Pajares, G. (2015). Overview and Current Status of Remote Sensing Applications Based on Unmanned Aerial Vehicles (UAVs). *Photogrammetric Engineering & Remote Sensing*, *81*(4), 281-330. doi:10.14358/pers.81.4.281

Parsnips Poison Pigs. (2014). *The Pig Site.* Retrieved November 18, 2017, from http://www.thepigsite.com/swinenews/38123/parsnips-poison-pigs/

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*(1), 217-222. Doi: 10.1080/01431160412331269698

Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., & Wilhelm, V. L. (2017). Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery. *European Journal of Remote Sensing*, *50*(1), 564-576. doi:10.1080/22797254.2017.1373602

Paola, J. D., & Schowengerdt, R. A. (1995). A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and remote sensing, 33*(4), 981-996.

PCI Geomatics. (2015). *Geomatica OrthoEngine Course Exerceses.* Markham, Ontario.

Pearson, R. L., & Miller, L. D. (1972, October). Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. In *Remote Sensing of Environment, VIII* (p. 1355).

Pepe, M., & Prezioso, G. (2016). Two Approaches for Dense DSM Generation from Aerial Digital Oblique Camera System. *Proceedings of the 2nd International Conference on Geographical Information Systems Theory, Applications and Management.* Doi:10.5220/0005774900630070

Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological economics, 52*(3), 273-288.

Price, K. P., Guo, X., & Stiles, J. M. (2002). Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas. *International Journal of Remote Sensing, 23*(23), 5031-5042.

Puissant, A., Hirsch, J., & Weber, C. (2005). The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. *International Journal of Remote Sensing, 26*(4), 733-745. Doi:10.1080/01431160512331316838

Puri, A., Valavanis, K. P., & Kontitsis, M. (2007, June). Statistical profile generation for traffic monitoring using real-time UAV based video data. In *Control & Automation, 2007. MED'07. Mediterranean Conference on* (pp. 1-6). IEEE.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote sensing of environment, 48*(2), 119-126.

Racoviteanu, A., & Williams, M. W. (2012). Decision tree and texture analysis for mapping debris-covered glaciers in the Kangchenjunga area, Eastern Himalaya. *Remote sensing*, *4*(10), 3078-3109.

Schowengerdt, R. A. (2011). *Remote sensing: Models and methods for image processing*. P.2. New Delhi: Elsevier.

Simactive User Manual. (2016). *Simactive Photogrammetry Software.* Software Version 6.5

Richards, J. A. and Jia, X., 2006. *Remote Sensing Digital Image Analysis: An Introduction*. Berlin, Germany: Springer.

Rietschel, R. L., Fowler, J. F., & Fisher, A. A. (2008). *Fisher's contact dermatitis*. PMPH-USA. Pp. 492-495, 526-528, 539-543.

Robertson, L. D., & King, D. J. (2011). Comparison of pixel- and object-based classification in land cover change mapping. *International Journal of Remote Sensing*, *32*(6), 1505-1529. Doi: 10.1080/01431160903571791

Rodriguez-Galiano, V., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P., & Jeganathan, C. (2012). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, *121*, 93-107. doi:10.1016/j.rse.2011.12.003

Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93-104. doi:10.1016/j.isprsjprs.2011.11.002

Rosenzweig, M. L. (2001). The four questions: what does the introduction of exotic species do to diversity? *Evolutionary Ecology Research*, *3*(3), 361-367.

Rouse Jr, J. W. (1972). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation.

Rouse Jr, J. W. (1974). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation.

Rouse, J. W. (1973). Monitoring the vernal advancement and retrogradation of natural vegetation [NASA/GSFCT Type II Report]. *Greenbelt, MD: NASA/Goddard Space Flight Center*.

Salamí, E., Barrado, C., & Pastor, E. (2014). UAV Flight Experiments Applied to the Remote Sensing of Vegetated Areas. *Remote Sensing*, *6*(11), 11051-11081. Doi: 10.3390/rs61111051

Sanjerehei, M. M. (2014). Assessment of spectral vegetation indices for estimating vegetation cover in arid and semiarid shrublands. *Range Management and Agroforestry*, *35*(1), 91-100.

Sax, D. F., Gaines, S. D., & Brown, J. H. (2002). Species invasions exceed extinctions on islands worldwide: a comparative study of plants and birds. *The American Naturalist*, *160*(6), 766-783.

Scott, B. R., Pathak, M. A., & Mohn, G. R. (1976). Molecular and genetic basis of furocoumarin reactions. *Mutation Research/Reviews in Genetic Toxicology*, *39*(1), 29-74.

Singh, K. K., Davis, A. J., & Meentemeyer, R. K. (2015). Detecting understory plant invasion in urban forests using LiDAR. *International Journal of Applied Earth Observation and Geoinformation*, *38*, 267-279.

Slobodkin, L. B. (2001). The good, the bad and the reified. *Evolutionary Ecology Research*, *3*(1), 91-105.

Sluiter, R. (2005). *Mediterranean land cover change: modelling and monitoring natural vegetation using GIS and remote sensing*. Utrecht University.

Søborg, I., Andersson, C., & Gry, J. (1996). *Furocoumarins in Plant Foods: Exposure, Biological Properties, Risk Assessment and Recommendations*. Nordic Council of Ministers.

Solis, R. R., Dotson, D. A., & Trizna, Z. (2000). Phytophotodermatitis: a sometimes difficult diagnosis. *Archives of family medicine*, *9*(10), 1195.

Szantoi, Z., Escobedo, F., Abd-Elrahman, A., Smith, S., & Pearlstine, L. (2013). Analyzing fine-scale wetland composition using high resolution imagery and texture features. *International Journal of Applied Earth Observation and Geoinformation,23*, 204-212. doi:10.1016/j.jag.2013.01.003

Tarabalka, Y., Benediktsson, J. A., & Chanussot, J. (2009). Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(8), 2973-2987.

Thomas, N., Hendrix, C., & Congalton, R. G. (2003). A comparison of urban mapping methods using high-resolution digital imagery. *Photogrammetric Engineering & Remote Sensing*, *69*(9), 963-972.

Torre, C.D. (2016, March). Python for geospatial data processing. Retrieved October 20, 2017, from Http://www.machinalis. Com/blog/python-for-geospatial-data-processing/

Torres-Sánchez, J., López-Granados, F., & Peña, J. M. (2015). An automatic object-based method for optimal thresholding in UAV images: Application for vegetation detection in herbaceous crops. *Computers and Electronics in Agriculture*, *114*, 43-52.

Torres-Sánchez, J., López-Granados, F., De Castro, A. I., & Peña-Barragán, J. M. (2013). Configuration and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. *PLoS One*, *8*(3), e58210.

Torres-Sánchez, J., Peña, J. M., De Castro, A. I., & López-Granados, F. (2014). Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. *Computers and Electronics in Agriculture*, *103*, 104-113.

Tsai, Y. H., Stow, D., & Weeks, J. (2011). Comparison of Object-Based Image Analysis Approaches to Mapping New Buildings in Accra, Ghana Using Multi-Temporal QuickBird Satellite Imagery. *Remote Sensing*, *3*(12), 2707-2726. Doi:10.3390/rs3122707

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, *8*(2), 127-150.

Tylianakis, J. M., Didham, R. K., Bascompte, J., & Wardle, D. A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecology letters*, *11*(12), 1351-1363.

Underwood, E., Ustin, S., & DiPietro, D. (2003). Mapping nonnative plants using hyperspectral imagery. *Remote Sensing of Environment*, *86*(2), 150-161.

Unmanned Aircraft Systems (UAS). (2010). *International Civil Aviation Organization*, 1-15. doi:10.1002/9780470664797.ch1

USDA ARS (United States Department of Agriculture, Agricultural Research Service). (2006). National Genetic Resources Program. Germplasm Resources Information Network (GRIN), National Germplasm Resources Laboratory. Retrieved October 27, 2017, from http://www.ars-grin.gov/cgi-bin/npgs/html/tax_search.pl?pastinaca+sativa

Valkealahti, K., & Oja, E. (1998). Reduced multidimensional co-occurrence histograms in texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(1), 90-94.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, *10*(5), 988-999.

Vapnik, V., 1979. Estimation of Dependences Based on Empirical Data. *Nauka, Moscow*, pp. 5165–5184, 27 (in Russian) (English translation: Springer Verlag, New York, 1982).

Verstraete, M. M., & Pinty, B. (1996). Designing optimal spectral indexes for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing*, *34*(5), 1254-1265.

Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, *115*(12), 3468-3478.

Wang, J., Zhang, J., & Zhao, J. (2016). Texture Classification Using Scattering Statistical and Cooccurrence Features. *Mathematical Problems in Engineering*, *2016*.

Wang, X., Wang, M., W, S., & Wu Y. (2015). Extraction of vegetation information from visible unmanned aerial vehicle images. *Transactions of the Chinese Society of Agricultural Engineering*, *31*(5).

Woebbecke, D.M., Meyer, G.E., Von Bargen, K., Mortensen, D.A., 1995. Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. ASAE 38*, 259–269

Xiao, X., He, L., Salas, W., Li, C., Moore Iii, B., Zhao, R., & Boles, S. (2002). Quantitative relationships between field-measured leaf area index and vegetation index derived from VEGETATION images for paddy rice fields. *International Journal of Remote Sensing*, *23*(18), 3595-3604.

Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., & Ojima, D. (2004). Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment*, *91*(2), 256-270.

Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of plant ecology*, *1*(1), 9-23.

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., & Schirokauer, D. (2006). Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. *Photogrammetric Engineering & Remote Sensing, 72*(7), 799-811. doi:10.14358/pers.72.7.799

Yuan, L., Fu, L., Yang, Y., & Miao, J. (2009). Analysis of texture feature extracted by gray level co-occurrence matrix. *Journal of Computer Applications, 29*(4), 1018-1021. doi:10.3724/sp.j.1087.2009.01018

Zangeri, A. R., Berenbaum, M. R., & Levine, E. (1989). Genetic control of seed chemistry and morphology in wild parsnip (Pastinaca sativa). *Journal of Heredity*, *80*(5), 404-407.

Zhang, C. (2008). An UAV-based photogrammetric mapping system for road condition assessment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci, 37*, 627-632.

Zohary, D., & Hopf, M. (2012). *Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley.* Oxford: Oxford Univ. Press.

Appendix A

Python Code of Out-of-bag (OOB) Plot

```
import numpy as np
import os

from osgeo import gdal
from sklearn import metrics
from sklearn.ensemble import RandomForestClassifier

from matplotlib import pyplot as plt
%matplotlib inline

# A list of "random" colors (for a nicer output)
COLORS = ["#000000", "#FFFF00", "#1CE6FF", "#FF34FF", "#FF4A46", "#008941"]
```

Define some useful functions that we are going to be using later. They are making heavy use of the GDAL api to manipulate raster and vector data.

```
def create_mask_from_vector(vector_data_path, cols, rows, geo_transform, projection, target_value=1):
    """Rasterize the given vector (wrapper for gdal.RasterizeLayer)."""
    data_source = gdal.OpenEx(vector_data_path, gdal.OF_VECTOR)
    layer = data_source.GetLayer(0)
    driver = gdal.GetDriverByName('MEM')  # In memory dataset
    target_ds = driver.Create('', cols, rows, 1, gdal.GDT_UInt16)
    target_ds.SetGeoTransform(geo_transform)
    target_ds.SetProjection(projection)
    gdal.RasterizeLayer(target_ds, [1], layer, burn_values=[target_value])
    return target_ds
def vectors_to_raster(file_paths, rows, cols, geo_transform, projection):
    """Rasterize all the vectors in the given directory into a single image."""
    labeled_pixels = np.zeros((rows, cols))
    for i, path in enumerate(file_paths):
        label = i+1
        ds = create_mask_from_vector(path, cols, rows, geo_transform, projection, target_value=label)
        band = ds.GetRasterBand(1)
        labeled_pixels += band.ReadAsArray()
        ds = None
    return labeled_pixels
def write_geotiff(fname, data, geo_transform, projection):
    """Create a GeoTIFF file with the given data."""
    driver = gdal.GetDriverByName('GTiff')
    rows, cols = data.shape
    dataset = driver.Create(fname, cols, rows, 1, gdal.GDT_Byte)
    dataset.SetGeoTransform(geo_transform)
    dataset.SetProjection(projection)
    band = dataset.GetRasterBand(1)
    band.WriteArray(data)
    dataset = None  # Close the file
```

```
raster_data_path = "7040/PixelAggregate/Resolution0.01/5bandsFINALImage.tif"
#output_fname = "classification.tiff"
train_data_path = "7040/PixelAggregate/Resolution0.01/ROITraining/"
```

Now, we will use the GDAL api to read the input GeoTiff: extract the geographic information and transform the band's data into a numpy array.

```
raster_dataset = gdal.Open(raster_data_path, gdal.GA_ReadOnly)
geo_transform = raster_dataset.GetGeoTransform()
proj = raster_dataset.GetProjectionRef()
bands_data = []
for b in range(1, raster_dataset.RasterCount+1):
    band = raster_dataset.GetRasterBand(b)
    bands_data.append(band.ReadAsArray())

bands_data = np.dstack(bands_data)
rows, cols, n_bands = bands_data.shape
```

Process the training data: project all the vector data, in the training dataset, into a numpy array. Each class is assigned a label (a number between 1 and the total number of classes). If the value v in the position (i, j) of this new array is not zero, that means that the pixel (i, j) must be used as a training sample of class v. training_samples is the list of pixels to be used for training. In our case, a pixel is a point in the 7-dimensional space of the bands. training_labels is a list of class labels such that the i-th position indicates the class for i-th pixel in training_samples.

```
files = [f for f in os.listdir(train_data_path) if f.endswith('.shp')]
classes = [f.split('.')[0] for f in files]
shapefiles = [os.path.join(train_data_path, f) for f in files if f.endswith('.shp')]

labeled_pixels = vectors_to_raster(shapefiles, rows, cols, geo_transform, proj)
is_train = np.nonzero(labeled_pixels)
training_labels = labeled_pixels[is_train]
training_samples = bands_data[is_train]
```

OOB Errors for Random Forests

```
clf = RandomForestClassifier(warm_start=True, oob_score=True, max_features="sqrt")

#dict of (estimator, error_rate)
error_rate = {}

#explore max 200
min_estimators = 10
max_estimators = 1500

for i in range(min_estimators, max_estimators + 1):
    clf.set_params(n_estimators=i)
    clf.fit(training_samples, training_labels)
    # Record the OOB error for each `n_estimators=i` setting.
    oob_error = 1 - clf.oob_score_
    error_rate[i] = oob_error

# Generate the "OOB error rate" vs. "n_estimators" plot.
er = sorted(error_rate.items())
x, y = zip(*er)
fig = plt.figure(figsize=(10,5))
plt.plot(x, y)

plt.xlim(min_estimators, max_estimators)
plt.xlabel("n_estimators")
plt.ylabel("OOB error rate")
plt.show()
fig.savefig('7040/PixelAggregate/Resolution0.01/n1500.png', dpi = 300)
```

Appendix B

Python Code of Input Variables Ranking

Feature Importance

```
feature_importance = classifier.feature_importances_
index = np.argsort(feature_importance)[::-1]

fig= plt.figure(figsize=(10,5))
plt.title("Feature Importances")
plt.bar(range(training_samples.shape[1]), feature_importance[index],
       color="r", align="center")
plt.xticks(range(training_samples.shape[1]), index)
plt.xlim([-1, training_samples.shape[1]])
plt.show()
fig.savefig('7040/PixelAggregate/Resolution0.05/5bandsfeatureImportance.png', dpi = 300)
```

Appendix C

Python Code of Random Forest

Training

Now, we will use the GDAL api to read the input GeoTiff: extract the geographic information and transform the band's data into a numpy array.

```
raster_dataset = gdal.Open(raster_data_path, gdal.GA_ReadOnly)
geo_transform = raster_dataset.GetGeoTransform()
proj = raster_dataset.GetProjectionRef()
bands_data = []
for b in range(1, raster_dataset.RasterCount+1):
    band = raster_dataset.GetRasterBand(b)
    bands_data.append(band.ReadAsArray())

bands_data = np.dstack(bands_data)
rows, cols, n_bands = bands_data.shape
```

Process the training data: project all the vector data, in the training dataset, into a numpy array. Each class is assigned a label (a number between 1 and the total number of classes). If the value v in the position (i, j) of this new array is not zero, that means that the pixel (i, j) must be used as a training sample of class v.

Training_samples is the list of pixels to be used for training. In our case, a pixel is a point in the 7-dimensional space of the bands.

Training_labels is a list of class labels such that the i-th position indicates the class for i-th pixel in training_samples.

```
files = [f for f in os.listdir(train_data_path) if f.endswith('.shp')]
classes = [f.split('.')[0] for f in files]
shapefiles = [os.path.join(train_data_path, f) for f in files if f.endswith('.shp')]

labeled_pixels = vectors_to_raster(shapefiles, rows, cols, geo_transform, proj)
is_train = np.nonzero(labeled_pixels)
training_labels = labeled_pixels[is_train]
training_samples = bands_data[is_train]
```

```
classifier = RandomForestClassifier(max_features='sqrt', n_estimators=1100, n_jobs=-1, oob_score=True, random_state=None)
classifier.fit(training_samples, training_labels)
```

Classifying

```
#input data for process

#image size
r = 2458
c = 2458

#process size, rows per process
offset = 100

#y_start row
y_offset = 0
#y remains in the third part
y = r%offset
if(y == 0):
    y = offset

#num of iterations in the second part
k = int((r-offset-y)/offset)     #(rows-first_rows-third_rows)/offset
```

```
#first part
h_bands_data = []

for b in range(1, raster_dataset.RasterCount+1):
    band = raster_dataset.GetRasterBand(b)
    band_data = []

    h_bands_data.append(band.ReadAsArray(0, 0, c, offset))

h_bands_data = np.dstack(h_bands_data)
rows, cols, n_bands = h_bands_data.shape

print("Each process bands data shape (rows, cols, n_bands):")
print(h_bands_data.shape)

n_samples = rows*cols
flat_pixels = h_bands_data.reshape((n_samples, n_bands))

# nan to 0
for i in range(0, flat_pixels.shape[0]):
    for j in range(0, flat_pixels.shape[1]):
        if(np.isnan(flat_pixels[i][j]) == True):
            flat_pixels[i][j] = 0

result = classifier.predict(flat_pixels)
classification = result.reshape((rows, cols))

print("Result classification shape (rows, cols):")
print(classification.shape)

y_offset += offset
```

```
#second part
for i in range(0, k):
    h_bands_data = []
    for b in range(1, raster_dataset.RasterCount+1):
        band = raster_dataset.GetRasterBand(b)
        band_data = []

        h_bands_data.append(band.ReadAsArray(0, y_offset, c, offset))

    h_bands_data = np.dstack(h_bands_data)
    rows, cols, n_bands = h_bands_data.shape

    n_samples = rows*cols
    flat_pixels = h_bands_data.reshape((n_samples, n_bands))
    result = classifier.predict(flat_pixels)
    classification2 = result.reshape((rows, cols))

    for i in range(0, classification2.shape[0]):
        classification = np.append(classification, [classification2[i]], 0)

    print("Result classification shape (rows, cols):")
    print(classification.shape)

    y_offset += offset

print
print("y_start = %d" % y_offset)
print("y_rows = %d" % y)
```

```
from matplotlib import pyplot as plt
%matplotlib inline

f = plt.figure(figsize=(15,15))
f.add_subplot(1, 2, 2)
r = bands_data[:,:,0]
g = bands_data[:,:,1]
b = bands_data[:,:,2]
rgb = np.dstack([r,g,b])
f.add_subplot(1, 2, 1)
plt.imshow(rgb/255)
f.add_subplot(1, 2, 2)
plt.imshow(classification)
```

```
#third part
h_bands_data = []

for b in range(1, raster_dataset.RasterCount+1):
    band = raster_dataset.GetRasterBand(b)
    band_data = []

    h_bands_data.append(band.ReadAsArray(0, y_offset, c, y))

h_bands_data = np.dstack(h_bands_data)
rows, cols, n_bands = h_bands_data.shape

n_samples = rows*cols
flat_pixels = h_bands_data.reshape((n_samples, n_bands))
result = classifier.predict(flat_pixels)
classification2 = result.reshape((rows, cols))

for i in range(0, classification2.shape[0]):
    classification = np.append(classification, [classification2[i]], 0)

print("result classification shape, (rows, cols):")
print(classification.shape)
```

Appendix D

Python Code of Object-based Image Classification

```
import numpy as np
import os
import scipy

from matplotlib import pyplot as plt
%matplotlib inline
from matplotlib import colors
from osgeo import gdal
from skimage import exposure
from skimage.segmentation import quickshift, felzenszwalb
from sklearn import metrics
from sklearn.ensemble import RandomForestClassifier
```

```
# input data
RASTER_DATA_FILE = "7040/PixelAggregate/window1/9features.tif"
TRAIN_DATA_PATH = "7040/PixelAggregate/ROITraining/"
TEST_DATA_PATH = "7040/PixelAggregate/ROITesting/"
Output_Seg = "7040/PixelAggregate/OBIA/segmentation.tiff"
Output_trainingSeg = "7040/PixelAggregate/OBIA/trainingseg.tiff"
output_fname = "7040/PixelAggregate/OBIA/classification.tiff"
```

Defien some functions

```
def create_mask_from_vector(vector_data_path, cols, rows, geo_transform, projection, target_value=1):
    """Rasterize the given vector (wrapper for gdal.RasterizeLayer)."""
    data_source = gdal.OpenEx(vector_data_path, gdal.OF_VECTOR)
    layer = data_source.GetLayer(0)
    driver = gdal.GetDriverByName('MEM')  # In memory dataset
    target_ds = driver.Create('', cols, rows, 1, gdal.GDT_UInt16)
    target_ds.SetGeoTransform(geo_transform)
    target_ds.SetProjection(projection)
    gdal.RasterizeLayer(target_ds, [1], layer, burn_values=[target_value])
    return target_ds
```

```
def vectors_to_raster(file_paths, rows, cols, geo_transform, projection):
    """Rasterize all the vectors in the given directory into a single image."""
    labeled_pixels = np.zeros((rows, cols))
    for i, path in enumerate(file_paths):
        label = i+1
        ds = create_mask_from_vector(path, cols, rows, geo_transform,
                                     projection, target_value=label)
        band = ds.GetRasterBand(1)
        labeled_pixels += band.ReadAsArray()
        ds = None
    return labeled_pixels
```

```
def write_geotiff(fname, data, geo_transform, projection):
    """Create a GeoTIFF file with the given data."""
    driver = gdal.GetDriverByName('GTiff')
    rows, cols = data.shape
    dataset = driver.Create(fname, cols, rows, 1, gdal.GDT_Byte)
    dataset.SetGeoTransform(geo_transform)
    dataset.SetProjection(projection)
    band = dataset.GetRasterBand(1)
    band.WriteArray(data)
    dataset = None  # Close the file
```

Extract the geographic information and transform the band's data into a numpy array

```python
raster_dataset = gdal.Open(RASTER_DATA_FILE, gdal.GA_ReadOnly)
geo_transform = raster_dataset.GetGeoTransform()
proj = raster_dataset.GetProjectionRef()
n_bands = raster_dataset.RasterCount
bands_data = []
for b in range(1, n_bands+1):
    band = raster_dataset.GetRasterBand(b)
    bands_data.append(band.ReadAsArray())

bands_data = np.dstack(b for b in bands_data)
```

Create original image

```python
img = exposure.rescale_intensity(bands_data)
rgb_img = np.dstack([img[:, :, 0], img[:, :, 1], img[:, :, 2]])
```

```python
plt.figure(figsize=(15,15))
plt.imshow(rgb_img)
```

Felzenszwalb segmentation

```python
# Apply the segmentation in each band and then combine the results
band_segmentation = []
for i in range(n_bands):
    band_segmentation.append(felzenszwalb(img[:, :, i], scale=85, sigma=0.25, min_size=9))
```

```python
# Put pixels in same segment only if in the same segment in all bands
# Combine the band segmentation to one number
const = [b.max() + 1 for b in band_segmentation]
segmentation = band_segmentation[0]
for i, s in enumerate(band_segmentation[1:]):
    segmentation += s * np.prod(const[:i+1])

_, labels = np.unique(segmentation, return_inverse=True)
segments_felz = labels.reshape(img.shape[:2])
```

```python
cmap = colors.ListedColormap(np.random.rand(len(np.unique(segments_felz)), 3))
plt.figure(figsize=(15,15))
plt.imshow(segments_felz, interpolation='none', cmap=cmap)
```

```python
write_geotiff(Output_Seg, segments_felz, geo_transform, proj)
```

```python
# Show three images
n_segments = max(len(np.unique(s)) for s in [segments_quick, segments_felz])
cmap = colors.ListedColormap(np.random.rand(n_segments, 3))
#SHOW_IMAGES:
f, (ax1, ax2, ax3) = plt.subplots(1, 3, sharey=True, figsize=(15,15))
ax1.imshow(rgb_img, interpolation='none')
ax1.set_title('Original image')
ax2.imshow(segments_quick, interpolation='none', cmap=cmap)
ax2.set_title('Quickshift segmentations')
ax3.imshow(segments_felz, interpolation='none', cmap=cmap)
ax3.set_title('Felzenszwalb segmentations')
plt.show()
```

```python
segments = segments_felz
segment_ids = np.unique(segments)
print("Felzenszwalb segmentation. %i segments." % len(segment_ids))
```

```
rows, cols, n_bands = img.shape
files = [f for f in os.listdir(TRAIN_DATA_PATH) if f.endswith('.shp')]
classes_labels = [f.split('.')[0] for f in files]
shapefiles = [os.path.join(TRAIN_DATA_PATH, f) for f in files if f.endswith('.shp')]
print(shapefiles)
```

```
ground_truth = vectors_to_raster(shapefiles, rows, cols, geo_transform, proj)
```

```
classes = np.unique(ground_truth)[1:]  # 0 doesn't count
len(classes)
```

```
segments_per_klass = {}
for klass in classes:
    segments_of_klass = segments[ground_truth==klass]
    segments_per_klass[klass] = set(segments_of_klass)
    print("Training segments for class %i: %i" % (klass, len(segments_per_klass[klass])))
```

Disambiguation

```
# Check if there are segments which contain training pixels of different classes.
accum = set()
intersection = set()
for class_segments in segments_per_klass.values():
    intersection |= accum.intersection(class_segments)
    accum |= class_segments
assert len(intersection) == 0
```

Paint in black all segments that are not for training. Training segments will be painted of a color depending on the class

To do that we'll set as threshold the max segment id (max segments image pixel value). Then, to the training segments we'll assign values higher than the threshold. Finally, we assign 0 (zero) to pixels with values equal or below the threshold.

```
train_img = np.copy(segments)
threshold = train_img.max() + 1
for klass in classes:
    klass_label = threshold + klass
    for segment_id in segments_per_klass[klass]:
        train_img[train_img == segment_id] = klass_label
train_img[train_img <= threshold] = 0
train_img[train_img > threshold] -= threshold
```

Training segments

```
plt.figure(figsize=(15,15))
cm = np.array([[ 1,  1,  1 ], [ 1,0,0], [ 1,0,1], [ 0,1,0], [ 0,1,1], [ 0,0,1]])
cmap = colors.ListedColormap(cm)
plt.imshow(train_img, cmap=cmap)
plt.colorbar(ticks=[0,1,2,3,4,5])
```

Training

Transform each training segment into a segment model and thus creating the training dataset

```python
# For each band, compute: min, max, mean, variance, skewness, kurtosis
def segment_features(segment_pixels):
    features = []
    n_pixels, n_bands = segment_pixels.shape
    for b in range(n_bands):
        stats = scipy.stats.describe(segment_pixels[:,b])
        band_stats = list(stats.minmax) + list(stats)[2:]
        if n_pixels == 1:
            # scipy.stats.describe raises a Warning and sets variance to nan
            band_stats[3] = 0.0  # Replace nan with something (zero)
        features += band_stats
    return features
```

Create all the objects

Compute the features' vector for each segment (and append the segment ID as reference)

```python
# time consuming
import warnings

with warnings.catch_warnings():
    warnings.simplefilter("ignore")

    objects = []
    objects_ids = []
    for segment_label in segment_ids:
        segment_pixels = img[segments==segment_label]
        segment_model = segment_features(segment_pixels)
        objects.append(segment_model)
        # Keep a reference to the segment label
        objects_ids.append(segment_label)

    print("Created %i objects" % len(objects))
```

Subset the training data

```python
training_labels = []
training_objects = []
for klass in classes:
    class_train_objects = [v for i, v in enumerate(objects) if objects_ids[i] in segments_per_klass[klass]]
    training_labels += [klass] * len(class_train_objects)
    print("Training samples for class %i: %i" % (klass, len(class_train_objects)))
    training_objects += class_train_objects
```

Train a classifier

```python
classifier = RandomForestClassifier(max_features='sqrt', n_estimators=1200, n_jobs=-1, oob_score=True,random_state=None)
classifier.fit(training_objects, training_labels)
```

Classify all segments

```python
predicted = classifier.predict(objects)
```

Propagate the classification

Now that each segment has been classified, we need to propagate that classification to the pixel level. That is, given the class k for the segment with label S, generate a classification from the segmented image where all pixels in segment S are assigned the class k.

```python
clf = np.copy(segments)
```

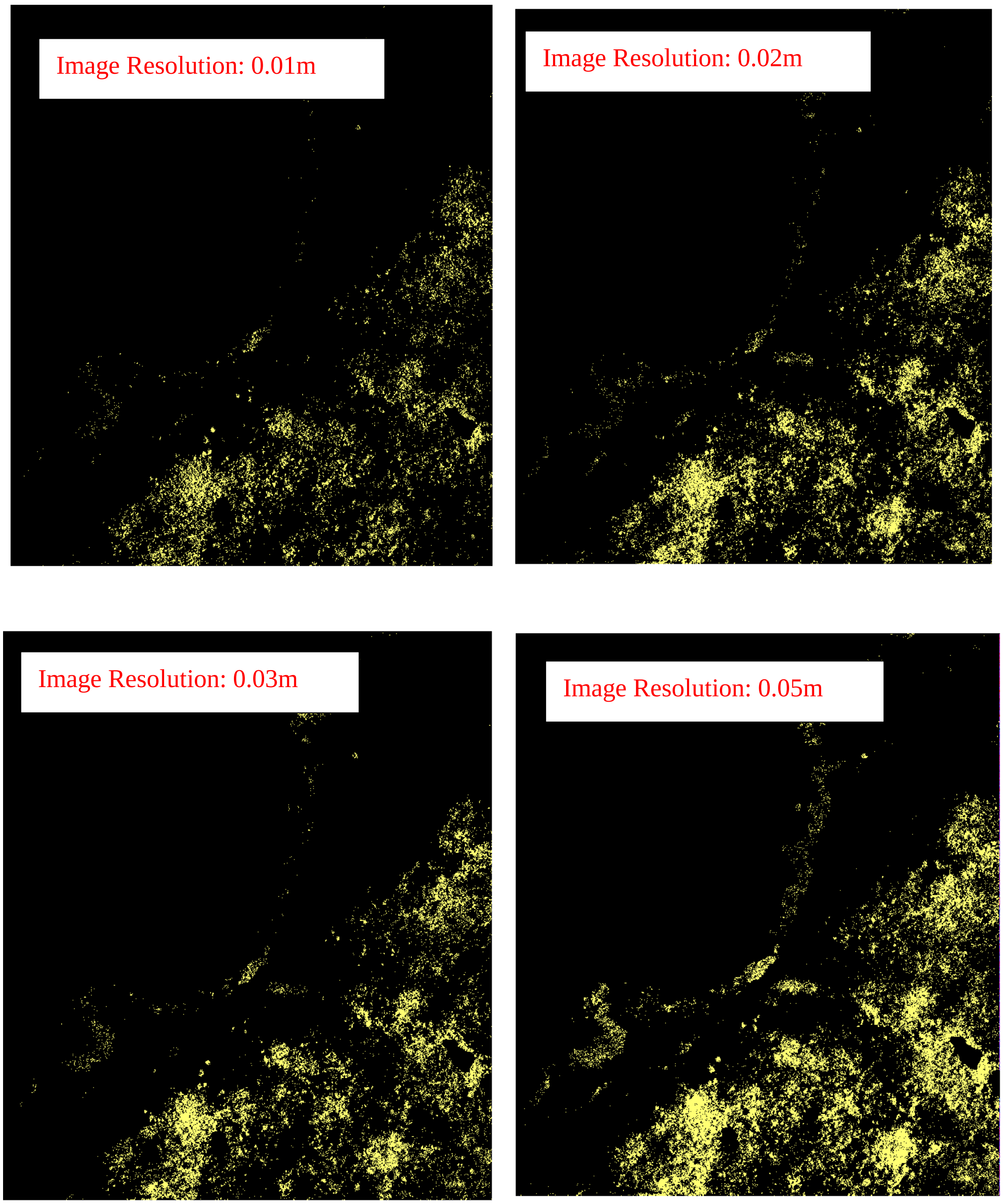
```python
for segment_id, klass in zip(objects_ids, predicted):
    clf[clf==segment_id] = klass
```

```python
f, (ax1, ax2) = plt.subplots(1, 2, sharey=True, figsize=(15,15))
ax1.imshow(rgb_img, interpolation='none')
ax1.set_title('Original image')
ax2.imshow(clf, interpolation='none', cmap=colors.ListedColormap(np.random.rand(len(classes_labels), 3)))
ax2.set_title('Classification')
```

```python
write_geotiff(output_fname, clf, geo_transform, proj)
```

Appendix E

The Comparison between Different Resolution Images



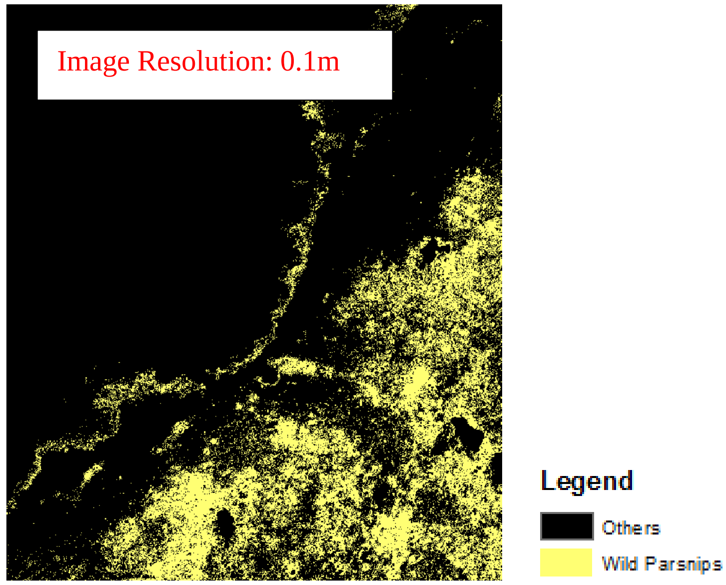


Image Resolution: 0.1m
Legend
Others
Wild Parsnips