final report

Project code: B.NBP.0663

Prepared by: Alistair Reid, Mitch Bryson, Salah Sukkarieh

Australian Centre for Field Robotics
University of Sydney

Date published: November 2012

ISBN: 9781741919622

PUBLISHED BY
Meat & Livestock Australia Limited
Locked Bag 991
NORTH SYDNEY NSW 2059

New Detection and Classification Algorithms for Mapping Woody Weeds from UAV Data

An extension of project B.NBP.0474

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

This publication is published by Meat & Livestock Australia Limited ABN 39 081 678 364 (MLA). Care is taken to ensure the accuracy of the information contained in this publication. However MLA cannot accept responsibility for the accuracy or completeness of the information or opinions contained in the publication. You should make your own enquiries before making decisions concerning your interests. Reproduction in whole or in part of this publication is prohibited without prior written consent of MLA.

Abstract

The detection and eradication of woody weed infestations in native and farmland environments is a difficult and costly management problem. Seeking an efficient solution for surveying these environments, the preceding MLA project B.NBP.0474 funded the development and testing of an Unmanned Aerial Vehicle (UAV) system at the Australian Centre for Field Robotics at the University of Sydney, focused on acquiring large-scale image datasets using a small, low-cost platform. The system was demonstrated in field trials, collecting accurately geo-referenced, high resolution red-green-blue images of woody weed infestations. The project also resulted in an initial concept of an algorithmic framework for building maps from the image data and using computer vision algorithms to detect different weed species in the imagery.

The objective of this project (B.NBP.0663) has been to further develop the algorithms for processing the UAV aerial data. Specifically, new algorithm research and development has been focused on accurate large scale mapping through globally consistent image mosaicking, and on improving the framework for automated detection and classification of woody weeds through the application of state-of-the-art computer vision techniques that learn important visual features from the data. High level processing algorithms have also been investigated to delineate and count tree crowns, and even suggest intelligent paths that could be taken to visit the weeds based on the resulting maps. The combined goal of these analyses is to demonstrate that automated processing can maximise the efficiency of a UAV surveillance system, producing key information in a condensed accessible format, while avoiding large amounts of manual analysis that would counteract the advantages of automated surveillance.

Executive summary

Woody weed infestations over large farmland regions pose a difficult and costly land management problem. The extent of an infestation or the presence of individual invasive trees can be difficult to detect from conventional remote-sensing data sources such as satellite images (due to their limited spatial resolution), making it difficult to rely on them for planning intelligent survey, control and eradication strategies.

This project is part of a larger body of research to demonstrate a low flying autonomous Unmanned Aerial Vehicle (UAV) as a low cost solution for remotely sensing weed infestations (and potentially other spatially distributed natural phenomenon). The proposed system consists of a robotic aircraft and ground station to acquire data, and a framework of computer analysis algorithms to process the data and provide detections, maps and statistics.

The preceding project (B.NBP.0474) was concerned with the development and testing of the robotic UAV system over a three year period from 2008-2010. The project concluded in field trials in 2009 and 2010, demonstrating small, low-cost UAVs operating effectively in a rugged farmland environment. The successful flight trials provided a number of large-scale aerial image datasets (consisting of high resolution red-green-blue image tiles with accompanying navigational information), and initial concepts were presented for classifying and mapping based on this data.

In this ensuing project, development has been focused on the analysis algorithms with the goal of extracting high level scientific information from the UAV sensor data. Three key research areas have been pursued.

Firstly, a large-scale bundle-adjustment framework has been designed to combine the UAV image tiles into a larger terrain representation with globally accurate geo-referencing. This allows the locations of features in the imagery to be accurately pinpointed in geodetic co-ordinates, and also stitches images across frame edges allowing the vision analysis that follows to classify trees that span multiple camera frames as a single object. The three dimensional stereo vision techniques underlying this reconstruction also use (and can visualise) additional information such as three dimensional structure.

Secondly, vision algorithms have been implemented for reliable detection of trees in the imagery, and even classification between species. This has involved identifying descriptors from the data for distinguishing between different types of objects, and applying machine learning techniques to automatically search the for the different classes in the data.

The final research component has been to conduct a high level analysis of the resulting maps, deriving further statistics such as tree crown counts, cover area, and producing easy-to-interpret maps with detection overlays. Algorithms can then be used to suggest efficient paths for inspection and treatment, either for an aerial vehicle, a ground vehicle, or to serve as waypoints on a PDA for a field team. It is expected that these condensed high-level results will be of the most value and interest to environmental officers and land managers who operate UAVs in the future.

While this project has used an experimental UAV system, it is setting the groundwork for a future of UAVs in Australian agriculture, and our results suggest that future management practices will benefit from the global trend of increasingly affordable and readily available commercial UAVs and sensor systems for civilian applications.

Table of Contents

1 Background

The detection and eradication of woody weed infestations in native and farmland environments is a difficult and costly management problem. The objective of the proceeding project B.NBP.0474 with the University of Sydney was to develop and test an Unmanned Aerial Vehicle (UAV) system for weed detection and control, acquiring a large dataset of aerial images over farmland in Northern Queensland. This project (B.NBP.0663) is a continuation of the research to further develop the automated processing system for the UAV data. This report outlines the new developments in the algorithms for terrain modelling, classification and mapping.

The data analysed in this project were acquired during field trials in 2009 and 2010, and focus on regions of known woody weed infestations of Parkinsonia, Prickly Acacia and Mimosa present amongst other species. Because the unmanned aircraft was able to operate safely at low altitudes, it acquired a very high pixel ground-resolution (approximately 4cm), while employing automated flight control systems to obtain consistent overlapping swaths over survey regions extending beyond the sight range of the operator. However, due to weight, cost and power restrictions on the small UAV, the aircraft was not able to acquire multi-spectral pixels, instead capturing standard red-green-blue bands. The properties of this data modality have motivated computer vision analyses including stereo vision and texture-based classification.

Figure 1: Left: The UAV platform landing at a survey site in Julia Creek, QLD, Australia.
Right: The platform was equipped with sensor hardware including a downward pointing colour camera, GPS, and Inertial Measurement Unit (IMU), an on-board computer to log the data to disk and an autopilot system for autonomous flight.

UAV navigation data was also captured during field trials from an on-board Inertial Measuring Unit (IMU) and a Global Positioning System (GPS) receiver. By combining the IMU/GPS information and the overlapping aerial photographs, it was demonstrated that objects identified in the images could be geo-referenced in their environment (i.e. their precise three-dimensional position computed) with an accuracy of approximately +/- 1m. Part of the objective of the current project was to develop methods for processing collected imagery into easily interpretable map formats, such as imagery mosaics.

Classification of vegetation in remotely sensed data has traditionally been approached using multi-spectral information from high altitude aircraft [Klinken07], commercial satellite data [Casady05], or public satellite data such as Landsat [Lawes08]. In spectrally rich data, reflection peaks in the visible and near infrared bands can be used as cellular and chemical fingerprints for different types of vegetation, allowing each pixel to be classified separately [Yu06]. The primary

drawback of spectral classification is that a coarse spatial resolution mixes the response with the background, so these methods cannot resolve small individual trees, or even low density clusters of trees from the background (for example, weed infestations of up to 30% ground cover have gone undetected in 4m/pixel data [Casady05]). Thus, the high spatial resolution of the UAV imagery is critical when the goal is to detect and map individuals of the weed population.

Instead of relying on pixel spectral profiles, the approach taken in this project has been based on algorithms from the fields of machine learning and computer vision, and an interested reader is directed to recent publications coming from this project research including [Bryson10] and [Reid11]. Local spatial pattern, using the red, green and blue colour components together, is used to quantify appearance. Machine learning algorithms can then identify classes of vegetation in the imagery based on the chosen descriptors. In other applications, this problem has received much attention from the computer vision community. The primary advantage of a learning approach is that the highly complex and subtle rules to describe appearance need not be devised or tuned manually – an algorithm can learn them from a set of informative examples [Everingham2011]. However, the subtlety of differences in appearance between different types of trees viewed from above, coupled with the natural variations in an unstructured natural survey environment, and the relatively low number of labels, makes for a unique and challenging vision application.

Finally, we have sought to further analyse the classified imagery to obtain statistics and high-level information that would be of interest to weed managers who adopt UAV mapping technologies. We take a novel approach to tree crown delineation (the problem of finding individual crowns in imagery [Wulder00]), and use this information to derive information about the distribution, cover area and population of individual trees of different species, even demonstrating computer-assisted planning to suggest how points of interest can be efficiently visited given their spatial structure.

2 Project Objectives

Intelligent algorithms have been investigated for the detection and mapping of invasive woody weeds, based on the high-resolution colour aerial photographs and geo-spatial information captured by the autonomous UAV during project B.NBP.0474. The focus of this new research presented has been on analysis algorithms for:

Estimating three dimensional scene structure using stereoscopic techniques:

- Accurately joining image frames into a large scene mosaic/map:
- Reconstructing visually consistent and geospatially consistent scene maps.
- Providing GIS compatible representations of the data

Using state of the art computer vision techniques for object classification:

- Investigating quantitative descriptors to discriminate foliage
- Training machine learning algorithms to detect trees of different species
- Providing Batch classification of flights for mapping vegetation

High-Level Analysis:

- Estimating Tree Crown Locations
- Computer Aided Path Planning for investigation/treatment.
- Building an orthographic map representation of a survey area

3 Methodology (Algorithms)

3.1 Overview of the Algorithm Framework

A new framework of processing algorithms has been developed for the detection of invasive woody weeds in UAV aerial image data. This pipeline is composed of state of the art algorithms that utilise 3D structure in images to produce orthorectified maps, transform the images from pixels into features, and then into probabilistic classifications. Where possible, the framework utilises supervised machine learning techniques, where the rules to discriminate between the different classes selected in the data are automatically learnt from the data (supervised by a set of training examples that are provided to the algorithms). Machine learning is a very elegant and powerful way of configuring analysis algorithms, as it is not necessary for an engineer to design and tune a set of rules. A block diagram of the data flow between algorithms in this project is depicted below:

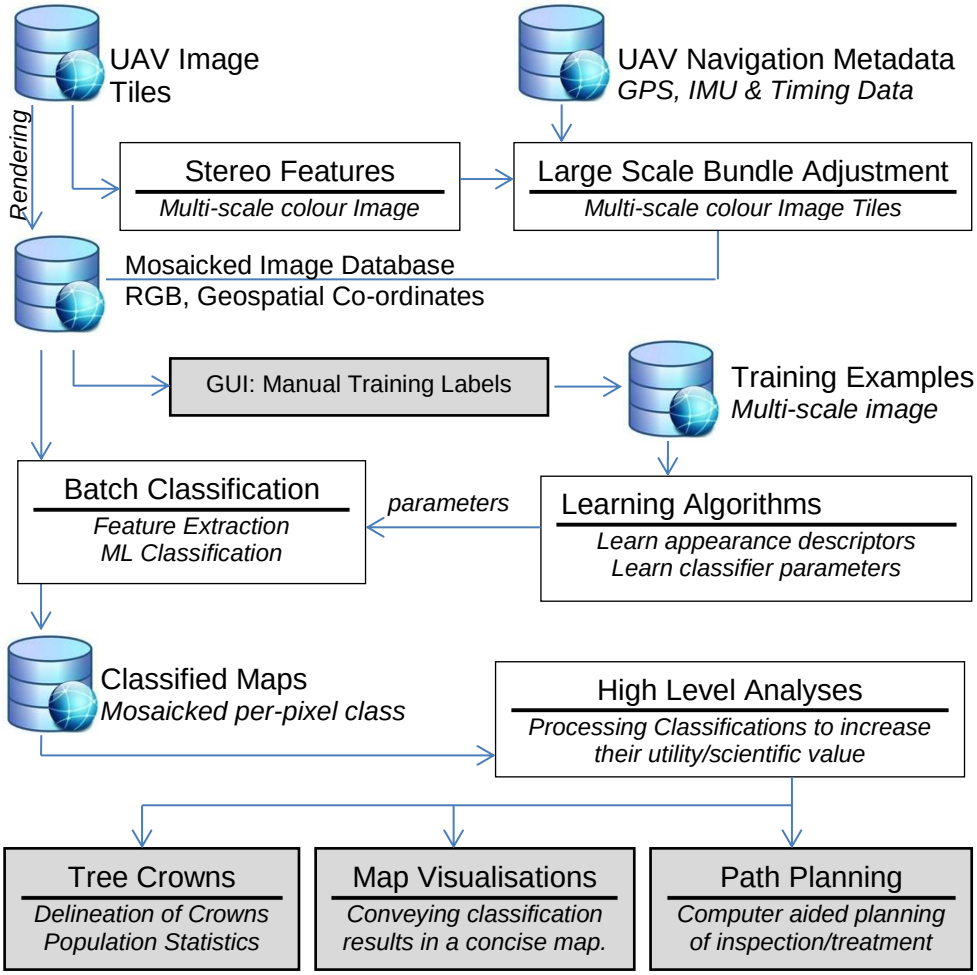

Figure 2: A data flow diagram of the data processing framework. The blue disks symbolise data stores, grey boxes indicate human interaction (manual inputs, or producing outputs)

3.2 Creation of Imagery Mosaics from UAV Data

Instead of performing classification on individual images captured by the UAV, a mosaic-processing pipeline was developed that created large-scale, geo-referenced mosaics from the UAV data that also contained three-dimensional terrain height

information. These mosaics allowed for the imagery to be presented per spatial area, with consistent spatial dimensions and without overlap or doubling counting of the objects classified in the imagery data.

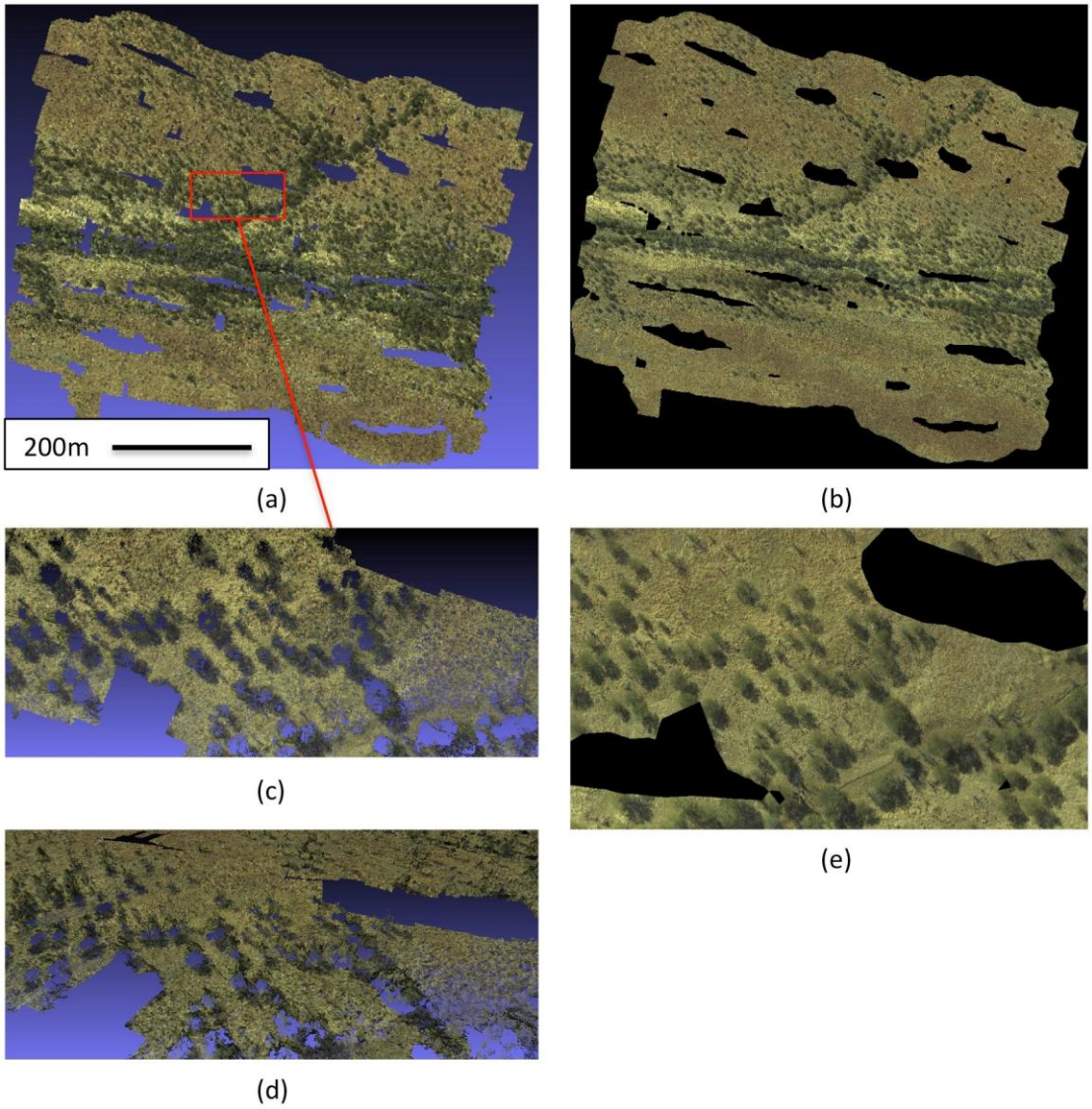

Figure 3: 3D point cloud and image mosaic data for section of Flight 11 (Carrum, 2009), using a local NE reference frame. (a) Colour 3D point cloud data and (b) final colour imagery mosaic. (c) and (d) zoomed-in views of 3D point clouds and (e) zoomed in view of mosaic.

The mosaic processing builds on the existing mapping pipeline developed in [Bryson10] starting from the bundle-adjusted UAV position and attitude information associated with each image captured from the UAV. Taking the pose data and the images, a multi-view stereo algorithm [Furukawa10] is then used to create a dense 3D point cloud of the terrain corresponding to corner points extracted from the image data. This point cloud is used to build a surface model via Delaunay triangulation [Barber96]. For each face in the surface model, the best image corresponding to this face is selected based on the closest distance from each image to the face. Each face of the triangulation is then transformed into a vertical orthogonal projection and a rendering algorithm [Johnson-Roberson10] is used to project the image data into a mosaic raster graph, resulting in the image mosaic.

Finally, the image mosaics are converted into geo-tiffs and are tagged with information about the spatial resolution of the image and its geographic coordinates,

allowing for the images to be interpreted by Geographic Information System (GIS) software. Figure 3 illustrates example outputs from the mapping pipeline including 3D point-clouds of the terrain and the final processed mosaic imagery maps.

3.3 Computer Vision for Woody Weed Detection

The image processing framework has been extensively developed to improve the detection and mapping of invasive woody weeds from remotely sensed red-green-blue aerial images from the project UAV. To achieve this, the computer vision pipeline for woody-weed detection uses the state of the art strategy of transforming the images from pixels into features describing a local region, and then applying a machine learning system to autonomously label the feature vectors based on supervised training labels.

3.3.1 Human Input for Algorithm-Training

A supervised classification system has been implemented, where knowledge is provided to the computer algorithms through a set of training examples. In this case, these examples contain both a local region of the input imagery, and a corresponding, manually chosen output label. The self-configuring algorithms are then trained on this data in order to learn to recognise the pre-selected classes. This is as opposed to an unsupervised method, which would cluster similar image patches together without providing control over how the object classes are divided and the approach may not allocate the woody weeds to different classes, for example.

In this work, the training data is used at two stages – firstly when considering feature selection (where the code to describe appearance is determined), and secondly to train a classification algorithm to recognise new examples of the classes in previously unseen query images. Because the classifier draws on information from its training data, correct labelling is critical to the success of the whole system. To prepare, the non-classified imagery was mosaicked onto a map, and overlaid with ground truth points collected with a hand held GPS receiver during the field trials. For example, Figure 4 shows this ground truth overlay for a region of the Carrum property, while Figure 5 shows an equivalent for a region of the Williams property.

The GPS data cannot provide the training examples directly, because both the UAV imagery and the handheld GPS will be subject to small errors in geo-referencing. In addition, when conducting the ground survey, it was not possible to stand over the trees for recording. Therefore, the surveyed points are typically offset from the trees by a few metres and manual analysis was used as an intermediate step to provide training data that was well registered with the images.

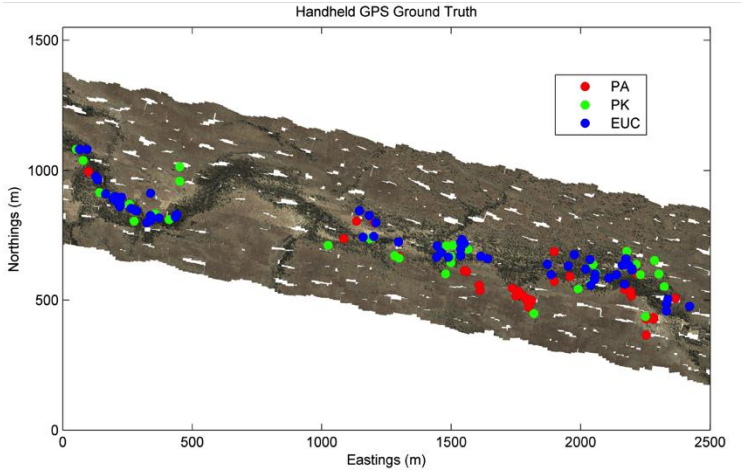

Figure 4: Overlay of ground truth data with aerial image mosaic for project Flight 11 (Carrum, 2009), using a local NE reference frame.

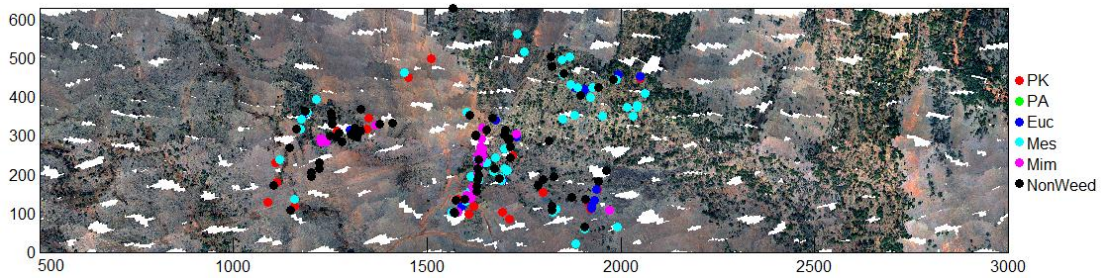

Figure 5: Overlay of ground truth data with aerial image mosaic for project Flight 27 (Williams Outstation, 2010), using a local NE reference frame.

Human input was in the form of a graphical user interface (GUI) where the user may click on the imagery to assign a particular class label. The example in Figure 6 shows how the ground truth data can be matched to the aerial imagery, allowing the GUI operator to recognise the appearance of different types of foliage (tree appearance can change markedly switching from an oblique to a downward looking perspective).

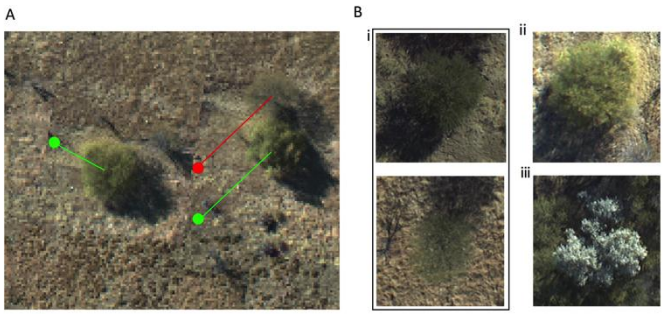

Figure 6: Inspection of ground truth data together with aerial data: (A) The ground points can be associated with tree crowns manually. (B) This particular survey region of the Carrum property in 2009 contained almost exclusively Prickly Acacia (i), Parkinsonia (ii) and Native Eucalypts (iii). It should be noted that the vegetation type was heavily dominated by Prickly Acacia making the dataset highly unbalanced.

Given the limited number of local vegetation types, for each property a number of classes were nominated. In the Carrum property, Prickly Acacia, Parkinsonia, and Eucalypts were each assigned a class. In the Williams property, Mimosa and Parkinsonia were nominated, while the varied remaining trees were aggregated into

a miscellaneous class. In addition, because most of the image pixels do not contain foliage, shadow and background classes were added to both class sets.

While it was originally expected that size and shape might help early on in the classification process (as feature descriptors), it was found that simply looking at local image pixels was more robust at the classification stage: many of the tree crowns are irregularly shaped, partly shadowed, or clustered seamlessly with adjoining tree crowns. In addition, the Eucalypts often have a separated canopy composed of small nearby clumps.

3.3.2 Feature Vectors to Describe Appearance

The first challenge in a computer vision framework is to define a numerical language to represent the appearance of objects in the images. The numerical description is not itself a classification, but seeks to provide a vector of **features** to discriminate between the classes defined in the problem. With a sufficiently rich description, a machine learning algorithm can be trained to label new query data.

As there are only three colour bands in the UAV images, the approach must analyse a local neighbourhood of pixels around each point to get a richer representation. In related applications, a wide range of patch-description strategies have been employed to describe vegetation texture, including co-occurrence matrix statistics [Yu06], multi-scale decompositions such as pyramids [Heeger95], and banks of Gabor Filters [Tang03]. In this project, we have sought to determine suitable descriptors from the data.

With the many possible visual patterns and corresponding feature dimensions, it would be an overwhelming task for a human to design a set of rules and features for this problem without computer assistance. Statistical analyses and learning algorithms provide an elegant solution to this problem because they can automatically form effective pattern recognition on the training data. The following section outlines our approach.

Neighbourhood Patch Sampling

Prior to classification, statistical whitening was used to de-correlate the spectral responses of the dataset's pixels without using spatial information. This does assist the analysis that follows, but it was important to do it prior to the manual labelling because it was found to improve human interpretation of the data. For example, the image on the left of Figure 7 has a strong colour cast that distracts from the informative colour extracted on the right, even though the transform is lossless and we can switch between them in both directions.

The whitening transform applied as a pre-processing step involves statistical analysis of the image pixels. An empirical covariance matrix Σ_{rgb} was assessed over the whole mosaicked flight describing the empirical dependence between a pixel's red, green and blue channels. Matrices containing the eigenvectors V and eigenvalues D of this covariance are then used to de-correlate the channels:

$$RGB_{out} = [V\,(D + \sigma I)^{-0.5}\,V^{T}]\,RGB_{in}$$

Following the whitening, a sampling procedure was specified for the extraction of neighbourhood patches from the aerial images to describe appearance around a given point. Instead of simply sampling a rectangle, to simultaneously consider both fine detail such as leaf structure, and larger detail such as canopy shading structure, a multi-scale sampling construction was defined, where image 5x5 pixel blocks were

sampled at 1x, 5x and 9x scale, looking at 5x5, 25x25 and 45x45 pixel regions. This sampling pattern is illustrated in Figure 8.

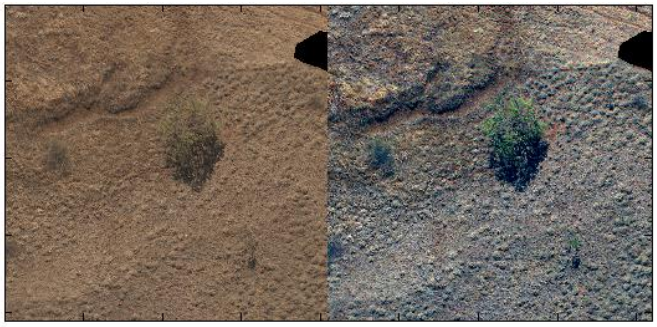

Figure 7: An example of the pixel spectral whitening that is applied as a pre-processing step before classification to assist both in feature extraction and in the supervised human interpretation of the imagery.

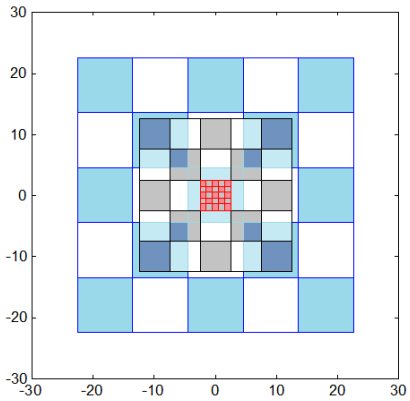

Figure 8: A multi-scale sampling pattern provides a compromise between fine local detail and larger scale structural detail with a linear rather than quadratic increase in input dimensionality.

This sampling was efficiently implemented by caching coarse block responses at 5x and 9x, allowing trivial extraction of the features by accessing cached image elements. Because the coarse blocks are potentially costly to extract from the imagery, a computationally efficient frequency space convolution has been employed. Some resulting multi-scale samples have been visualised below, demonstrating that the fine detail in the centre of a patch can capture foliage texture, while there is still enough detail in the coarse blocks to capture shadows, crown shape and surroundings to put the texture into context.

Stable Features for High Dimensional Data

Using the above sampling strategy, a patch is described by 225 dimensions resulting from 5x5 samples over three bands at three scales. On the other hand, this system is intended to start operating effectively from dozens of training labels. In the case that there are fewer examples of the rare classes in the training sets than the number of dimensions, many learning algorithms will fail to find a data explanation that can generalise to new data. A high dimensional feature space is challenging for a classification algorithm, because the algorithm needs to use increasingly extensive training sets and complex modelling to identify which combinations of dimensions are important to the problem. Consequently, it is common practise in vision problems to compact the raw pixel dimensions into sets of features.

In the literature, the D dimensional sampled vector x is often projected into new d-dimensional feature vector f using a linear weighting matrix W such that f = Wx. A common technique to select W is Principal Component Analysis (PCA) [Jolliffe86] that has been used in vision problems such as face recognition [Tsalakanidou03]. PCA takes an unsupervised approach, and focuses on minimising the reconstruction error by extracting high variance eigenvectors from the empirical training covariance, and throwing away the least noticeable detail. However, with the subtle differences between classes over all the multi-scale dimensions included in this data, PCA was discarding discriminative information.

To reduce the dimensionality of the problem from D=225 down to d<<D, a non-linear transformation was instead employed. A visual dictionary of a few reference patches $P_1 \ldots P_d$ has been learnt (unsupervised) using the training data x. The patches were selected by clustering the training data (using k-means clustering) into d groups, and then using the squared distance to each cluster's mean to calculate the d distance features. In this data, d = 30 reference patches was found to give strong predictive performance. An example dictionary is shown below in Figure 9.

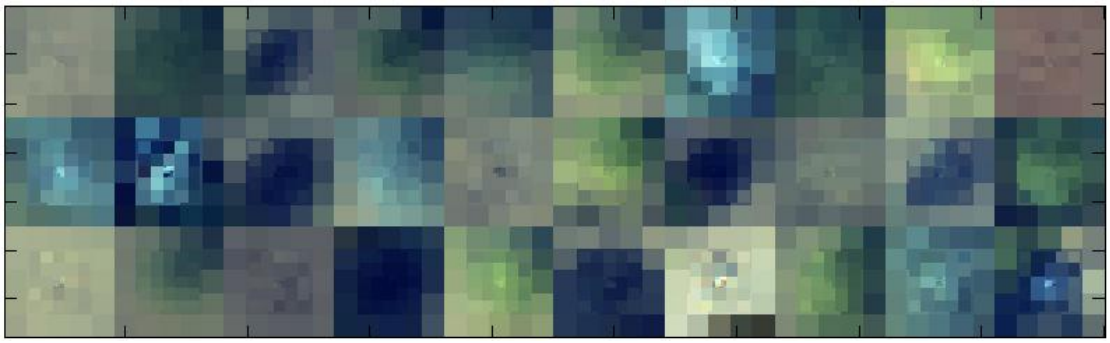

Figure 9: The means of clusters from the training data –these multi-scale patches allow the number of samples to be linear with the width of the patch without sacrificing fine detail in the immediate vicinity of the sample. The distance from the clusters to the query patches form a lower dimensional space where the classifier can find simple data explanations.

Calculating the squared distance is actually as computationally inexpensive as a PCA projection. By expressing $f = \sqrt{\Sigma(x-P_i)^2}$ as $f = \sqrt{(\Sigma P_i^2 + \Sigma x^2 - 2\Sigma P_i x)}$, it can be computed using on only three linear projections where the first two can be pre-computed, and the last uses the cluster centroids as weights.

The vector of distances generalises in a stable fashion from potentially few training vectors, as was needed for this problem. If a query point lies between two of the reference patches, its distances will lie between the distance patterns of the patches. The features were then projected into an over-complete frame, which is equivalent to a basis of a vector space, but allows more basis vectors than the dimensionality of the space. This acts to make the spatial responses sparse (only having a few non-zero features at a time), which helps a classifier form simple rules:

$$W = (f_{train}{}^{T} f_{train})^{-1} f_{train} \quad , \quad F_{query} = f_{query}\, W;$$

3.3.3 Classification

Having obtained a rich set of discriminative features, a classification algorithm is required to map them into class probabilities. Many classifiers in the literature (including techniques such as support vector machines and nearest neighbour techniques) output a 'guess' of the class, while others (including LogitBoost and

Gaussian Process classifiers) predict the class as a set of probabilities. The latter is the more flexible approach, as it allows the user to identify classifier confidence, or even tune the detection process afterwards by applying different thresholds to the probability (although the predicted class is usually equivalent)

The probabilistic LogitBoost algorithm has been applied for its strong probabilistic predictive performance, coupled with its low computational costs. LogitBoost belongs to the boosting algorithm family that improves the performance of a weak learner by training an ensemble and weighting their outputs [Freund99]. The model complexity is varied through selection of the number of learners in the ensemble, and probabilistic outputs are obtained by statistically minimising a logistic loss function. In this work, the LogitBoost algorithm has been applied to boost the one level decision stump (equivalent to an ensemble of if-then-else rules).

The complexity of the decision rules encapsulated in a LogitBoost classifier is controlled by the number of weak learners in the ensemble. The number of stumps is optimised using cross validation (withholding some of the training data for testing). Too few rules will cause the classifier performance to be poor, while too many rules will allow over-fitting, where the classifier can perform well on the training data but poorly on the query data.

3.4 Training Set Cross-Validation

Ten-fold cross validation has been measured over training examples collected from the aerial data. Data points were divided into ten random groups. In rounds, nine groups have been used for training the classifier, and one group has been withheld for testing. The most likely class from the cross validated predictions have then been used to build confusion matrices that contain the actual class along rows, and the predicted class down columns. These matrices allow us to identify the likely sources of error in the open-loop prediction when the trained classifier is either employed to batch-classify a region of interest, or an entire survey area. Within the confusion matrix, an ideal classification result is to place all outputs in the correct class, creating a fully diagonal matrix. To assist in the interpretation of these matrices, precision (the fraction of positive detections that are true) and recall (the fraction of true class occurrences that were detected) statistics have been derived.

We note that the survey regions have been pre-selected for their heavy weed infestations, which has led to an oddly unbalanced classification problem where the data contains far more woody weeds than native trees (in fact, our results later in the analysis suggest that at the time of our 2009 field trials, 98% of the tree crowns in the chosen Carrum survey region were woody weeds). Unbalanced data is a difficult problem for a classification analysis to overcome because the ‘best accuracy’ behaviour is to forget about rare classes and learn a model for the frequent classes only. For example, the classifier that says ‘all trees are woody weeds’ would score a 98% accuracy in the Carrum data, but clearly the results would not be meaningful if we train for this criterion. If (as is the case here) all classes are important to detect, then this can be prevented through prioritised sampling from the less from the frequent classes. By sampling a more balanced training distribution, the classifier has been able to consider the possibility of the infrequent classes occurring, at the expense of a higher exposure to false positives and average error rate.

Carrum Property *- 10 fold cross validation*

Table 1: Confusion Matrix from 10 fold cross validation

Decision \ Actual Class	PA	PK	EU	SH	BG
PA	66	5	0	0	4
PK	5	24	0	0	0
EU	0	0	30	0	1
SH	0	0	0	33	0
BG	0	1	0	0	49

Key: PA (Prickly Acacia), PK (Parkinsonia), EUC (Eucalyptus), SH (Shadow), BG (Background)

Table 2: Precision and Recall Statistics

Mean Accuracy 92%	PA	PK	EU
Precision %	88	83	97
Recall %	93	80	100

We conclude that there is excellent discrimination between the natives, woody weeds, shadows and background. A fraction of the prickly acacia have been missed and labelled as background – examination of the image tiles later suggests the foliage can be very transparent when viewed from above and likely the background was showing through due to lighting conditions.

Between the woody weeds, the classifier performance is not as strong, but it still achieves precisions of 88 and 83% for PA and PK respectively. These woody weeds are occasionally confused due to their similar size, shape and colour, although the classifier is correct the majority of the time. Later inspection of the images shows this confusion lies largely in the shadowed parts of the tree crown and can be post processed using techniques such as median filtering.

From the ground truth labels, we were able to confidently cross reference sufficient labels of mimosa and parkinsonia woody weeds to support classifier training. A variety of other unlabelled species were present in this survey region and have been grouped into a general class for classifier training purposes.

Williams Outstation – 10 fold cross validation

Table 3: Confusion Matrix

Decision \ Actual Class	MIM	PK	OTHR	SH	BG
MIM	22	3	1	0	0
PK	2	23	0	0	1
OTHR	2	1	20	1	0
SH	0	0	0	12	0
BG	0	0	1	0	21

Key: MIM (Mimosa bush), PK (Parkinsonia), OTHR (other trees), SH (Shadow), BG (grass/background).

Table 4: Precision and recall statistics

Mean Accuracy 89%	MIM	PK	OTHR
Precision %	85	88	83
Recall %	85	85	91

The confusion matrix shows that the classifier has performed very well in separating tree crowns from the background, and also well at separating the nominated species (mimosa and parkinsonia) effectively. A richer set of ground truth labels would enable the third general class to be split into additional categories, and it is also noted that the slightly lower overall accuracy compared to the previous validation is likely due to the less extensive label data available. It is expected that more labelling would have improved the classifier performance to the level of the previous scenario over the Carrum site.

3.5 Comparison to Project B.NBP.0474

The results from the Williams Outstation and Carrum Property are compared to those from the previous project (B.NBP.0474). The vegetation classes are extracted from the confusion matrices to compute the overall precision and recall statistics.

Table 5: Precision and recall statistics

	PK	PA	EUC	ME	MIN	Average
2011 Precision %	48	-	84	86	15	58.3
2011 Recall %	53	-	85	83	20	60.3
2012 Precision %	85	88	97	-	85	88.8
2012 Recall %	82	93	100	-	85	90.3

Some classes are missing due to the slightly different evaluation sets used. Overall there is a clear improvement in the classification accuracy in the current approach.

The overall precision improves from 58.3% to 88.8%, whereas the recall improves from 60.3% to 90.3%. The improvement is due to the more expressive feature descriptors used in this project: in B.NBP.0474 only the local pixel colour were used for classification whereas, in this project, multi-scale image patches were used to encapsulate the regional image properties.

3.6 Visualisations of Classified Images

While validation data does not exist for all the pixels in the image datasets, the visualisation of classified outputs can still be very informative in assessing the strengths and weaknesses of the classification framework. Efficient Matlab code has been implemented to load mosaicked image tiles, extract features from the imagery (intelligently handling tile edges by loading adjacent views), and pass these features to the pre-trained classification algorithm. The software currently takes approximately 3 seconds to fully process a raw UAV frame on a consumer-grade laptop, or more relevantly, about 3 seconds to process a 35x35m tile of mosaicked imagery. This time includes image loading, feature extraction and classification. This gives us the option of either processing a raw flight (taking approximately 12 hours), or mosaicking prior to classification taking an equivalent of 3 hours because redundant overlap has been removed. The smaller regions of interest only take a fraction of this time. We also note that the processing of image tiles is easily parallelisable because different computer cores can load and classify frames simultaneously. Distributing the classification work onto a cluster or even a multi-core desktop has not been implemented, but would easily allow the mosaicked-tile classifications to be completed within 1-2 hours of the image mosaic being completed.

The images below in Figures 10-16 are a selection of outputs from the batch classifier. In these images, there are more classes than available display channels. Thus, target species have been mapped to red (Prickly acacia in Carrum and Mimosa in Williams), green (Parkinsonia) and blue (Eucalypt in Carrum, general OTHR class in Williams). In addition, shadows have been mapped to a black output, and background to a white output. The most likely class is used to colour the output image, overlaid with a 25% weight on the original RGB shading:

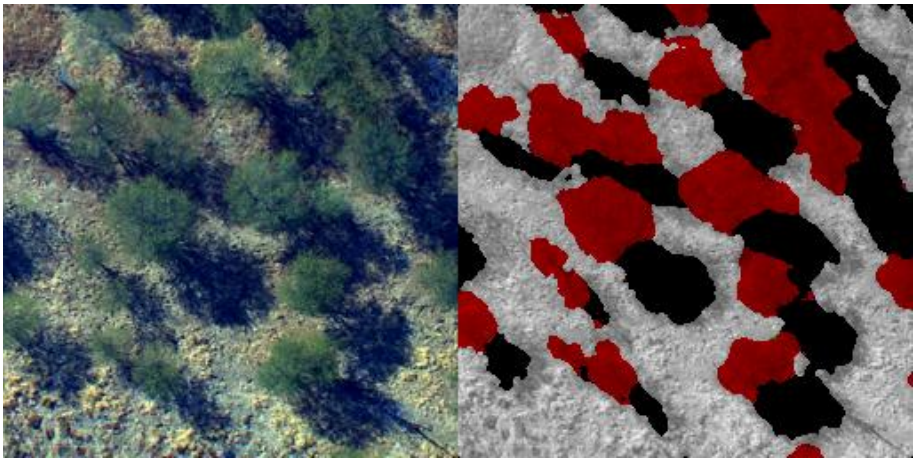

Figure 10: Successful detection of prickly Acacia crowns and their shadows. (Left: original image, right: classifier visualisation) in the Carrum dataset.

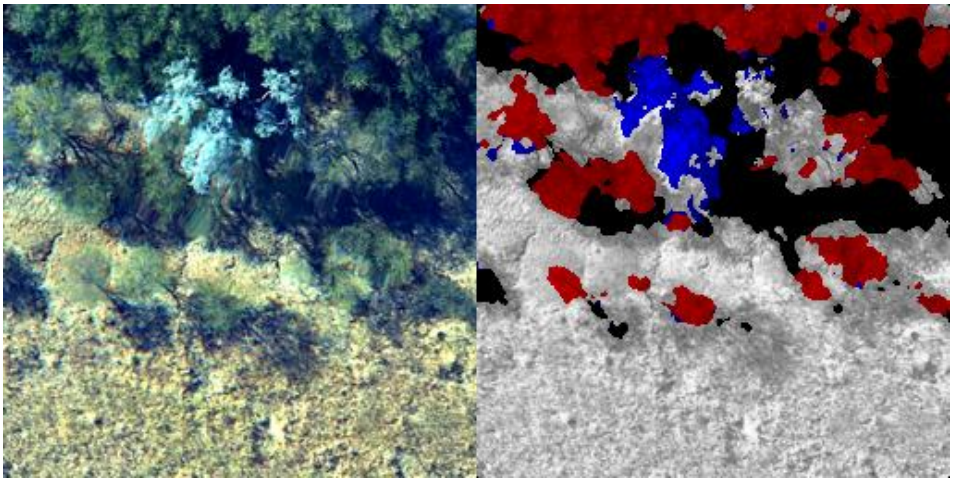

Figure 11: Detection of a eucalypt crown (blue) amongst shadow and a joined canopy of prickly acacia (red). The speckled detection pattern of the eucalypt is related to the patchy appearance of the canopy.

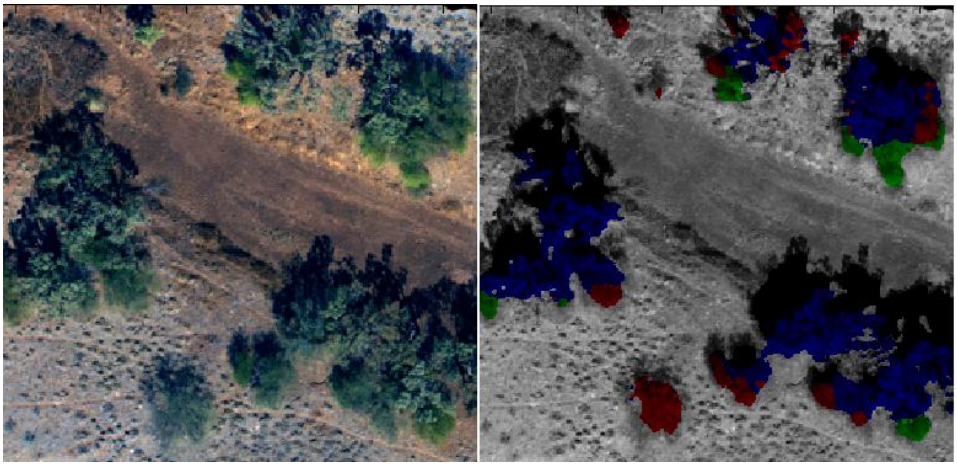

Figure 12: Detection of general trees (blue) against mimosa (red) and parkinsonia (green) in the Williams dataset.

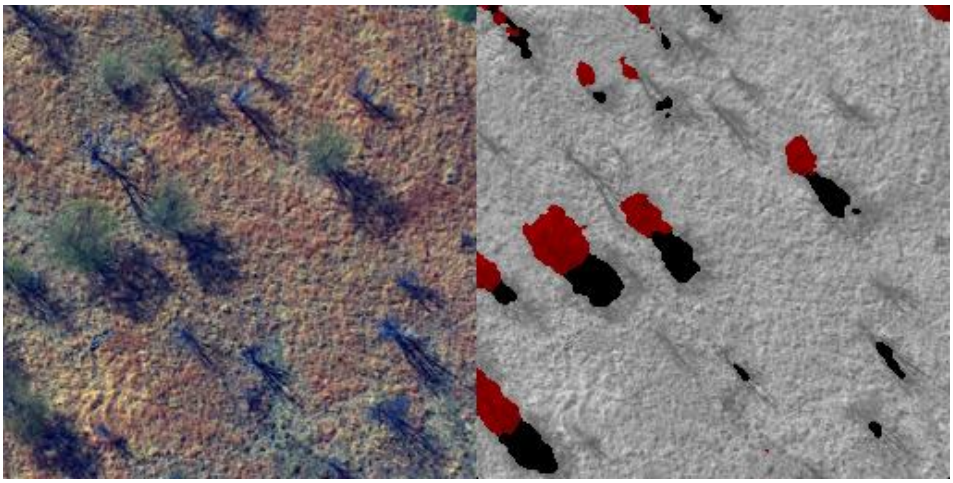

Figure 13: Correct classification of prickly acacia in a low contrast scene. Also note some trees appear to be dead or defoliated, and have been detected as shadows with no associated crowns. However, we also note that because of the multi-scale spatial sampling, the shadows actually strengthen the detection of these free standing crowns.

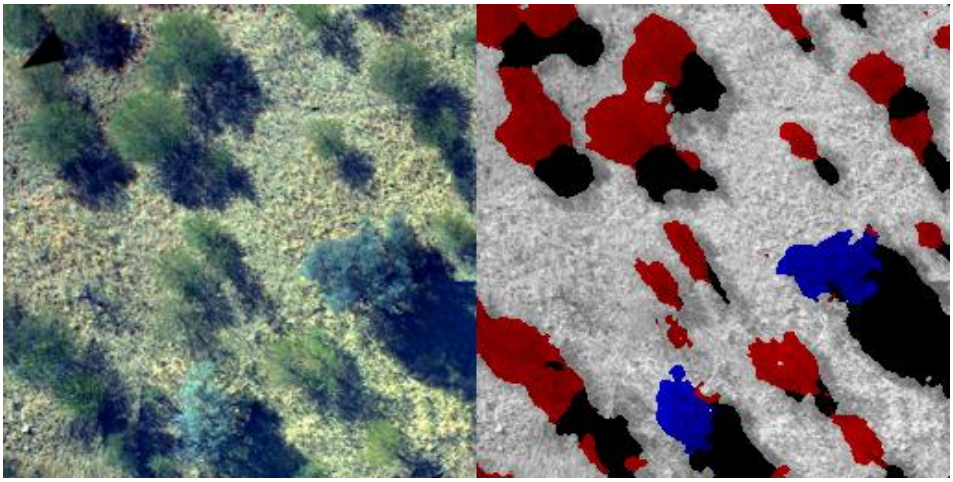

Figure 14: Correct classification of Eucalypt, prickly acacia, shadow and ground.

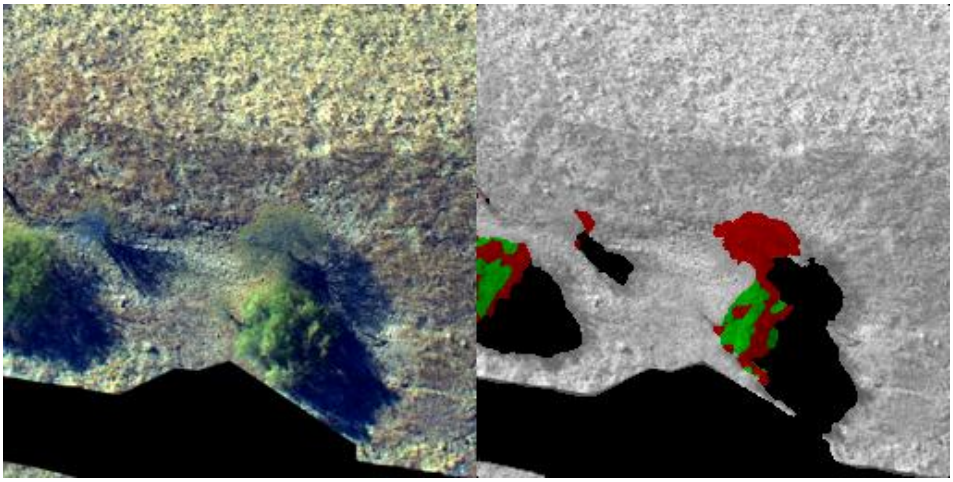

Figure 15: The fraction of parkinsonia in the training data has been reduced to prevent false positives on the highly unbalanced prickly acacia class. This means they are still detected (green), but part of the crown will usually appear as prickly acacia (red). This particular phenomenon has been addressed in post processing.

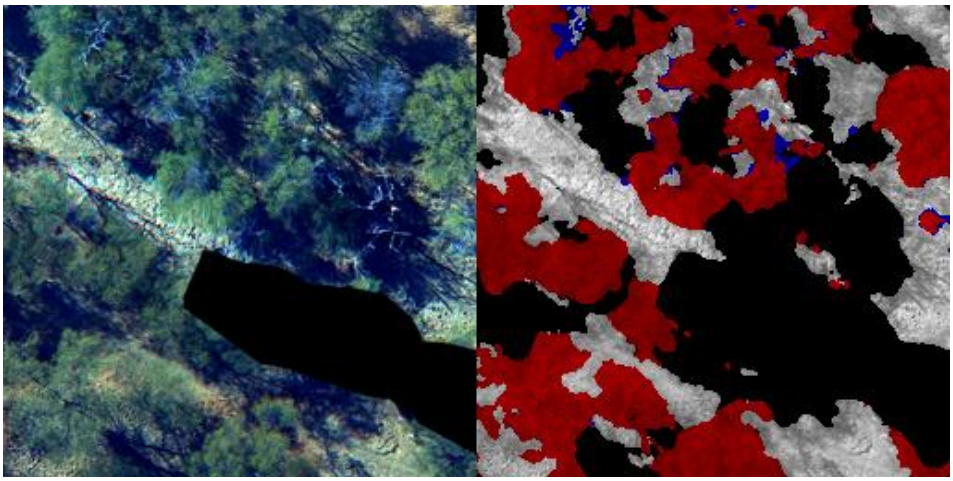

Figure 16: Dead trees can have a similar appearance to an isolated patch of eucalypt. For example, these dead trees have caused the classifier to exhibit a small response to the Eucalypt class (which has been subdued with median filtering later), although the majority of these are classified correctly to begin with.

3.7 High Level Analyses

The above Figures 10-16 show that the classifier is generalising its detection behaviour reliably beyond the labelled points in the training data. Once the targets of interest have been classified over the image pixels, it becomes possible to propagate the information into more accessible information modalities such as maps and survey plans. Prior to doing these analyses, a very simple post-processing was applied to the classifier outputs – the standard closing morphological operator was applied to close the gaps in each class output to handle problems such as the patchy Eucalypt crowns, or the partial mixing of Parkinsonia crowns.

3.7.1 Classified Regions of Interest

This section presents batch classified imagery, both integrated and decoupled from the mosaicking results.

Figures 17 and 18 show a map, and the corresponding classifications produced using the raw UAV navigation data to project each frame to a flat ground model. If classification is done pre-mosaicking then the narrow field of view makes it difficult to capture multi-scale data, so a central patch with principal component features has been used for this specific case.

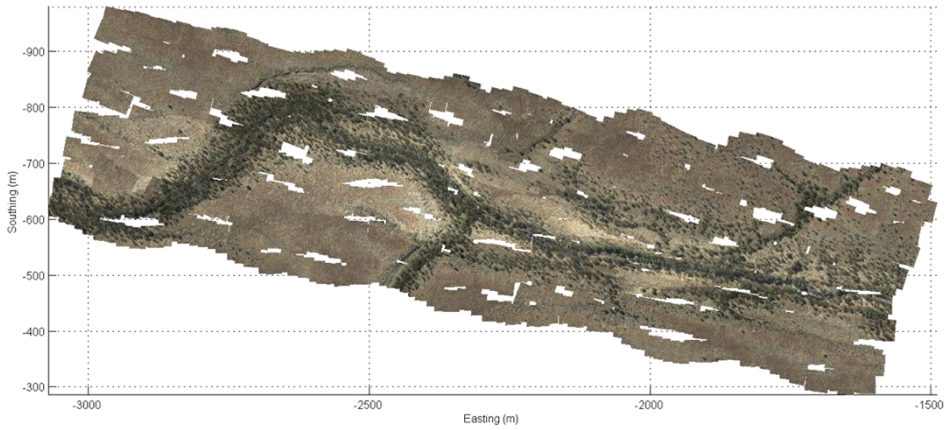

Figure 17: Projected aerial imagery of the Flight 11 survey region. This is simply a flat ground projection from the raw navigation information. The new mosaicking system used in the following figures has further improved the alignment quality using visual cues together with the UAV navigation solution.

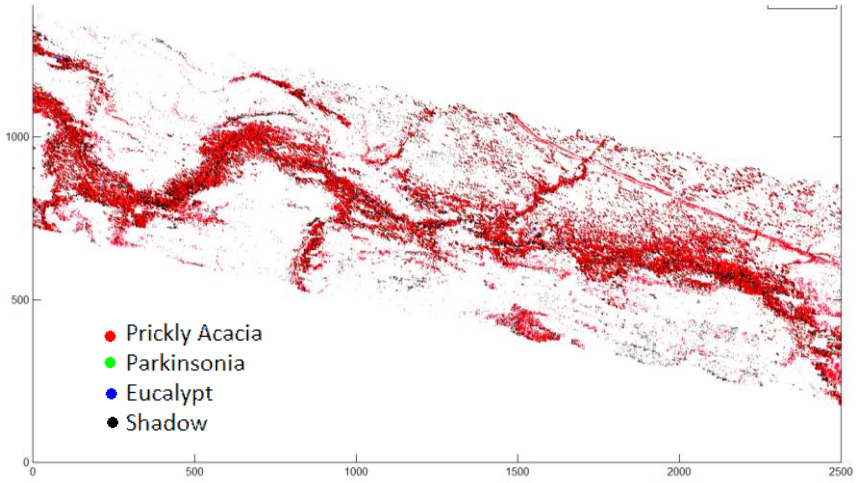

Figure 18: A large scale visualisation of individual UAV-frame classifications clearly indicates the unbalanced nature of the classification problem – the survey site is dominated by prickly acacia. However, there are some other problems with tree crown alignment at the fine scale.

The scenarios below have used the multi-scale features. They also benefit from mosaicking to align the images. If we first mosaic the imagery (aligning the imagery using both navigation data and visual cues as in Section 3.2), then the map output is spatially consistent, the amount of redundant classification processing is reduced, and we can extract consistent, larger scale features across the edges of tiles. This has produced the high fidelity regions of interest (Figures 19 and 20)

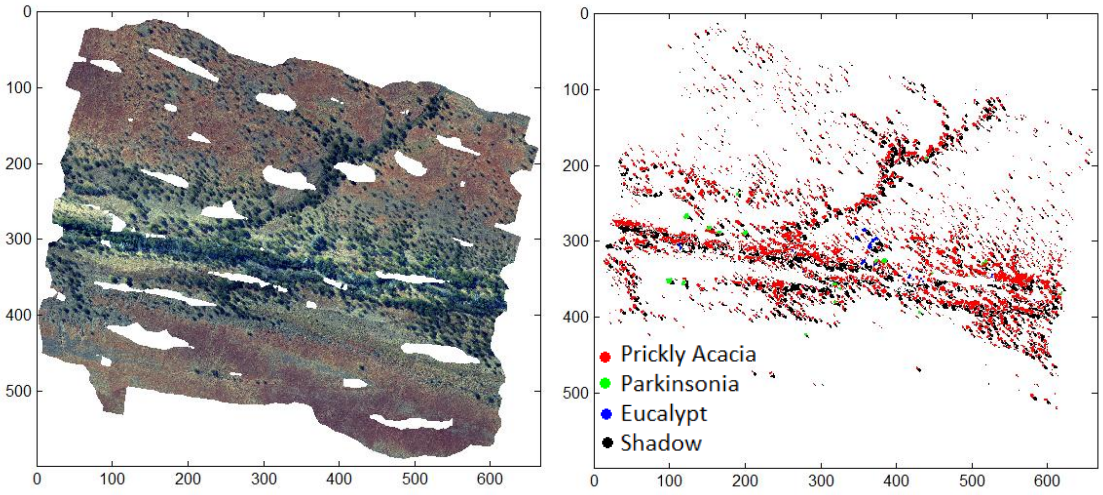

Figure 19: A mosaicked and classified region from the Carrum site. Prickly Acacia (red), Parkinsonia (Green) and Eucalypt (Blue) are marked on the map. Coordinates are given in metres in a local reference frame.

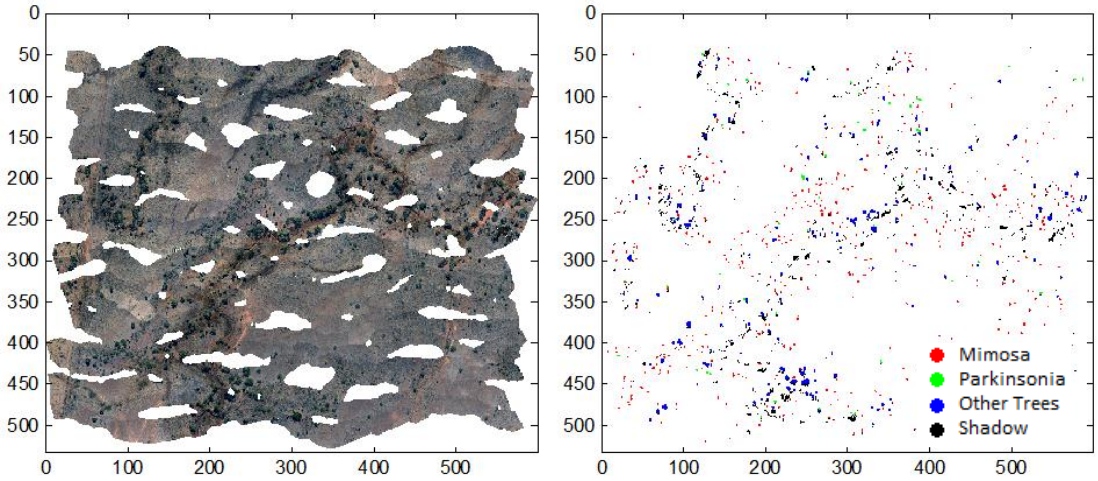

Figure 20: A mosaicked and classified region from the Williams outstation site. Mimosa (red), Parkinsonia (Green) and Other trees (blue) are marked on the map.

3.7.2 Delineation of Tree Crowns (enabling Tree Counting)

As discussed previously, the problem of finding individual tree crowns in a fully connected canopy is known as tree crown delineation. This is a topic of interest in the remote sensing community and a number of publications have used heuristics such as finding local maximum shading peaks [Wulder00].

Tree crown delineation is critical to estimating tree counts, as well as providing spatial co-ordinates in terms of latitude and longitude. An effective technique has been found by investigating the above mapping results and considering two distinct cases: some trees are sparse and fully separated, while others are part of a connected canopy. These cases are depicted in Figure 21:

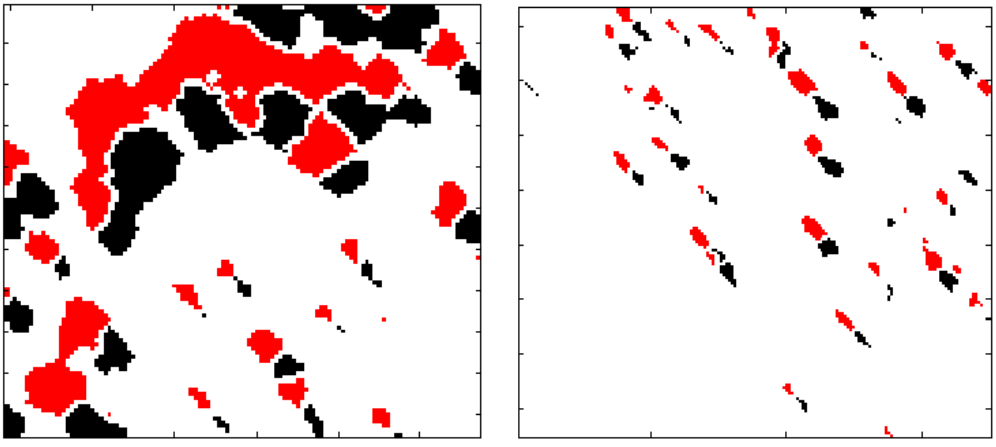

Figure 21: The two distinct tree crown delineation cases are handled differently: (Left) Connected Canopy Silhouettes and (Right) The relatively trivial sparse case.

In the sparse case, it is relatively easy to traverse sets of connected pixels using a standard algorithm such as breadth first search. Figure 22 below depicts the classifier detections, drawing red dots onto the tree crowns:

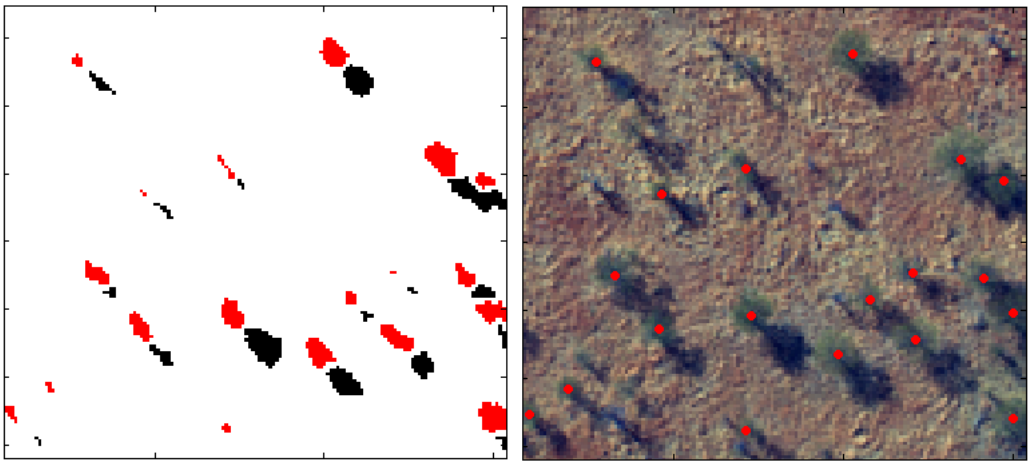

Figure 22: For the simple case of sparsely connected tree crowns, their centroids can be obtained with image processing algorithms to extract the centroid. Left: raw classifier outputs, Right: Centroids marked onto the RGB image.

The second case is much more difficult. We are faced with the task of breaking the silhouettes into crown regions – illustrated below in Figure 23.

Figure 23: The goal of the cluster delineation algorithm is to split a classified silhouette of a cluster of trees into estimates of the locations of the tree crowns that compose it.

An effective algorithm has been developed for this second case. The grouped canopy is detected by running an area analysis on all the connected regions. Those with abnormally large area (compared to the mode of the detections) are assumed to be clusters, and the number of constituent trees is estimated based on the group canopy area. The spatial coordinates of all the constituent pixels are clustered using the k-means clustering algorithm, relying on the principle that the pixels of each crown will be closer to its centre than to any other crown's centre. This approach has proven highly effective – these estimated crowns are drawn with red * markers rather than dots so the user can see which algorithm has been applied in each case.

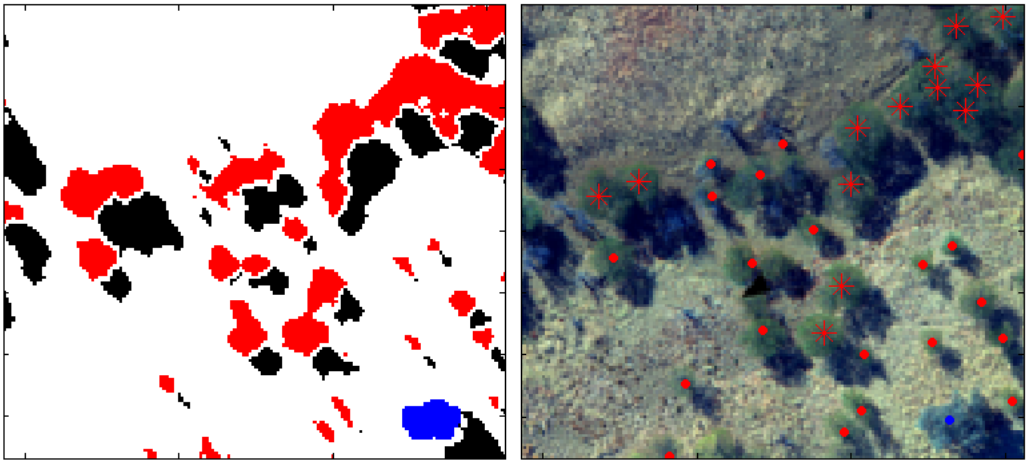

Figure 24 - a combination of single and connected tree crowns has been split into an estimated set of tree crown coordinates. Those obtained with the centroid algorithm are marked with red dots; those with the clustering algorithm are marked with stars.

Inspection of Figure 24 shows this algorithm is producing state-of-the-art delineation. It is particularly effective on almost-separated trees. This analysis relies on the mosaicking process to avoid counting tree fragments multiple times, and has been applied to the regions of interest displayed previously to obtain some tree count estimates. Cover area can also be easily estimated by counting pixels of each class.

Carrum ROI (264300m$^{2)}$ Williams ROI (225300m^2)

Class	Trees	Cover %	CLASS	Trees	Cover %
PA	2230	13.2	MIM	583	0.9
PK	20	0.3	PK	49	1.4
EUC	16	0.2	OTHR	1007	3.6

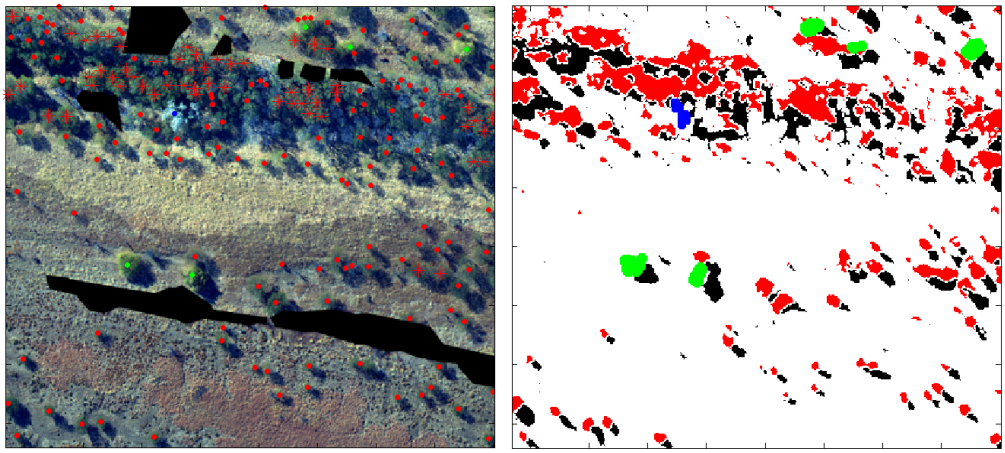

Figure 25: A general view of part of the region of interest labelling tree crowns of different types. Note that there are some gaps in the UAV coverage that are not classified – this cross track error is caused by wind gusts rolling the UAV.

3.7.3 Path planning case studies

Once tree crowns in the classifier output images have been delineated into spatial coordinates, the possibility of computer-aided planning comes into play. Depending on the desired application, computer aided path planning has the potential to program waypoints into a handheld GPS system for human officers to conduct further inspection and treatment, but also opens the possibility of a robotic system planning actions responding to the survey information, for example sending the project's hovering aerial vehicle in for a closer look, or navigating a ground robot to administer treatment. These possibilities are explored here through a number of planning scenarios built around the mosaicked region of the Carrum property.

In all the below cases, the path planning problem has been constructed by forming a graph over the survey region. Each target tree is assigned a node, and edges between all pairs of nodes represent potential paths the human or robot would take to traverse between them. Consequently, these edges have associated cost penalties. For example, to visit all the non-acacia trees in this scenario, and assuming we travel in straight lines between them, the planning graph resulting is depicted in Figure 26.

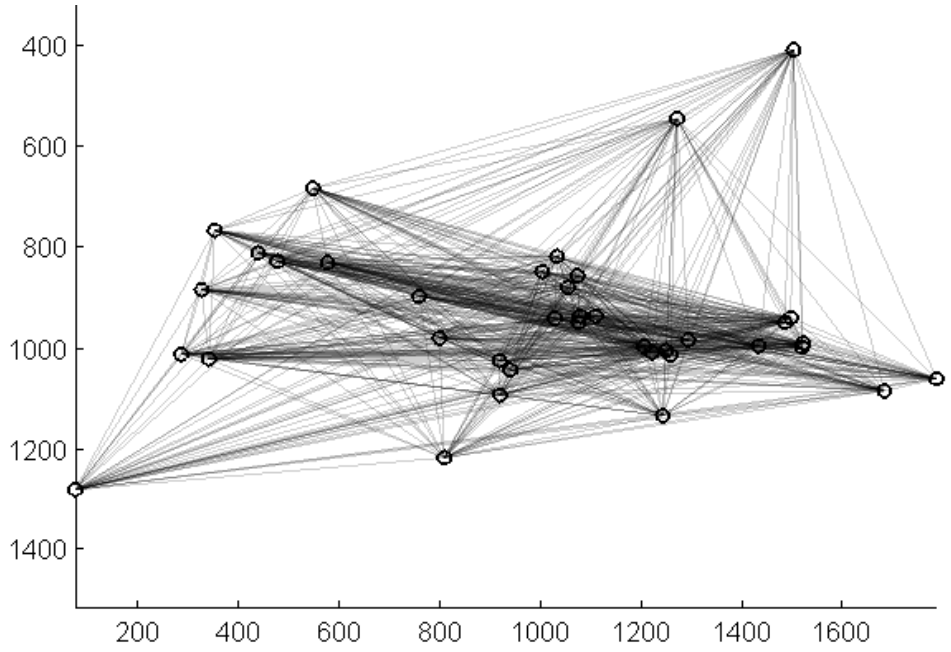

Figure 26: A graph over a set of target trees (nodes) in the constructed path planning problem, with direct edges between them. This forms a large space of paths from which to select a tour that visits each target once.

Even in its simplest form, selecting a tour of the graph is a challenging problem. The core task is to visit each of the targets using a path that minimises the total cost. For more than a handful of targets, this becomes a daunting combinatorial problem of N! possible orders to visit N nodes. In fact, this is an applied form of the well understood travelling salesman problem from the computer science literature, for which it has been proven that it is not computationally possible to solve the problem in polynomial time. Instead, we must approximate the optimal path.

There have been many efficient approximations proposed for this problem [Hassin00], some with performance guarantees, and others with a high probability of finding a good path. In this work, both are applied. The deterministic Christofides approximation algorithm is initially employed to generate an approximate path guaranteed to be at worst 50% longer than the cost-optimal path. This is then improved by using the 3-opt algorithm to consider randomly re-ordering parts of the best known sequence searching for a cheaper solution. This typically untangles the path from crossing over itself and can further improve this bound close to 5% above optimal. Thus the resulting paths are expected to be near optimal while guaranteed to be within a certain excess.

A complete tour of all detections

Our initial treatment is to assume that distance (calculated by the length of the line between nodes) is the cost of a path. This allows us to assess the problem of visiting all the trees. To avoid an excessive waypoint list, the tree crowns have been collected over a 20 metre grid, determining the mean position of trees in each cell (if there is only one tree, the point will lie over the crown). It is assumed the human or robot will want to start and finish at a station which has been added as a node in the graph marked with an X. The resulting tour is shown in Figure 27.

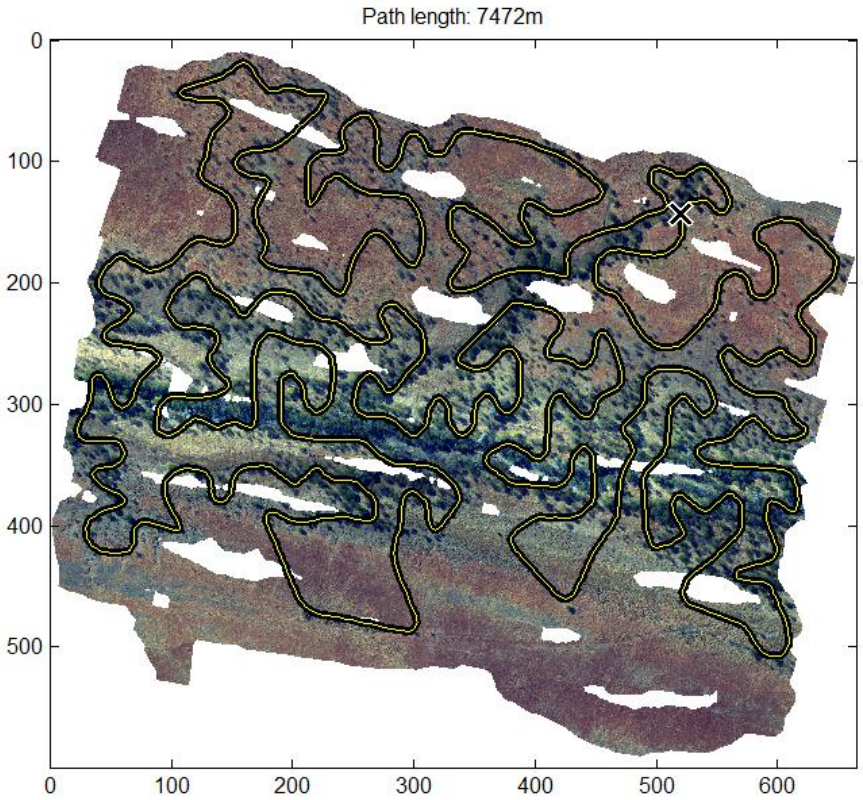

Figure 27: A naïve path visiting all tree detections produces an inticrate tour of the region. This is certainly shorter than a systematic raster scan, but in many cases would not be an ideal strategy.

Firstly, it is immediately apparent that the tour is still daunting from an operations perspective. It is guaranteed to be shorter than a simple raster scan, but perhaps too complicated for a human to realistically follow. In this case, a UAV collecting a secondary modality of data could be deployed, or the path could be simplified. In this case, the path has been smoothed by generating the straight line path in one metre

steps, and applying a weighted average filter independently in the x and y dimensions. This leads to a smooth continuous path that is still guaranteed to fall within a certain range (in this case approximately 5) metres of the target points.

A prioritised path for a hovering vehicle

A hovering UAV such as the secondary, rotary wing platform used in the original project has been considered. It would be useful to send this vehicle for additional image data, but not efficient to re-visit all detections. Strategies such as pointing a gimballed camera would allow the vehicle to travel close to targets, while it is important that the path is smooth and continuous, so the smoothing has been increased so the path lies within 15 metres. We consider that a reasonable task would be to fly over the areas where the less common classes were found (which in this case corresponds to visiting all the trees that are believed to not be prickly acacia). The resulting tour is shown in Figure 28.

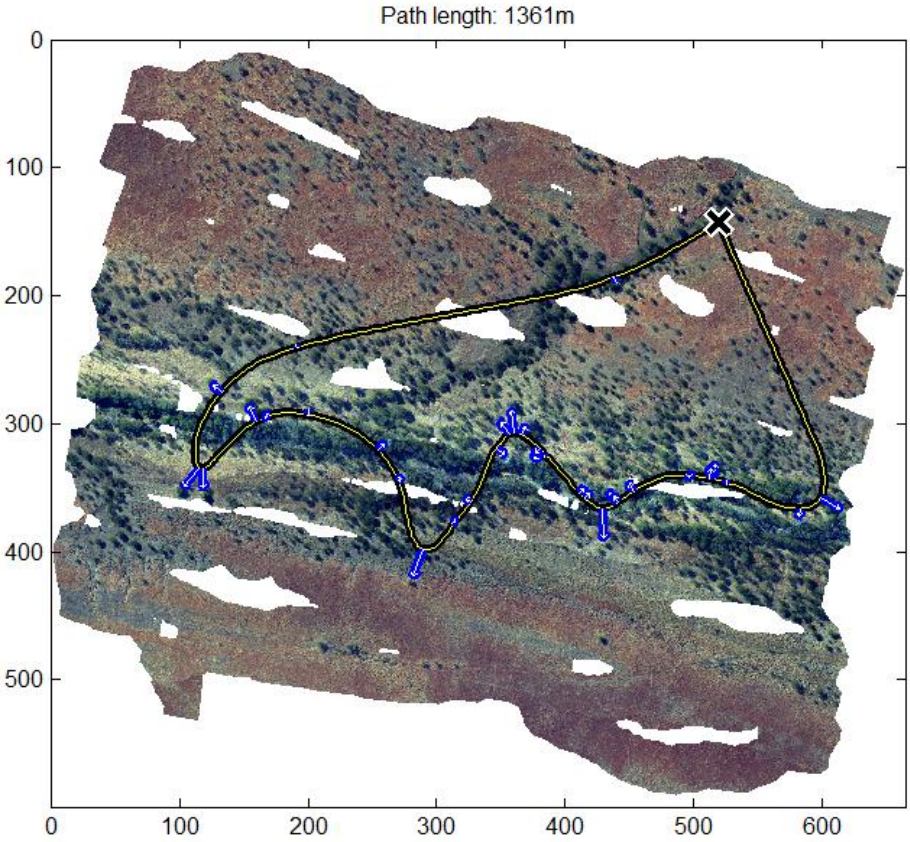

Figure 28: a smoothed travelling-salesman solution to visit a subset of the targets, assuming that traversibility is dominated by distance and path smoothness. The path was not constrained to exactly pass through the tree crowns.

A 100-waypoint path for a handheld PDA

A human can be expected to investigate their surroundings, so the main advantage of a planning system for use with a human survey team would be to identify an efficient ordering of waypoints, and to place virtual GPS markers to prevent the field team from missing locations (which is very easy to do in a large forested region). This can be achieved simply with a coarser block resolution, and omitting the path smoothing which would be difficult to convey over a GPS waypoint list, resulting in the ordered waypoints (black dots) shown in Figure 29.

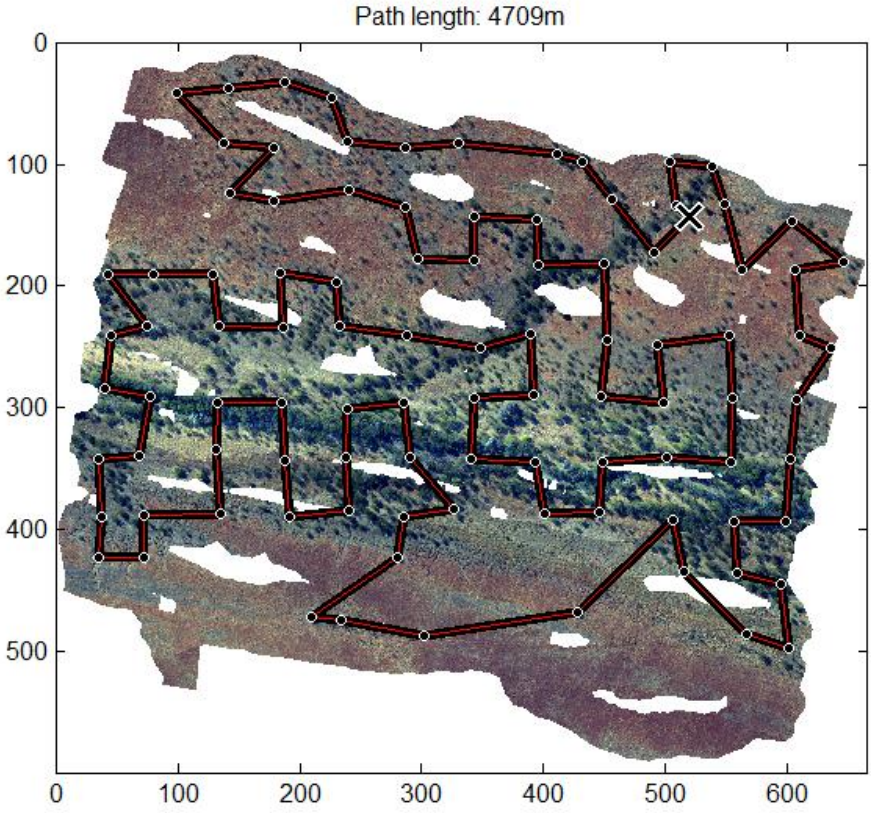

Figure 29: A path with reduced waypoints (100 not including the start location) to ensure it is manageable on a GPS system and no regions are missed.

A traversable path for a ground vehicle:
In the previous cases, the cost of traversal has been associated with distance alone. However, it is possible to consider the terrain structure based on the mapping data collected. In this example scenario, a ground robot with motion constraints is considered. For example, the vehicle is primarily expected to navigate around trees. In addition, it has a secondary goal of information gathering and should avoiding travelling over empty ground when an alternative path that gathers visual information is available.

These two criteria have been constructed as a simple spatial traversal cost function penalising movement through one pixel to the next in the map's image-space. Path cost can then be evaluated as the cost as the shortest path through the image graph. Connections have been simulated at the pixel level by treating the image as a very large (but very sparse) 8-connected graph (with diagonal connections weighted to accommodate the additional length). This function, and the connectivity used to apply it to the image, is shown in Figure 30 below.

From this non-linear connectivity mapping, it is possible to infer the cost-optimal paths by following the minimal cost cells back from the destination to the source of the simulation. This simulation has been run from each planning graph node to obtain a graph with equivalent structure, but non-trivial edges and edge costs.

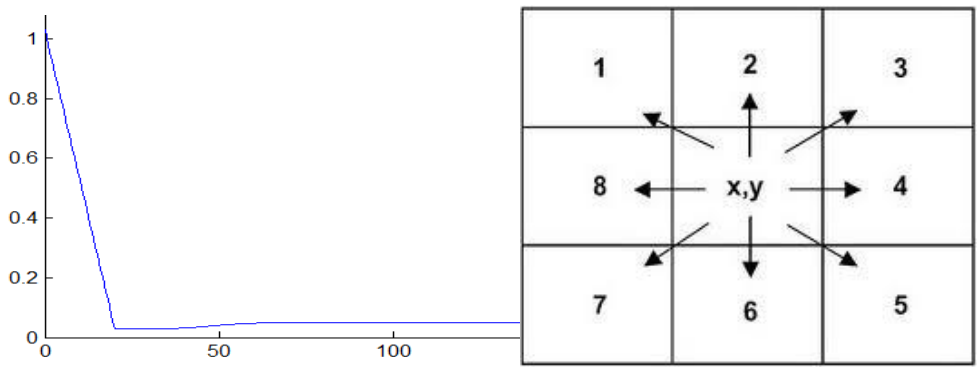

Figure 30: example cost function dependent on distance to the nearest tree. It heavily penalises distances shorter than 10m, also slightly penalises distances longer than 40m (empty space). Unexplored areas have also been made costly to traverse (to avoid going through trees that weren't surveyed). There is also a small positive base cost penalising distance travelled.

The cost function and pixel connectivity are then applied to the image using Dijkstra's shortest-path algorithm to propagate paths outwards from each target to all other points in the image space. This produces a non-trivial edge cost map between the source point and any other point:

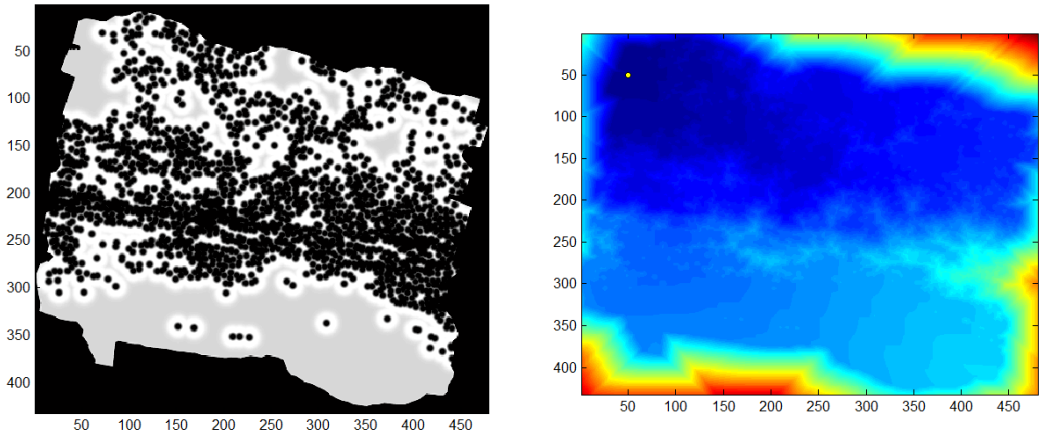

Figure 31: Left: The cost-function image, and Right: the result of propagating paths from the yellow dot in the top left corner to arbitrary positions in the cost map. Note the highly nonlinear patters induced by the tree positions.

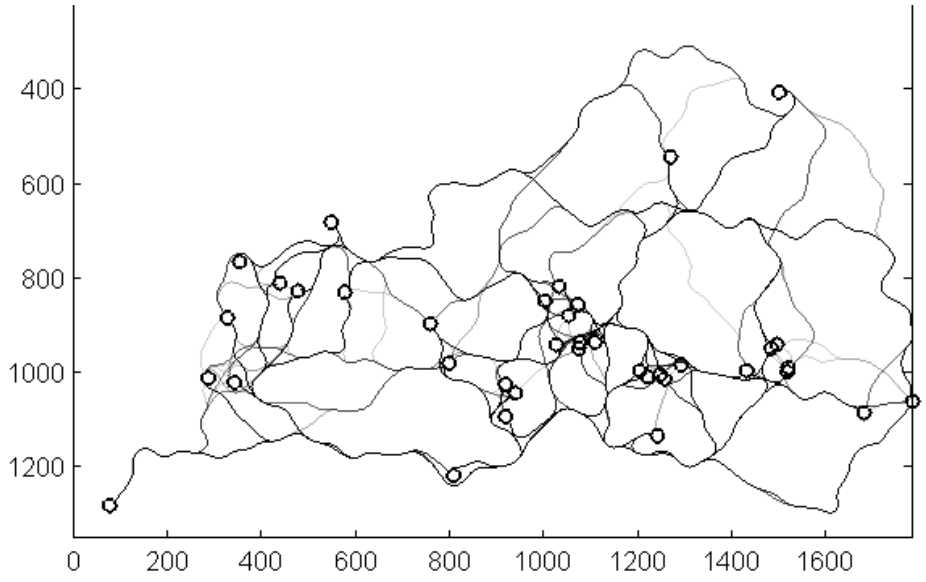

Figure 32: The cost-optimal paths between nodes (targets) are no longer straight lines due to our constructed cost function, but the travelling salesman algorithms can still find a tour provided these costs are reflected in the graph edge weightings.

Interestingly, it is apparent that with the chosen cost function formulation, the path finding tends to converge into common 'roads' between the nodes, differing mainly at the end points. We will look at these roads shortly. It was also considered that the ground vehicle will be able to start and finish at different points (such as gates or roads) without having to return to the origin like an aircraft. This can be achieved by placing a start and finish node into the graph with a zero cost edge connecting them.

A path through this graph can then be found using the same travelling salesman approximation techniques as were applied before:

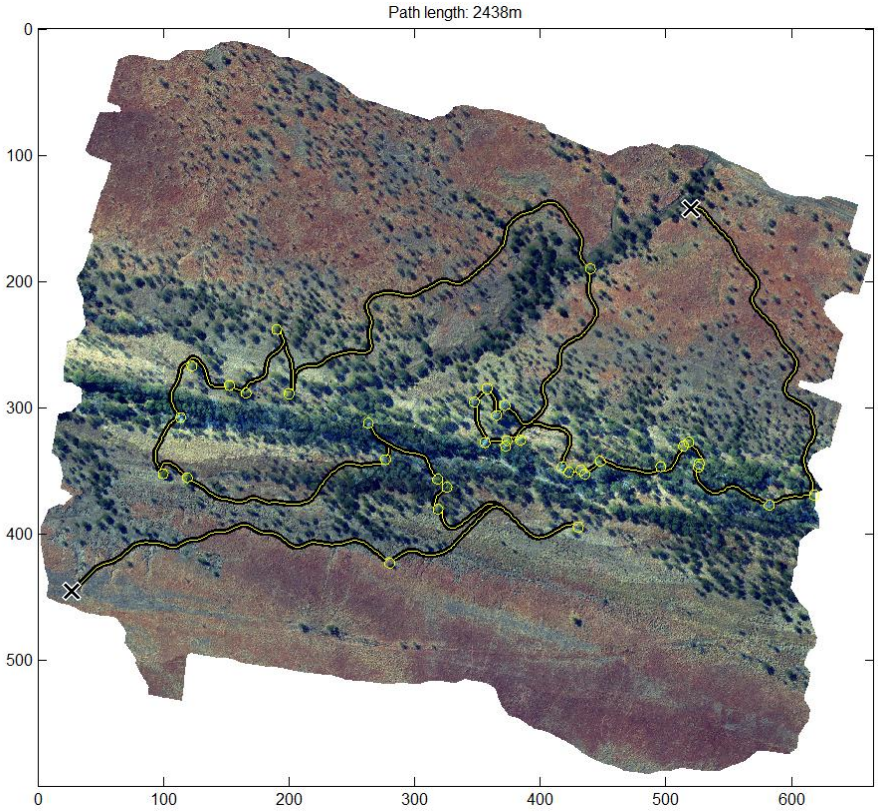

Figure 33: a cost-optimal tour of all the target nodes from start to finish, following non-linear paths based on the constructed cost function. Note that it is built of edges from the above graph representation.

Now, a close inspection of the path in Figure 33 (a closer view is shown in Figure 34 below) explains the converging roads between nodes - the paths are locally optimal with respect to the cost function, visiting nodes in the target set (which is compulsory), but also opportunistically following the other trees that were not directly included in the waypoint list. The paths weaving between trees to avoid collisions, but the high cost of taking a longer path leads to behaviour such as using the perimeter of the dense infestation for quick traversal.

Clearly this path would be better than the straight line plan when a ground robot is used.

Potentially, a simple cost function like this could be tuned to be a good approximation for the traversal goals of using a quad-bike (manned or autonomous), for example.

Figure 34: a close view of the tour path selected in this planning scenario. Left:. Paths weave around trees due to proximity cost. Right: Paths follow the tree line rather than going through the trees or using empty space, avoiding both collisions and wasted data collection opportunities (because empty space is slightly penalised the robot is willing to take a slightly longer path to pass by trees of opportunity along the way).

4 Conclusions

The computer analysis framework presented in this report reflects progress towards the robust and reliable classification of woody weeds in aerial imagery. We also extend the analysis to large scale mapping, and high level analyses of the classified images. The research produced by this project could potentially provide support for management tasks such as early detection, assessing the extent of an infestation through detailed up to date mapping of survey regions, or even planning an efficient route to inspect and treat the detected targets. Illustrative results have been presented for UAV imagery collected over the Carrum and Williams properties in Julia Creek, Queensland Australia.

A mosaic-processing pipeline was developed to render large-scale, geo-referenced mosaics from the UAV data, incorporating a three-dimensional terrain structure model to register the images rather than projecting them to a flat ground. These structural models allowed the imagery to be rendered as map mosaics with consistent spatial area, consistent spatial dimensions and without redundant overlap or doubling counting of the objects classified in the imagery data.

A framework to compress the image textures into features relevant to the UAV aerial data has been demonstrated. The proposed features have been coupled with a popular machine learning approach to obtain estimates of the class of unlabelled objects seen in the aerial imagery. This learning framework has been designed to work with a relatively small number of example image patches that can be easily specified by a human user exploring the image data through a graphical user interface on their computer. In this case, the training/validation data was cross referenced with a limited ground survey of the region conducted by ACFR personnel.

Following the design and training process, efficient software has been created to tractably batch-process multiple image frames. While the classification algorithm is performing well in cross validation with up to 92% accuracy, it can't interpret the scenes in the aerial images at the comprehension level of a human. Humans can

intelligently handle unexpected out-of-training data such as animals, artefacts and gaps in coverage. The primary advantage of an automated vision classifier is therefore not in its flexibility, but in the fast processing of overwhelming data that would be time consuming and tiring for a human. This will both reduce analysis costs, and provide consistent attention to detail over large databases.

This classification procedure has been integrated with the large scale spatial reconstruction pipeline to efficiently produce classification maps that are accurately aligned across the edge of image frames. This has allowed for simple and accurate delineation of the detected tree crowns, enabled the estimation of statistics such as tree counts, and allowed the exploration of efficient traversal routes between targets of interest using computer aided planning. An added advantage of integrating the classification and mosaicking systems is that the total amount of classification is reduced by avoiding the redundant processing of overlapping imagery - although the equally expensive cost of mosaicking overlapping imagery is introduced, we do not need to double up on both.

There are some cases in the data where it is apparent that more discriminative spectral features would be helpful. Because sensor technology is continuously advancing and evolving, an updated UAV payload system in the future could include bands targeted to the vegetation discrimination problem, such as LIDAR sensors to measure height or near-infrared bands that are known to simplify vegetation discrimination and provide additional reliability. It has also been found that gaps occur in the mosaicked imagery despite the design intentions of the flight plans. These are primarily caused by cross-track error induced by wind gusts rolling the UAV platform. The sensor footprint control could be improved by either researching flight control algorithms that consider both the sensor footprint and the UAV dynamic response when following the flight plan, or by investigating an independently gimballed camera for full control over the sensor footprint. There is also scope in the future to consider other forms of vehicle such as an autonomous quad-bike to obtain ground perspectives and directly access the trees for treatment, using the aerial survey information for mission planning.

In this project, our findings suggest that trees can be very reliably distinguished from the ground, and within the detections woody weeds can be effectively distinguished from native species. While not as reliable, in the majority of cases, types of woody weeds can be distinguished from each other. The impact of automating the analysis of UAV data is that, after a small amount of training from a human, the classification system can automatically interpret the data for potentially much larger areas, reducing the dependence on costly manual analysis, and offering affordable, up to date surveying, mapping, detection and planning capabilities in the field.

5 Potential Future Work

5.1 Algorithm Applicability to other Environments

The algorithm pipeline developed and implemented in this project can be applied to different datasets collected at different locations and/or at different times.

If there are no major changes to the environmental settings (similar lighting conditions, same type of vegetation species and background) the algorithm can be applied without retraining.

If the environmental settings have changed significantly, or a new vegetation species is detected, then retraining is required, although no changes are required of the algorithm itself. That is, a selection of images from the new dataset is provided to the expert who chooses the features of interest and the algorithm simply learns the new model. Further extensions of this work are in active learning, where the system itself chooses features to build a classification model and only queries the expert when it wants to improve on the confusion matrix that it self-generates. These techniques are based on statistical information gain principles and remove even further human effort required in trawling through the images.

Thus a wide range of training data (i.e. different environments, different species, etc) creates a broader system implementation profile. As the number of models increase then a manner by which appropriate model selection occurs is required. New approaches to addressing this challenge are in the use of hierarchical classification. This is where a library of different models is learnt from different datasets representing new environments and/or species (as discussed above); and a scene classification algorithm based on logic and rule-based approaches is developed that decides on the best model to use for a particular environment automatically (based on more general environmental characteristics such as the overall colour of a large scene or texture changes).

As observed in this project the classification performance relies heavily on the feature descriptors selected. The multi-scale feature descriptor outperforms the pixel colour based features used in B.NBP.0474. Another area worth investigating is to apply feature learning on the image dataset. Feature learning captures the feature descriptors from the dataset itself instead of requiring a human operator to perform feature selection and calibration. This would improve the classification performance further and at the same time reduce human effort required in calibration across different environments.

5.2 UAV with Hyperspectral Sensors

It is technically possible to mount a multi or hyperspectral sensor on the current UAV platform. There are a number of new sensors on the market that could easily be implemented.

The multispectral sensor typically provides an extra band in the near infrared (NIR) spectrum, thus allowing for the calculation of Normalised Difference Vegetation Index (NDVI) that can be very useful in vegetation segmentation.

The hyperspectral sensor can collect much higher spectral resolution data, in the order of hundreds of channels compared to the three-channel colour images collected currently. The hyperspectral data can be used to generate a spectral shape of different vegetation, and can potentially increase the discriminative power of the classifier. More interestingly would be to fly over vegetation during flowering. This would easily provide discriminative power beyond current approaches but would increase the difficulty of flight operations because of the short temporal window when flowering occurs.

While multi and hyperspectral data can potentially provide better classification results it is also more difficult to collect repeatable and comparable data. This is because the sensor is more sensitive to changes in ambient lighting conditions compared to a three-channel colour camera, and this can potentially result in greater effort required for pre-mission calibration. Furthermore, the pixel resolution of multi and

hyperspectral sensors are orders of magnitude less than those of the three-channel colour camera (for comparable cost) and hence would limit its applicability in farm and agricultural situations for the time being.

It should also be noted that in B.NBP.0474 we demonstrated that one could combine the high pixel power of a three-channel colour camera, obtained from a UAV, with lower resolution hyperspectral data obtained, from high altitude manned aircraft, to allow for classification over large areas. This itself would be interesting to conduct as a follow up study over larger areas and different environments, or it could also be the method applied in a single UAV flight operation. That is, flying at higher altitudes with the hyperspectral camera on a UAV would provide greater flight range capabilities (a larger portion of the farm is mapped), whilst providing very similar classification results to the high resolution three-channel colour camera after the correlations between the two sensors systems are learnt.

5.3 Unmanned Ground Vehicles

Unmanned Ground Vehicles (UGVs) will become a significant tool for the farm of the future. Immediate applications include greater discrimination of vegetation (because of its closer proximity and its greater capability of carrying more sensors); an ability to potentially treat weeds autonomously (which would also feed off the path planning algorithms presented earlier in this document); tracking and potentially even herding cattle; and the monitoring of infrastructure. Two examples are presented in Figures 35 and 36. Both systems have been developed at the ACFR. The Argo vehicle is used in many defence and infrastructure projects, and the Segway system is currently being used in a horticulture project to look at individual tree classification and monitoring of soil conditions, and in a dairy farm project to look at soil and fodder condition as well as the monitoring of cows.

Figure 35: the Argo autonomous vehicle was developed at the ACFR with the aim of building an intelligent multi-purpose platform that could be used in unstructured and undulating terrain. The system is equipped with various laser, radar and thermal sensors, and is capable of long distance and beyond line-of-sight operation.

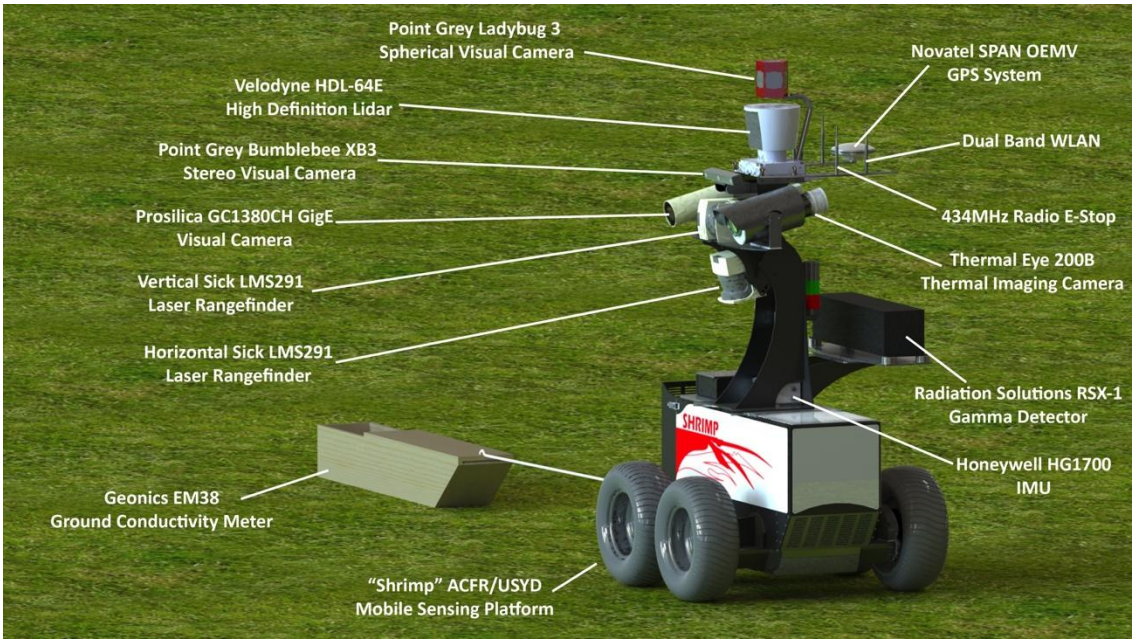

Figure 36: Shrimp is an electric powered UGV that was developed at the ACFR to carry multiple sensors systems and conduct real-time segmentation and classification of the environment. It has a wide sensor capability that allows for the development of simultaneous in-ground and out-of-ground environment models. It is currently being used in a project sponsored by Horticulture Australia Limited for tree and fruit segmentation, and in an internal University of Sydney project to look at monitoring of ground nutrients and cows on a dairy farm.

Acknowledgements

The authors would like to thank Dr. Matthew Johnson-Roberson for his assistance in the software rendering algorithm used in the construction of imagery mosaics and the mosaic geo-tagging. In addition, the UAV field operations of the previous trials would not have been possible without the support of ACFR's aerospace group members, particularly the engineering team of Jeremy Randle, Steve Keep, and Muhammad Esa Attia.

6 Appendices

6.1 UAV Flight Platform Summary

The UAV platform used to acquire data in the proceeding project was equipped with an imaging payload system including a downward pointing camera, and a navigation system for geo-referencing. Specifications of the payload components are provided below in Table A1:

Vision Camera	Hitachi HV-F31	**IMU**	Honeywell HG1900
Sampling Rate	3.75Hz	Sample Rate	600Hz, pre-processed to 100Hz
Field of View	28^o x 22^o	Accel. Noise (1σ)	$0.05m/s^2$
Resolution	1024 x 768 pix	Gyro Noise (1σ)	$0.05^o/s$
Angular Resolution	0.0285^o	Accel. Bias(1σ)	$0.05m/s^2$
Ground Resolution	3.7cm/pix @ 100m, 18.6cm/pix @ 500m	Gyro Bias (1σ)	$0.05^o/s$
Ground Footprint	38x30m @ 100m, 190x150m @ 500m		

GPS Receiver	Novatel OEM5, differentially corrected
Sample Rate	5Hz
Position Error (1σ)	1m
Velocity Error (1σ)	10cm/s

Table A1: Sensor Payload Specifications: The sensor payload consists of an IMU, GPS receiver and downwards-mounted colour monocular camera.

6.2 Julia Creek Flight Trial Patterns

The two sites surveyed in 2009 and 2010 belong to the Carrum (Figure A1) and Williams (Figure A2) properties respectively. Given limited prior knowledge about the regions, UAV flights were planned to cover highly vegetated regions (such as along riverbeds). These flights have been conducted at 100m (small swath, 4cm/pixel resolution) and 500m (large swath, lower resolution). For the classification analysis conducted in this project, the high resolution flights have been used. Specifically, we have focused on one low altitude flight from each site as case studies.

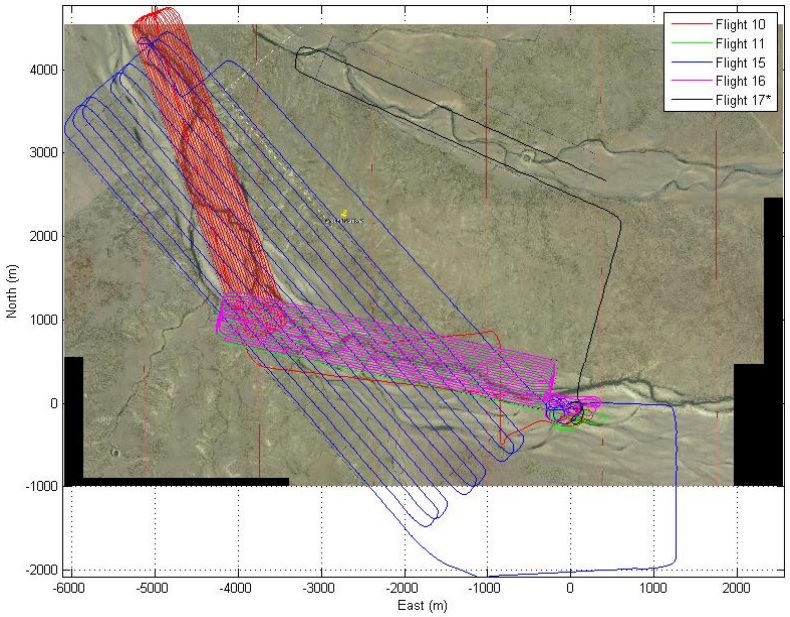

Figure A1: Flight paths for Flights 10, 11, 15, 16 and 17 at the Carrum farm site. Shown underneath the flight paths is low resolution map imagery of the area available from Google Earth.

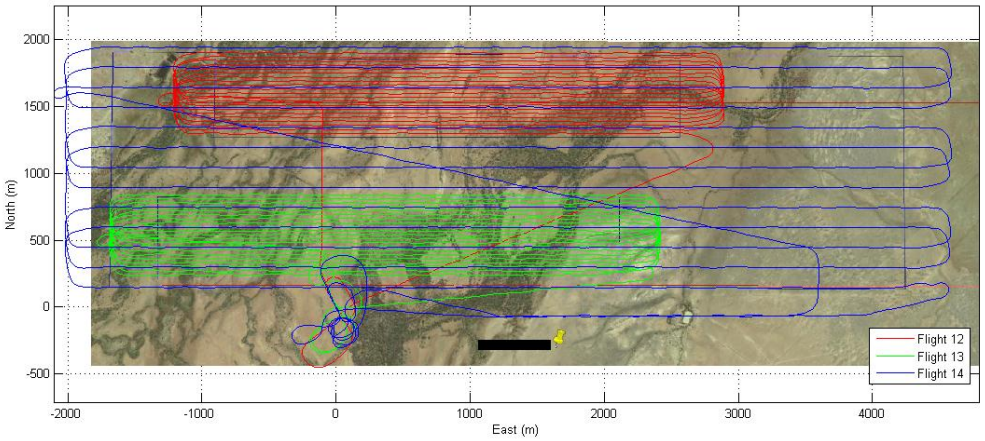

Figure A2: Flight paths for Flights 12, 13 and 14 at the Williams outstation site. Shown underneath the flight paths is low resolution map imagery of the area available from Google Earth.

6.3 In-Field Weed Identification

Manual ground truth collection with a handheld GPS unit has been conducted within the survey regions (concurrent with flying). Surveyors were familiarised with the appearance of the weeds and other tree species (Figure A3), and data points were collected using a handheld GPS unit.

Figure A3 – Ground based photographs of different woody weeds taken by the ground truth survey team: left, Prickly Acacia and right, Parkinsonia.

Trees that were labelled were recorded and cross referenced with the aerial imagery to investigate their appearance from above (which can be surprisingly different to on the ground). An example of these class labels is shown below in Figure A4:

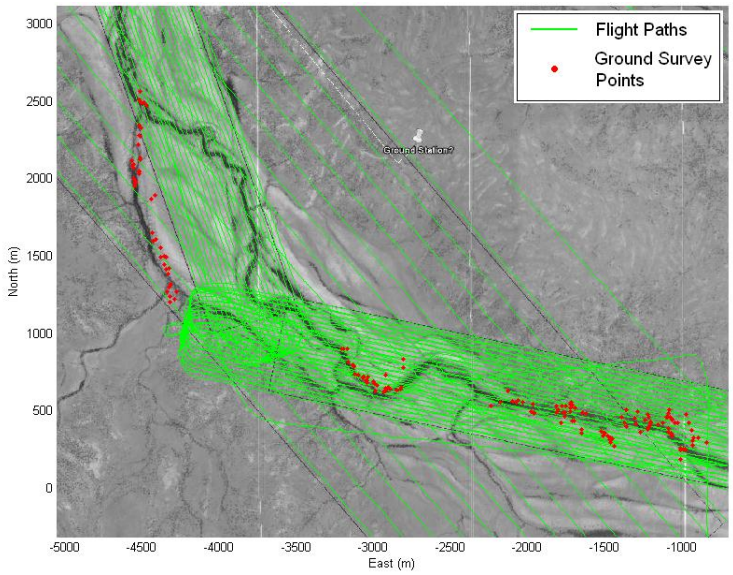

Figure A4 – Ground Survey Data mapped onto the Carrum Farm Site: The green lines indicate the paths of different flights in the Carrum area while the red points indicate the locations of surveyed tree crowns.

7 Bibliography

[Barber 96] C. Barber, D. Dobkin and H. Huhdanpaa, “The Quick- hull algorithm for convex hulls,” ACM Trans. on Mathematical Software vol. 22, no. 4, pp. 469–483, 1996.

[Belhumeur97] P. Belhumeur, P. Hespanha, D. Kriegman, "Eigenfaces vs. isherfaces: recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, Vol 19, Issue 7, pp 711-720

[Bryson10] M. Bryson, A. Reid, F. Ramos, and S. Sukkarieh, “Airborne vision-based mapping and classification of large farmland environments," Journal of Field Robotics (Special Issue on Visual Mapping and Navigation Outdoors), vol. 27, no. 5, pp. 632-655, 2010.

[Casady05] G. Casady, R. Hanley, and S. Seelan, “Detection of leafy spurge (Euphorbia esula) using multidate high-resolution satellite imagery," Weed technology, vol. 19, no. 2, pp. 462-467, 2005.

[Everingham2011] M. Everingham, L. Van Gool, C. Williams, J. Winn and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2011 (VOC2011) Results", http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html

[Freund97] Y. Freund and R. Schapire, “A decision-theoretic generalization of online learning and an application to boosting," Journal of Computer System Sciences, vol. 55, no. 1, pp. 119-139, 1997.

[Freund99] Y. Freund and R. Schapire, “A short introduction to boosting," Journal ofJapanese Society for Artificial Intelligence, vol. 14, no. 5, pp. 771-780, 1999.

[Furukawa 10] Y. Furukawa and J. Ponce, “Accurate, Dense, and Robust Multi-View Stereopsis,” IEEE Trans. on Pattern Analysis and Ma- chine Intelligence, vol. 32, no. 8, pp. 1362–1376, 2010.

[Hassin00] Hassin, R.; Rubinstein, S. (2000), "Better approximations for max TSP", Information Processing Letters 75 (4): 181–186

[Heeger95] D. Heeger and J. Bergen, “Pyramid-based texture analysis/synthesis," in Proceedings of the 22nd International Conference on Computer Graphics and Interactive Techniques, 1995.

[Johnson-Roberson 10] M. Johnson-Roberson, O. Pizarro, S. Williams and I. Mahon, “Generation and Visualization of Large-scale Three-dimensional Reconstructions from Underwater Robotic Surveys,” Journal of Field Robotics, vol. 27, no. 1, pp. 21–51, 2010.

[Joliffe86] Jolliffe, I. T. (1986). Principal Component Analysis. Springer-Verlag. pp. 487

[Klinken07] R. Klinken, D. Shepherd, R. Parr, T. Robinson, and L. Anderson, “Mapping mesquite (prosopis) distribution and density using visual aerial surveys", Rangeland Ecology Management, vol. 60, pp. 408-416, 2007.

[Lawes08] R. Lawes and J. Wallace, “Monitoring an invasive perennial at the landscape scale with remote sensing," Ecological Management and Restoration, vol. 9, no. 1, pp. 53-58, 2008.

[Leung01] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision, 43(1):29-44, 2001.

[Meyer94] Fernand Meyer, "Topographic distance and watershed lines," Signal Processing , Vol. 38, July 1994, pp. 113-125.

[Reid11] A. Reid, F. Ramos, and S. Sukkarieh, “Multi-class classification of vegetation in natural environments using an unmanned aerial system," in Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2953-2959, 2011.

[Tang03] L. Tang, L. Tian, and B. Steward, “Classification of broadleaf and grass weeds using Gabor wavelets and an artificial neural network," Transactions on ASAE, vol. 46, no. 4, pp. 1247-1254, 2003.

[Tsalakanidou03] F Tsalakanidou, D Tzovaras, M.G Strintzis, Use of depth and colour eigenfaces for face recognition, Pattern Recognition Letters, Volume 24, Issues 9–10, 2003, Pp 1427-1435

[Wulder00] Wulder et al., Local Maximum Filtering for the Extraction of Tree Locations and Basal Area from High Spatial Resolution Imagery.

[Yu06] Q. Yu, P. Gong, N. Clinton, G. Biging, M. Kelly, and D. Schirokauer, “Object based detailed vegetation classification with airbourne high spatial resolution remote sensing imagery" Photogrammetric Engineering and Remote Sensing, vol. 72, no. 7, pp. 799-811, 2006.