

DEVELOPING SPECIES SPECIFIC VEGETATION MAPS USING MULTI-SPECTRAL HYPERSPATIAL IMAGERY FROM UNMANNED AERIAL VEHICLES

Christoph Strecha ^{a,c}, Andrew Fletcher ^b, Alex Lechner ^b, Peter Erskine ^b and Pascal Fuà ^a

^a EPFL-I&C-CVLAB, 1015 Lausanne, SWITZERLAND - (christoph.strecha, pascal.fua)@epfl.ch

^b University of Queensland, Centre for Mined Land Rehabilitation, Level 5 Building 47A, Brisbane, Queensland, AUSTRALIA 4072 - (a.fletcher, p.erskine, a.lechner)@cmlr.uq.edu.au

^c Pix4D Sàrl, Chemin des Clos 177b, 1024 Ecublens, SWITZERLAND - christoph.strecha@pix4d.com

WG III/4: Complex Scene Analysis and 3D Reconstruction

KEY WORDS: Integration of IR and Visible sensors, condition, shrub swamps, UAV

ABSTRACT:

In remote, rugged or sensitive environments ground based mapping for condition assessment of species is both time consuming and potentially destructive. The application of photogrammetric methods to generate multispectral imagery and surface models based on UAV imagery at appropriate temporal and spatial resolutions is described. This paper describes a novel method to combine processing of NIR and visible image sets to produce multiband orthoimages and DEM models from UAV imagery with traditional image location and orientation uncertainties. This work extends the capabilities of recently developed commercial software (Pix4UAV from Pix4D) to show that image sets of different modalities (visible and NIR) can be automatically combined to generate a 4 band orthoimage. Reconstruction initially uses all imagery sets (NIR and visible) to ensure all images are in the same reference frame such that a 4-band orthoimage can be created. We analyse the accuracy of this automatic process by using ground control points and an evaluation on the matching performance between images of different modalities is shown. By combining sub-decimetre multispectral imagery with high spatial resolution surface models and ground based observation it is possible to generate detailed maps of vegetation assemblages at the species level. Potential uses with other conservation monitoring are discussed.

1. INTRODUCTION

1.1 General Instructions

Protection of ecological communities from anthropogenic activities such as mining has increased in recent years in developed countries such as Australia as a result of public awareness, media scrutiny and increasing competition between different land uses. As a result it is necessary for companies seeking license to develop and continue operations to provide highly accurate and transparent evidence that negative impacts to natural environments are not significant. The footprint of activities such as mining is generally clearly defined; however, the metrics for assessing impact are less clear. In this study, shrub swamps present above longwall underground mining operations provided a case study of the interaction of spatial, spectral and temporal resolution in effectively monitoring potential mining impacts. Shrub swamps are characterised by a complex mosaic of dense shrubby vegetation that is difficult to access and/or assess by ground based methods (Elzinga C., Salzer D., Willoughby J. and Gibbs J., 2001). These communities are hydro-geologically restricted in distribution and follow drainage lines. Shrub swamp communities in south eastern Australia are variously listed as endangered or vulnerable ecological communities by various government regulators eg. (DEH, 2005). As a result, companies extracting resources from beneath listed communities must demonstrate that all possible steps to protect these shrub swamp communities have been taken.

The application of high resolution satellite imagery to the assessment of shrub swamp communities has shown a limited capacity to detect impacts beyond formation of significant areas of bare ground (Jenkins and Frazier, 2010). This level of impact is unlikely to be appropriate to the effective conservation of these ecological communities which requires early detection of impact. The limitation of commercially available remotely sensed products to effectively assess shrub swamps is a function of the shape, size and distribution of these communities. Shrub swamps are commonly small in size and linear shaped with a high perimeter area ratio and thus a high proportion of the total swamp area is in close proximity to community boundaries. This introduces significant measurement uncertainty as small, linear features have higher classification errors due to the higher probability of pixels lying on the boundary between classes (A. Lechner, A. Stein, S. Jones and J. Ferwerda, 2009) or boundaries between classes may be indeterminate (Burrough 1996). Application of traditional spatial and temporal resolution imagery to these communities will necessarily introduce significant errors to the analysis. Remote sensing of shrub swamp communities at meaningful scale for identification and assessment therefore requires imagery that contains spatial resolutions appropriate to the features of interest.

The application of Unmanned Aerial Vehicles (UAV) was therefore examined as a means of achieving the high spatial resolution required for vegetation assemblage mapping and condition analysis given the small total area (up to 20ha) represented by individual swamp communities. This technique allows routine swamp scale image capture at

ground sample resolutions less than 10cm. Image location and orientation uncertainties resulting from GPS receiver weight restrictions, limited sensor integration and inertial measurement sensitivity inherent to small UAVs limit the quality of image metadata. Restricted sensor payload and low altitude wide angle imagery contribute Bayer interpolation and geometric errors to colour imagery respectively. Technical quality of images is therefore lower than commercial off the shelf imagery and often requires substantial effort to develop quantifiable data (A. Laliberte and A. Rango, 2011).

High near infrared (NIR) reflectance is a feature of green vegetation and thus is a commonly collected spectrum for the assessment of vegetation health. The use of NIR and red or green spectral ratios is often used in the determination of vegetation health of which normalised difference vegetation index is a commonly applied index. At sub-decimetre spatial resolution these indices will allow individual plant canopies within a shrub swamp to be identified and assessed for condition. Currently it is only possible to collect red and NIR using separate cameras. In this study we present a novel method to develop multispectral imagery of shrub swamps from small UAV image sets. This method involves the merging of multiple image sets collected in both RGB and NIR spectra. This method achieves both sub-decimetre ground sampling resolution and multispectral image generation.

2. DATA AQUISITION

2.1 Hardware

For each project, two sets of imagery were captured using a 2m wingspan, 3.5kg battery powered glider (Kahu, Skycam UAV, New Zealand) carrying a Sony NEX5 14Mp camera with 16mm fixed lens providing the equivalent of 24mm full frame image. One camera was modified to collect panchromatic full spectrum imagery by removal of IR and Bayer filters. NIR wavelengths were isolated by using a filter (Hoya R72).

2.2 Acquisition Sessions

Imagery was collected over six swamps on the Woronora and Newnes Plateaus. Imagery was collected between 10am and 3pm to minimise shadow impact on the imagery. A total of twelve sets of imagery were collected in both RGB and NIR spectra. Image sets from 150-900 single photos were collected in each flight depending on shrub swamp size and wind conditions. Images were collected at 2.2sec intervals in flight lines with >80% forward overlap. Imagery was collected at approximately 300ft above ground level and camera location and attitude at time of image capture were recorded on board throughout the flight. Imagery was collected in RAW (.ARW file) format and converted to JPG format prior to photogrammetric processing using Pix4D (Pix4D, Switzerland).

3. AUTOMATED DATA PROCESSING

3.1 Software Description

The software we use for our experiments was primarily designed for imagery in the visible spectrum. It can process up to 10000 images, is fully automated with a high accuracy (O. Küng, C. Strecha, A. Beyeler, J-C. Zufferey, D. Floreano, P. Fua, F. Gervais, 2011). A geo-referenced ortho mosaic and DSM can be obtained in principle without the need for ground control points. However, as shown here, more accurate geo-referencing requires Ground Control Points.

The software performs the following steps:

- I. The software searches for matching points by analyzing all images. Most well known in computer vision is SIFT (D.G.Lowe) feature matching. Studies on the performance of such feature descriptors are given in K. Mikolajczyk and C. Schmid (2002). We use here an improved version of the binary descriptors proposed in Strecha et al. (2012), which are very powerful to match keypoints quickly and accurately.
- II. Those matching points as well as approximate values of the image position and orientation provided by the UAV autopilot are used in a bundle block adjustment, e.g. (B. Triggs and P. McLauchlan and R. Hartley and A. Fitzgibbon, 2000) and (R. Hartley and A. Zisserman, 2000), to reconstruct the exact position and orientation of the camera for every acquired image (Tang, L. and Heipke, C., 1996).
- III. Based on this reconstruction the matching points are verified and their 3D coordinates calculated. The geo-reference system is WGS84, based on GPS measurements from the UAV autopilot during the flight.
- IV. Those 3D points are interpolated to form a triangulated irregular network in order to obtain a DEM. At this stage, construction of a dense 3D model, e.g. (D. Scharstein and R. Szeliski, 2002), (C. Strecha and T. Tuytelaars and L. Van Gool, 2003), (C. Strecha and W. von Hansen and L. Van Gool and P. Fua and U. Thoennessen, 2008), (Hirschmüller, 2008), increases the spatial resolution of the triangle structure.
- V. This DEM is used to project every image pixel and to calculate the geo-referenced orthomosaic (also called true orthophoto) (C. Strecha and L. Van Gool and P. Fua, 2008)

3.2 UAV Specific Problems

By applying the above steps to the datasets in the visible and NIR spectrum independently one would retrieve two ortho images, which, due to the large errors on the GPS positions of the individual UAV images are not well aligned. Thus, even though this can be done without manual intervention, a simple combination of both is insufficient to obtain precisely aligned multispectral information for the objects in the scene. The results of processing the datasets individually can be seen in Figure 1. This figure shows clearly the disadvantage of UAV based GPS measurements, which have been earlier reported in (H. Eisenbeiss and W. Stempfhuber, 2009).

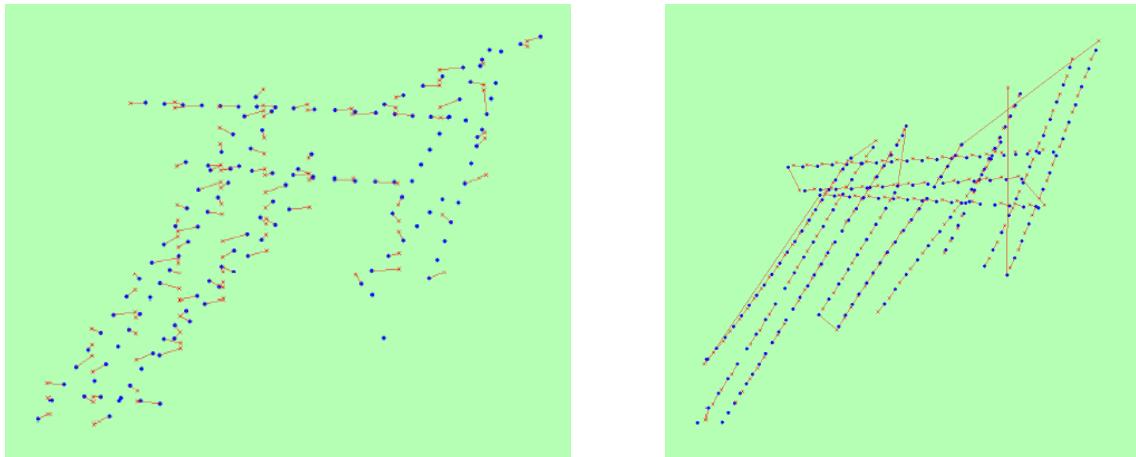


Figure 1: Image positions for the flights with the visible (left) and NIR sensor (right) for the dataset in Figure 3 (right). In blue one can see the refined camera centers as computed from the software described in section 3.1. Red lines connect those camera centres to the original GPS positions. These datasets show that the GPS measurements of the UAV are of low accuracy, contain missing GPS positions (left figure) and could potentially be mistagged or shifted (right figure)

3.3 Combined Processing of Visible and NIR Imagery

The solution to the problem consists of treating both datasets at once and applying the above steps to the compilation. All images will then be precisely aligned and the multi spectral information is spatially consistent. Since this requires matching keypoints with different modalities this is not obvious. This problem has been studied for the panoramic stitching of visible and NIR imagery where the correspondences are constrained by a single homography (D. Firsmanich, M. Brown and S. Susstrunk, 2011). Here we develop results on a real 3D case in which the images are related by general 3D motion.

We performed three experiments for which we have NIR and visible datasets available from the same area and from approximately the same time. All have been taken by a UAV. In Figure 3 we show the number of matches that could have been established between images from different modalities and between images of the same modality. These figures can be interpreted as follows.

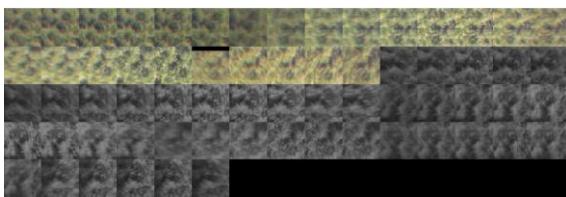


Figure 2: Feature track that corresponds to a single 3D point. Show are the image patches for which the keypoints could be automatically matched and verified by their geometric consistency of the overall reconstruction. One can see image patches from the images in the visible spectrum and NIR image patches.

One output of the automatic reconstruction pipeline are the 3D points that are used to bundle adjust the overall

reconstruction (B. Triggs and P. McLauchlan and R. Hartley and A. Fitzgibbon, 2000). Each 3D point corresponds to a feature track between images. Such a feature track can be seen in Figure 2.

A reconstruction does contain many feature tracks of different sizes. The size of the feature track depends on the ability of the keypoint detector to detect structure with high repeatability and of the descriptor to match these keypoints between images. Figure 4 evaluates this for the case of multi-modal image sets. It shows the cumulated number of matches between images as a function of the track size.

These figures have been obtained by taking all feature tracks and counting the number of matches within these tracks that correspond to NIR-NIR image pairs, NIR-visible and image pairs in the visible spectrum.

From this figure we can see that one can find sufficient matches between NIR and visible images such that both modalities can be automatically combined. Of note also is the difference between the scenes. Whereas in the first two scenes (corresponding to the left and middle part in Figure 4 and Figure 6) many matches can be found between the modalities, however we find far fewer in the swamp dataset (right in Figure 4 and 5), that contains vegetation only. This is consistent with the divergent reflectance properties of vegetation between visible and NIR spectra and the general paucity of reliable automatically detectable features in vegetation. Improvements to this situation may be achieved by concurrent image collection in both modalities and by increasing image overlap.

Figure 3 shows the results of combining the two datasets (from Figure 1) into a single reconstruction. Again as in Figure 1 by showing the images base and refined camera centers together with the input GPS positions that have been measured by the UAV.

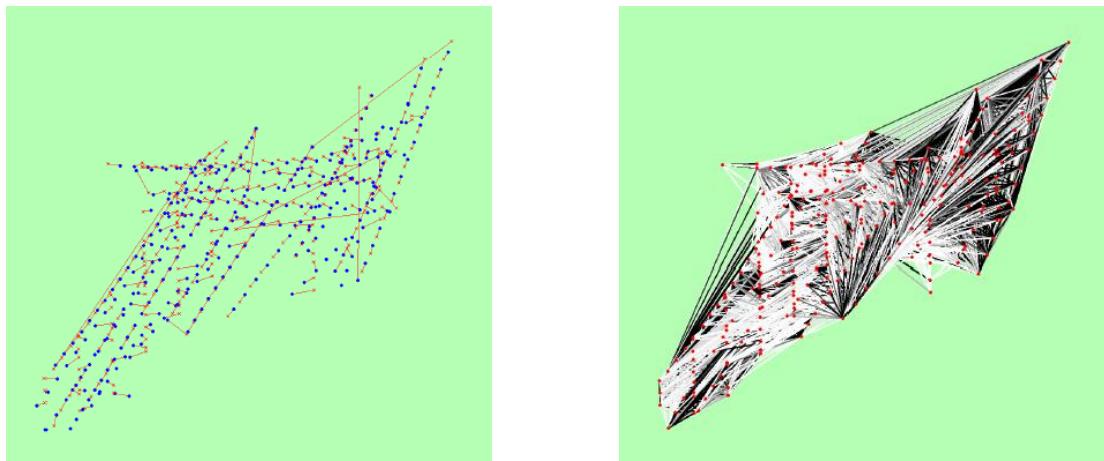


Figure 3: Computed camera centers for the combined processing of visible and NIR images for the example in Figure 1 (left Figure). The reprocessed camera centers (blue dots) are connected to the corresponding GPS measurements by red lines.

The Figure on the right shows the number of matches between all images color coded (black is indicating many matches between image pairs).

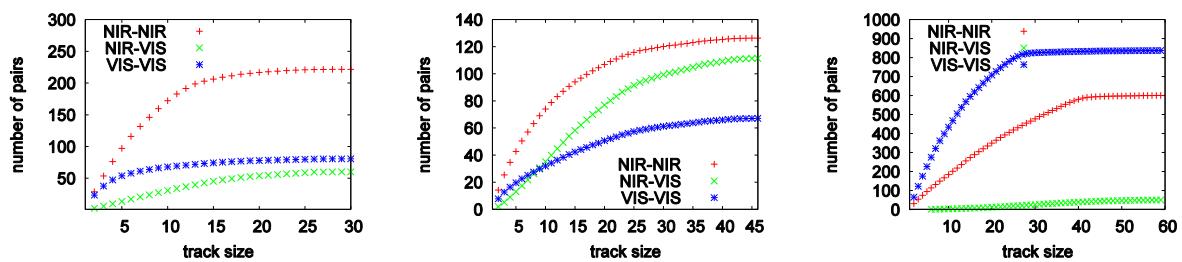


Figure 4. Matching statistics for NIR-NIR, NIR-visible and visible-visible image pairs for two different datasets.
 The figures contain cumulative number of matches as a function of the track size.

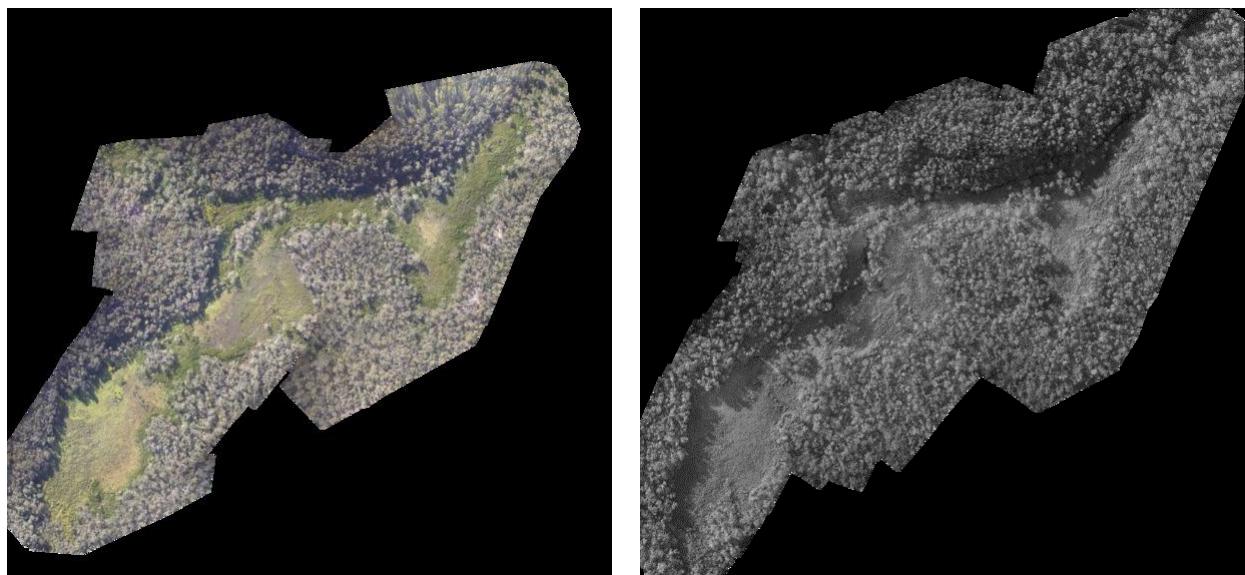


Figure 5: Ortho images for the visible (left) and the NIR spectrum (right) that have been obtained by combining both datasets into a single reconstruction.

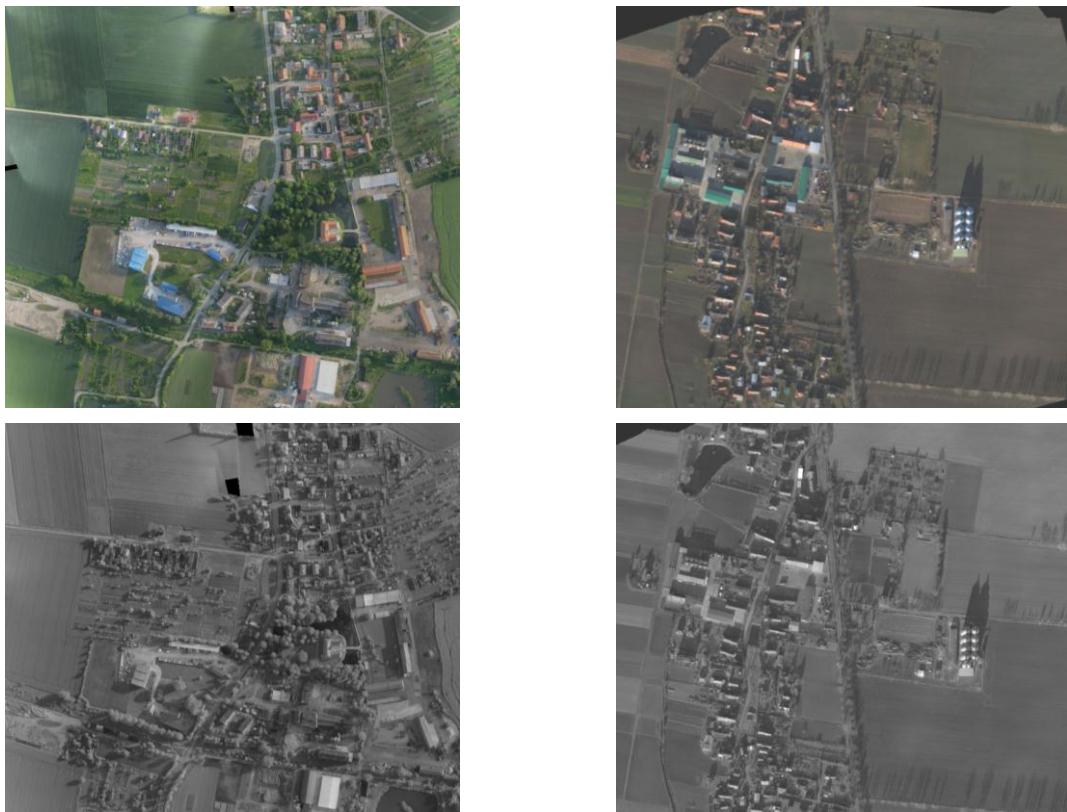


Figure 6: Ortho images for the visible (top) and the NIR spectrum (bottom) that have been obtained by combining both datasets into a single reconstruction. Images obtained by Pteryx UAV, courtesy of Trigger Composites.

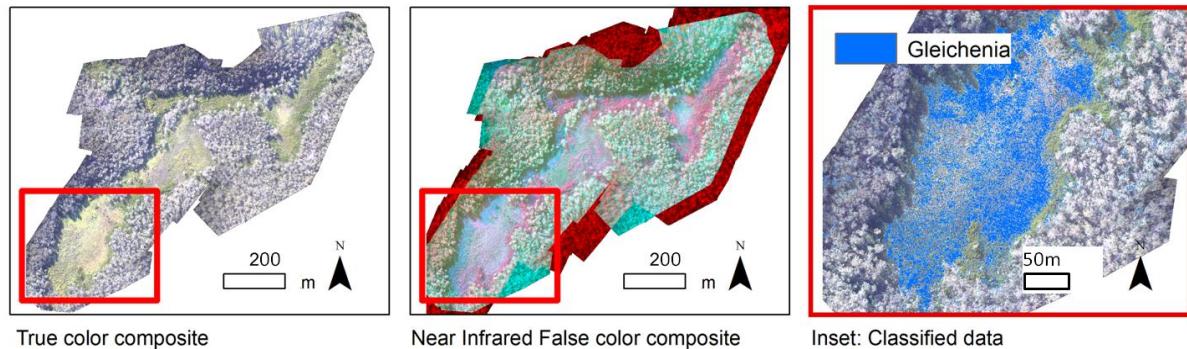


Figure 7. (left) True color orthomosaic; (middle) near infrared false colour composite (Band 1: Green, Band 2: Blue and Band 3: Near Infrared). (right) Subset of total swamp with classified *Gleichenia microphylla* overlaid on true color composite.

4. VEGETATION MAPPING AT NEAR PLANT LEVEL

We used a supervised maximum likelihood classifier to identify *Gleichenia dicarpa* in the ERDAS image software.

For this paper we demonstrated that certain plant species with distinct spectral signatures may be extracted using pixel based classifiers. *Gleichenia* is a small fern often with a unusually bright green colour that makes it easily distinguishable with remote sensing imagery at appropriate spatial resolution. The maximum likelihood classifier was

trained using a range of classes describing vegetation other than *Gleichenia dicarpa*. These other classes were then combined to produce a binary map of *Gleichenia* versus other vegetation classes (Figure 7).

The classifier performed well in extracting the vegetation species, however, there were numerous errors of commision. Further research in using object basited image analyses (OBIA) methods are likely to yield better results as they are suited to very high spatial resolution imagery (Blaschke 2009). However, there is great potential in using pixel based classification methods to classify plants when they are flowering. Particularly in swamps, the flowers of certain plant species exhibit high contrast with the surrounding vegetation due to bright, pink, yellow or orange pigmentation. Only through high spatial resolution imagery can such flowers be found as a single pixel. In lower resolution remote sensing imagery these components of plants only exist as mixels.

5. DISCUSSION

It is difficult to detect and map change in dense, spatially heterogeneous and bio-diverse vegetation communities such as shrub swamps by using ground observation or traditional remote sensing products. Accurate measures of vegetation structure and condition require spatial resolutions that exceed the granularity of individual plants and/or plant assemblages by at least eleven fold (Lechner et al 2009). This paper describes a novel methodology to converge numerous sets of imagery to develop raster cohesive datasets capable to assess complex vegetation communities at spatial resolutions appropriate to the features of interest.

The capacity of UAV platforms to provide rapid revisit and highly redundant imagery at spatial resolutions appropriate to fine scale vegetation assemblages can be readily demonstrated. However these capabilities are tempered by uncertainty in location and orientation of image captures introduced by light weight GPS receivers, low resolution attitude information and minimal integration of sensor and metadata sources. The process described in this paper addresses the limitations of small UAV imagery while also allowing imagery from multiple captures and modalities to be integrated despite the inherent uncertainties.

Conservation decisions are required to assess impacts of anthropogenic activities at site and landscape scales. Where particular communities or assemblages/habitats are susceptible to key impacts of activities the requirement for confidence in the absence of impact or in the assessment of severity and extent of impact is paramount. This paper demonstrates a novel method to achieve these criteria with very difficult photogrammetric target material that is also difficult to assess by other methods.

6. BIBLIOGRAPHY

A. Laliberte and A. Rango Image Processing and Classification Procedures for Analysis of Sub-decimeter Imagery Acquired with an Unmanned Aircraft over Arid

Rangelands [Journal]. - [s.l.] : Giscience and Remote Sensing, 2011. - Vols. 48(1) p. 4-23.

A. Lechner, A. Stein, S. Jones and J. Ferwerda Remote Sensing of Small and Linear Features: Quantifying the Effects of Patch Size and Length, Grid Position and Detectability on Land Cover [Conference] // Remote Sensing of Environment. - 2009. - Vols. 113(10) p. 2194-2204.

B. Triggs and P. McLauchlan and R. Hartley and A. Fitzgibbon Bundle Adjustment -- a Modern Synthesis [Book]. - [s.l.] : Vision Algorithms: Theory and Practice, 2000.

Blaschke T. Object based image analysis for remote sensing [Journal]. - [s.l.] : ISPRS Journal of Photogrammetry and Remote Sensing In Press, Corrected Proof., 2009.

C. Strecha and L. Van Gool and P. Fua A Generative Model for True Orthorectification [Journal]. - [s.l.] : Proceedings of the International Society for Photogrammetry and Remote Sensing (ISPRS), 2008.

C. Strecha and T. Tuytelaars and L. Van Gool Dense Matching of Multiple Wide-baseline Views [Journal]. - [s.l.] : Proceedings of the International Conference on Computer Vision (ICCV), 2003.

C. Strecha and W. von Hansen and L. Van Gool and P. Fua and U. Thoennessen On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery [Journal]. - [s.l.] : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008.

C. Strecha, A. M. Bronstein, M. M. Bronstein and P. Fua LDAHash: Improved Matching with Smaller Descriptors [Journal]. - [s.l.] : IEEE Transactions On Pattern Analysis And Machine Intelligence. - Vol. 34, 2012

D. Firmenich, M. Brown and S. Susstrunk Multispectral Interest Points for RGB-NIR Image Registration [Conference]. - [s.l.] : International Conference on Image Processing, 2011.

D. Scharstein and R. Szeliski A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms [Journal]. - [s.l.] : IJCV, 2002. - Vol. 47.

D.G.Lowe Distinctive Image Features from Scale-Invariant Keypoints [Journal]. - [s.l.] : IJCV. - Vol. 2.

DEH Department of Environment and Heritage, Data Sheet, Temperate Highland Peat Swamps on Sandstone. [Book]. - 2005.

Elzinga C., Salzer D., Willoughby J. and Gibbs J. Monitoring Plant and Animal Populations [Book]. - [s.l.] : Blackwell Science, Massachusetts, 2001. - Vol. p. 229.

Hirschmüller H. Stereo Processing by Semiglobal Matching and Mutual Information [Journal] // TPAMI. - 2008. - pp. 328-341.

K. Mikolajczyk and C. Schmid An Affine Invariant Interest Point Detector [Book]. - [s.l.] : ECCV, 2002.

O. Küng, C. Strecha, A. Beyeler, J-C. Zufferey, D. Floreano, P. Fua, F. Gervais The Accuracy of Automatic Photogrammetric Techniques on Ultra-light UAV Imagery [Conference] // UAV-g. - Zürich : [s.n.], 2011.

R. Hartley and A. Zisserman Multiple View Geometry in Computer Vision [Book]. - 2000.

Tang, L. and Heipke, C. Automatic relative orientation of aerial images [Journal]. - [s.l.] : Photogrammetry and Remote Sensing, 1996. - Vol. 62.