Vision-based Shadow-aided Tree Crown Detection and Classification Algorithm using Imagery from an Unmanned Airborne Vehicle

C. Hung [a],*, M. Bryson [a], S. Sukkarieh [a]

[a] Australian Center for Field Robotics, The University of Sydney, 2006, NSW Australia
(c.hung, m.bryson, s.sukkarieh)@acfr.usyd.edu.au

***Abstract*— We propose a novel algorithm for tree crown detection and classification in aerial images. The algorithm utilises statistical learning and computer vision techniques to identify various types of woody weeds in a natural landscape with aid of their shadow. In remote sensing context, colour spectrum and texture features are commonly used as cues for classification, the shape property of the target is not explored as often due to the image resolution constrain; however due to our relative lower altitude flights the shape of the target and the corresponding shadow could be resolved and is used to provide extra shape feature for detection. The proposed algorithm is divided into detection and classification stages, the purpose of the detection stage is to identify the interest points correspond to high likelihood of tree crown existence, and therefore reduces the search space of the classification stage from the entire image data set to significantly smaller region of interests. The detection stage is further divided into two stages, the first stage segments the image using colour and texture features, the second stage utilises template matching using shape information related to projected shadow of the woody weed, relying on information about the time of day, sun angle and UAV pose information collected by the onboard navigation system. The classification stage uses supervise learning training on the features collection from the region of interests. We present experimental results of the approach using colour vision data collected from a UAV during operations in the Julia Creek in July 2009 and August 2010.**

***Keywords:* —Aerial Imaging, Computer Vision, Unmanned Aerial Vehicle, Weed Detection, Weed Recognition**

I. INTRODUCTION

This paper concentrates on remote sensing object recognition in the context of aerial surveying using unmanned aerial vehicle (UAV). Various methods have been proposed for detection and recognition of objects from aerial images (Thompson and Mundy, 1987), (Brooks et. al, 1981), however these techniques mostly deal with well defined man-made objects whose visual signature could be accurately modeled. Unlike man-made objects, natural objects, such as trees, are less uniform and difficult to generalise with geometric models. A related problem exists in the field of forest research, where various tree counting algorithms have been proposed,

A few tree detection algorithms have been proposed in the forest counting research field over the past few years to detect natural vegetation from outdoor unstructured aerial images. The commonly used detection algorithms are: the valley finding algorithm (Gougeon, 1995), the region growing method (Erikson, 2003) and the template matching method (Pollock, 1996), (Larsen, 1997), (Olofsson, 2006), and recently the approach with Object Based Image Analysis (OBIA) for tree crown delineation has been explored (Bunting and Lucas, 2006). The paper uses template matching approach because this approach uses both the low level vision feature of colour and texture and also the intermediate level feature of shape. We also incorporate the shadow direction as and extra context cue.

This work is supported in part by Meat and Livestock Australia (MLA) under project code B.NBP.0474, ``UAV Surveillance Systems for the Management of Woody Weeds". This work is supported by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

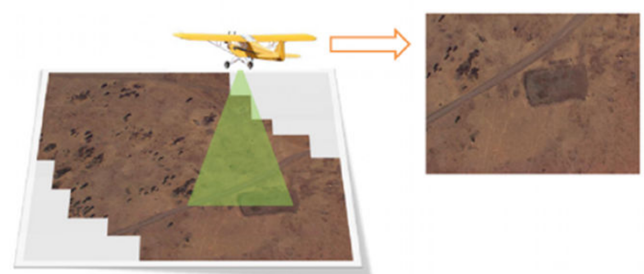

Figure 1. Scaled down J3 Cub: This platform is used to collect aerial image data over Julia Creek, Queensland, Australia. The mission area is flat with trees distributed sparsely, the data set is ideal to test the object detection algorithm.

In this paper we treat the shadow as additional information, similar approaches have been used in aerial surveying (Larsen, 1997) and robotics (Agrawal, 2010). An image frame is a two-dimensional projection of the real world from the camera point of view; this loss of one dimension is one of the limiting factors in computer vision using monocular cameras. There are techniques from computer vision developed to retrieve the lost dimension using shading and shadow. The former is used in photometric stereo (Brooks, 1989) and the latter is applied as a projection of an object from the point of view of the light source.

The first stage of the proposed algorithm is image segmentation. Patterns such as colour and texture which are difficult to quantify directly are learned statistically from the data set using Support Vector Machines (SVM) (Vapnik, 2000), the original image is then divided into several meaningful labelled regions. In the second stage a predictive template is generated using an object appearance model based on the navigation solution and a solar position model. The simulated appearance of the target includes the target with shading and shadow, and is used as a shape feature. The relative position of the target object and its shadow is treated as the context information and can be used as supporting evidence of detections. The third stage performs tree crown species classification on the detection points.

We evaluated the detection algorithm with aerial survey data collected from northwest Queensland, Australia during July 2009 and August 2010 on an ecological monitoring mission. We applied the algorithm to recognise certain types of vegetation.

The contribution of this work is an object recognition algorithm that uses the predictive template constructed with the navigation solution from the robotic platform, the sunlight direction obtained using time information and the prior knowledge of target geometry

II. Algorithm

The object recognition algorithm is divided into three steps: firstly an image segmentation step that takes the colour and texture information into account, secondly a model-object matching step that takes the shape, scale and context information into account and finally a classification step which utilises the features around the detection point to differentiate tree crown species. The three stage approach breaks down the otherwise difficult to solve vision problem into manageable sub-problems. This also allows each distinctive module to be evaluated separately, therefore future improvements can be made independently. This also allows us to generalise the algorithm for different applications with similar structure by re-learning the statistical models and using other prior knowledge.

A. Image Segmentation

The image segmentation stage divides the original image into three different classes based on the colour and texture features. The aim is to change the representation of images from arbitrary colours into meaningful labels that can be analysed by later stages. The three classes are the target object class, the shadow class and the background class.

The first step of image segmentation is to extract the colour and texture information. The images are collected in Red Green Blue (RGB) colour space, they are then transformed into a Hue, Saturation and Value (HSV) colour space to reduce the sensitivity towards change in light intensity. To extract texture information, the MPEG-7 texture descriptors are used (Manjunath1996).

After feature extraction the colour and texture features are grouped into one single feature vector consisting of three colour channels and thirty texture channels. Each feature vector is then assigned with a label representing its class. The aim is to segment the original image into three different classes, object, shadow and background. SVM is used as the classifier to predict the labels of the feature vectors.

.

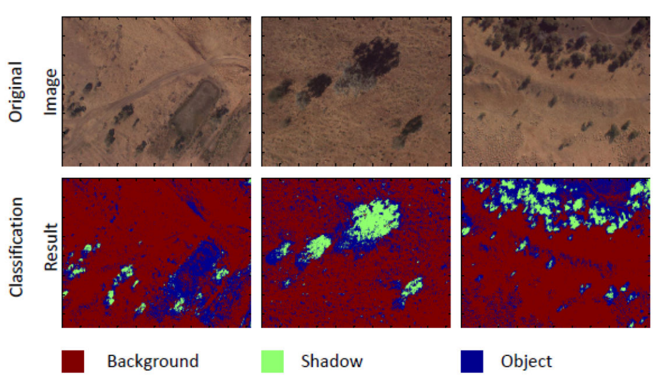

Figure 2: Image Segmentation: The classifier convert the original colour images into images with meaningful class labels.

B. Object Detection Using Predictive Template

Object detection algorithms based purely on statistical information have their performance limited by the quality of the training data. Additionally, prior knowledge such as shape and position of shadow based on the input light direction is too subtle to be incorporated into the statistical learning process. This necessitates separation of image segmentation based on statistical learning and object appearance model based on prior knowledge.

The algorithm used to generate a target object appearance model is discussed in this section. The platform navigation data and the target object outline are used as the prior knowledge in this algorithm; geometric model is constructed based on the target object outline, platform position and time are used to estimate the direction of the incident sun light and then used to predict the direction and outline of the shadow. At the end the platform pose is used to estimate the appearance of the object and shadow in each observed image frame. The object appearance model encapsulates both the object shape and context information.

In this algorithm simple geometric models are used to approximate the outline of the target objects. In contrast to industrial applications where more complicated CAD models are used, it is not possible to generalise and model the exact shape of a natural object, therefore a simple geometric shape or a combination of geometric shapes can be used as a good approximation. In this paper the algorithm an ellipsoid is used to approximate the shape.

In addition to the object model which provides the shape prior information, the shading and shadow cast by the object can provide extra shape information. More importantly, the shading and shadow orientation actually provide context information. The orientation of shading and shadow can be predicted using a solar model and knowledge of the vehicle pose. Any potential object detections with the wrong shading and shadow orientation can thus be rejected.

A sun path model is used to estimate the position of the sun in the sky at defined times and locations where images are collected. The model returns a light vector which represents the orientation of the incident light from the sun. The images time stamps are used in the sun path model to predict the exact position of the sun in the sky; combining this with the platform position allows prediction of the direction of the shadow. The platform pose is used to estimate the camera pose. Combined with the solar model we are able to predict the shadow position with respect to the target object in each image frame.

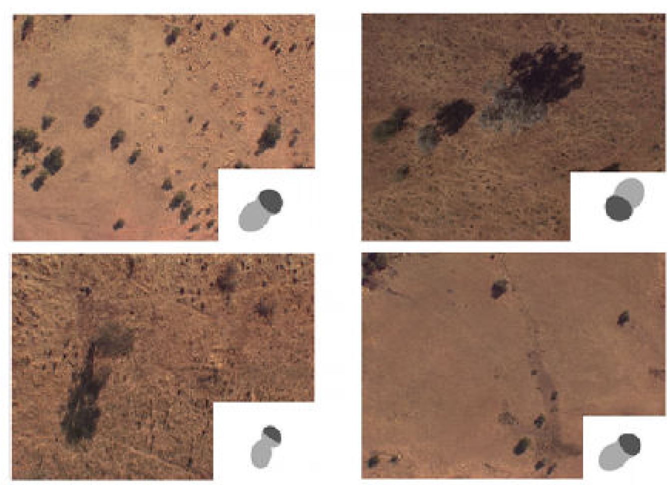

Figure 3: Predictive Template: The examples shown here are taken from different times during the flight. On the appearance model the darker area correspond to the object whereas the lighter area corresponds to the shadow. It can be seen that the relative position of the object and shadow is predicted correctly, also the ellipsoid shape assumption of the target is valid. Note that the templates are not plotted to the true scale.

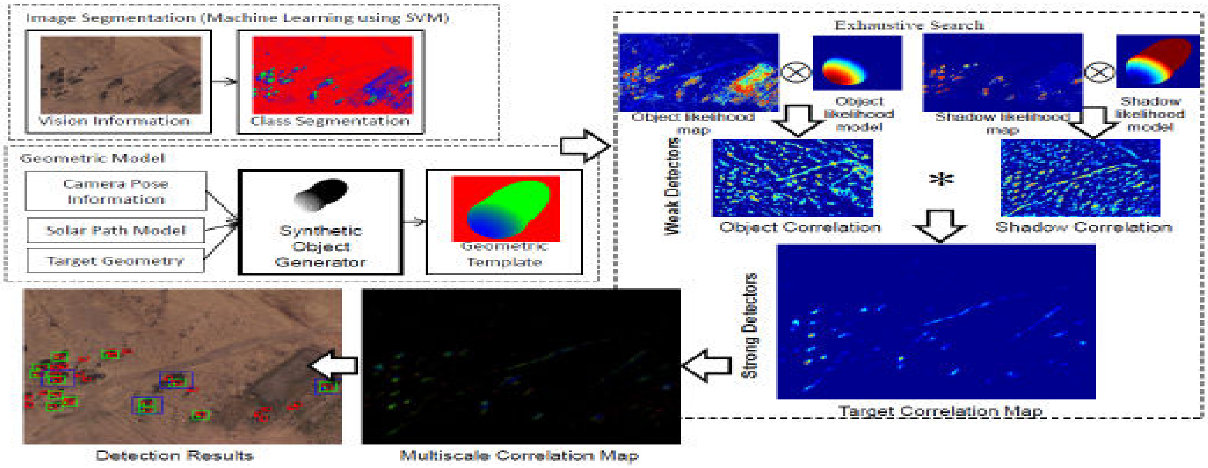

Figure 4: Summary of the Detection Stage

The object and shadow template are matched with the corresponding class segments. Two correlation maps are generated from each frame, these correlation maps use the shape information and act as weak detectors. The two individual correlation maps are then combined using the Hadamard product to generate the final detection map. This step encapsulates the context information, that is, the relative positions of the object and shadow. The final correlation map can be seen as a strong detector resulting from the combination of two weak ones.

C. *Classification*

The detection stage identifies the potential location of the tree crowns, colour and texture information are extracted around these locations for further analysis. SVM is applied to differentiate the tree crown species. Training data is generated according to the tree species and location ground truth data collected during the flight trial.

III. Experimental Setup

The study sites were at two cattle farms near the township of Julia Creek, northwest Queensland, Australia. The area is at about 650 km west of Townsville and 270 km east of Mount Isa. The survey area lies on the border of the Mitchell Grass Downs and the Gulf Plains bioregions, the area is flat and is dominated by Mitchell grass (*Astrebla* spp.), trees and shrubs are distributed sparsely within the area. The population density of the trees is typically higher along waterways.

The aim of the survey was to determine the distribution of different vegetation species around the area, mainly focused on the invasive woody weed species including prickly acacia (*Acacia nilotica*), parkinsonia (*Parkinsonia aculeata*), mesquite (*Prosopis pallida*) and mimosa bush (*Acacia farnesiana*). These weeds cause significant damage to the environment by out competing with the native species including Coolibah (*Eucalyptus coolabah/microtheca*) and whitewood (*Atalaya hemiglauca*), they also reduce the productivity of the farming industry.

The robotic platform used to collect data was a one third scale J3 Cub aircraft with a sensor payload consisting of a 3CCD camera with 200m by 140m field of view. The sensor spectral is the three visible band red, green and blue. The image resolution was 1024 by 768 pixels. During level flight each pixel represented ground coverage of approximately 20 cm by 20 cm area. The navigation data was collected using onboard GPS and inertial sensors. Multiple flights were performed during the mission at an altitude setting of either 500 metres. The survey was carried out at two separate mission, one on July 2009 and the second one on August 2010. In total 10 flights were carried out, each flight lasted around one hour, with approximately 16000 frames collected during each flight.

IV. Results and Discussions

Image segmentation was performed using SVM, cross validation was used to test the model learned from the data to prevent overfitting. A five-fold cross validation was used where three classes of data, object, shadow and background were collected from the image data set. 100 training instances were collected for each class. The optimised segmentation accuracy was found to be at 95.6%. The confusion between object and background occurred because in this data set there were a few instances when the object trees and the background grass had very similar colour and texture features. Humans are good at distinguishing the difference between tree and background because humans take additional context information into account. In this case a human observer would know shadows only exist next to the actual tree whereas the segmentation algorithm, based on only colour and texture information, is unable to do so. The object detection algorithm based purely on colour and texture is therefore not robust, this was one of the main reasons the second stage of the detection using the context information was required.

In this data set the target objects were trees, they varied slightly in size depending on environmental conditions. Three different template sizes were defined according to the prior knowledge on the typical size of the trees, these templates were created to capture the most targets within the size range.

The output of the algorithm was a continuous value which indicated the likelihood of detection, to evaluate the overall performance of the algorithm, the correlation map was thresholded to produce regions of interest, the centroids of each region were used as the position of detection. To optimise the threshold value and evaluate the performance of the object recognition algorithm, 30 images were selected throughout the entire duration of the flight as the ground truth data set. This wide range of images takes into account different factors including intrinsic properties such as lighting conditions, and extrinsic object related properties such as concentration of vegetation, size of vegetation, shadow orientation and shape of the shadow projection at different times. This evaluation was used to indicate the robustness of the entire algorithm. The target objects were hand labelled in each selected ground truth image. The overall detection accuracy was at 80%.

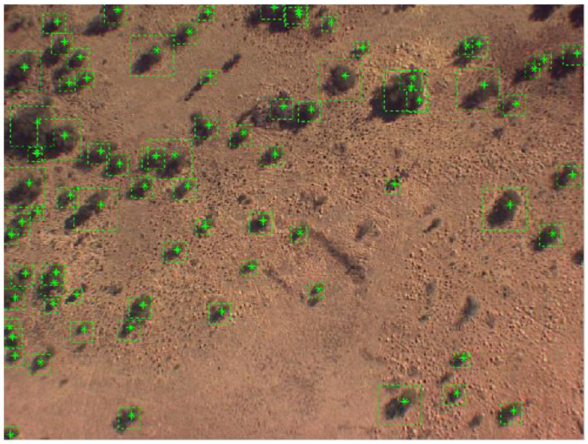

Figure 5: Detection results in a single image: the centroid of the crowns are ploted with stars and the rectangular bounding boxes indicates the size of the detected tree crowns.

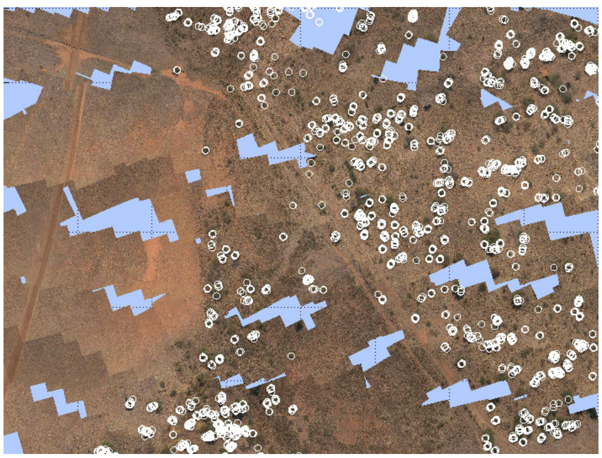

Figure 6: Detection results over a selected mission area, the white circles are the points of detection, multiple circles occasionally shows up for the same tree because the same tree crown was detected in separated frames. The individual image frames are joint together using algorithm described in (Bryson, 2010)

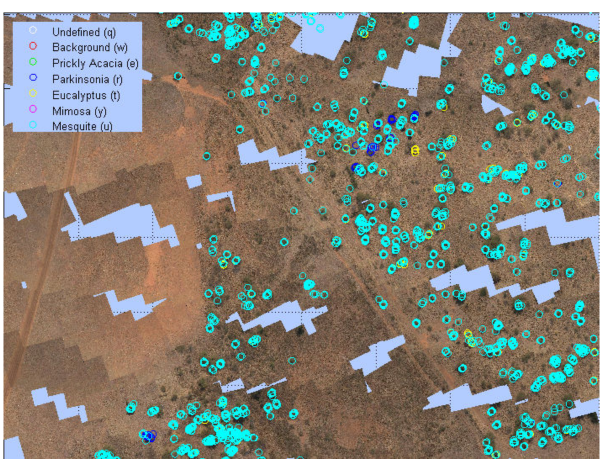

Figure 7: The species classification results of the detected tree crowns.

V. Conclusions

This paper presents a tree crown recognition algorithm using aerial photos taken from a UAV. The algorithm utilises two aspects of pattern recognition, a statistical model learnt from the data and prior knowledge from the understanding of the problem. The algorithm learnt colour and texture features of the target object and matched the labelled image to the predictive template created using prior knowledge. A strong detection map is generated by combining both the object and shadow detection maps, encapsulating the context information. Finally a SVM classifier is applied to further differentiate the tree crown species.

There are a few limitations of the algorithm, the most important one is that the algorithm requires minimal cloud coverage to provide identifiable shadows; the mission terrain should be relatively flat to ensure proper shadow projection.

For future work, this algorithm could also be extended to solve other object recognition problems within other UAV aerial imaging scenarios, such as power-line surveying, traffic monitoring and wildlife population monitoring.

References

A. Agrawal, Y. Sun, J. Barnwell, and R. Raskar. Vision-guided Robot System for Picking Objects by Casting Shadows. *The International Journal of Robotics Research*, 2010.

M.J. Brooks and B.K.P. Horn. Shape and source from shading. *Shape from shading*, 1989.

RA Brooks. Symbolic reasoning among 3d objects and 2D models. *Artificial Intelligence*, 17:285–348, 1981.

M. Bryson, A. Reid, F. Ramos and S. Sukkarieh. Airborne Vision-Based Mapping and Classification of Large Farmland Environments. *Journal of Field Robotics,* 27(5): 632-655, 2010

P. Bunting and R. Lucas. The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data. *Remote sensing of Environment*, 101(2):230–248, 2006.

M. Erikson. Segmentation of individual tree crowns in colour aerial photographs using region growing supported by fuzzy rules. *Canadian Journal of Forest Research*, 33(8):1557–1563, 2003.

M. Erikson and K. Olofsson. Comparison of three individual tree crown detection methods. *Machine Vision and Applications*, 16(4):258–265, 2005.

FA Gougeon. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing*, 21(3):274–284, 1995.

A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern recognition*, 24(12):1167–1186, 1991.

M. Larsen and M. Rudemo. Using ray-traced templates to find individual trees in aerial photographs. In *Proceedings of the Scandinavian Conference on Image Analysis*, volume 2, pages 1007–1014. Citeseer, 1997.

B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(8):837 –842, August 1996.

K. Olofsson, J. Wallerman, J. Holmgren, and H. Olsson. Tree species discrimination using Z/I DMC imagery and template matching of single trees. *Scandinavian Journal of Forest Research*, 21:106–110, 2006.

R.J. Pollock. The automatic recognition of individual trees in aerial images of forests based on a synthetic tree crown image model. *The University of British Columbia (Canada)*, 1996.

D. Thompson and J. Mundy. Three-dimensional model matching from an unconstrained viewpoint. In *1987 IEEE International Conference on Robotics and Automation. Proceedings*, volume 4, 1987.

V.N. Vapnik. *The nature of statistical learning theory*. Springer Verlag, 2000.