21st International Congress on Modelling and Simulation, Gold Coast, Australia, 29 Nov to 4 Dec 2015
www.mssanz.org.au/modsim2015

Automated detection and segmentation of vine rows using high resolution UAS imagery in a commercial vineyard

A.P. Nolan [a], S. Park [a], M. O'Connell [b], S. Fuentes [c], D. Ryu [a] and H. Chung [d]

[a] *Department of Infrastructure Engineering, The University of Melbourne, Parkville, Victoria, Australia*
[b] *Department of Economic Development, Jobs, Transport and Resources, Tatura, Victoria, Australia*
[c] *Department of Agriculture and Food Systems, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia*
[d] *Department of Mechanical and Aerospace Engineering, Monash University, Clayton, Victoria, Australia*

Email: anolan1@student.unimelb.edu.au

Abstract: Climate models predict increased average temperatures and water scarcity in major agricultural regions of Australia over the coming decades. These changes will increase the pressure on vineyards to manage water and other resources more efficiently, without compromising their high quality grape production. Several studies have demonstrated that high-resolution visual/near-infrared (VNIR) vineyard maps acquired from unmanned aerial systems (UAS) can be used to monitor crop spatial variability and plant biophysical parameters in vineyards. However, manual segmentation of aerial images is time consuming and costly, therefore in order to efficiently assess vineyards from remote sensing data, automated tools are required to extract relevant information from vineyard maps. Generating vineyard maps requires separating vine pixels from non-vine pixels in order to accurately determine vine spectral and spatial information. Previously several image texture and frequency analysis methods have been applied to vineyard map generation, however these approaches require manual preliminary delineation of the vine fields. In this paper, an automated algorithm that uses skeletonisation techniques to reduce the complexity of agricultural scenes into a collection of skeletal descriptors is described. By applying a series of geometric and spatial constraints to each skeleton, the algorithm accurately identifies and segments each vine row. The algorithm presented here has been applied to a high resolution aerial orthomosaic and has proven its efficiency in unsupervised detection and delineation of vine rows in a commercial vineyard.

Keywords: *Photogrammetry, image processing, precision viticulture*

1. INTRODUCTION

The Precision Viticulture (PV) approach utilizes crop phenological information and vineyard performance attributes to maximize yield and quality of grapes. Remote Sensing is one of the major tools used in PV for multi-temporal monitoring size, shape and vigor of grapevine canopies (Comba, et al. 2015). In order to efficiently evaluate vineyard performance attributes from remotely sensed data, automated tools are required to rapidly extract vineyard maps of relevant information from aerial imagery. The production of vineyard maps requires the separation of vine pixels from non-vine pixels for the determination of spectral information (e.g., photosynthetic-active, crop water stress) (Rabatel, et al. 2008) and spatial information such as a quantitative description of crop structure (canopy design, row plant spacing and row orientation) (Wassenaar, et al. 2002). For the production of accurate vineyard maps, image features such as roads, buildings and all non-vine row vegetation needs to be identified and removed to aid in the accurate estimation of plant biophysical parameters.

Various spectral and spatial approaches for vine field and vine row detection have been proposed for aerial imagery in general. A simple spectral approach is to assume that all vine canopy pixels will have a reflectance or vegetation index value greater than a threshold (Hall, et al. 2003). However, the similarities in the spectral response of inter row grass and other vegetation with that of vines make it difficult to differentiate between them. Alternatively, vine fields have a very specific and clearly defined spatial pattern that should allow for very effective filtering using texture or frequency analysis (Rabatel, et al. 2008). The detection of amplitude peaks and Hough transformations have been used for the accurate evaluation of inter-row width and row orientation (Wassenaar, et al. 2002). However, these methods require manual preliminary delineation of vine fields due to frequency filtering alone not being selective enough to separate vine fields for other agricultural fields (Rabatel, et al. 2008). In addition, the performance of textural analysis methods degrades when the periodic pattern of the rows is disrupted by row discontinuities caused by missing vines and other vineyard structures (e.g. sheds, irrigation infrastructure and native vegetation).

The objective of our research is to design a robust algorithm suitable for the automated delineation of vine-rows from aerial imagery for segmentation purposes. In this paper, a novel automated algorithm that uses skeletonisation techniques to reduce the complexity of agricultural scenes into a collection of skeletal descriptors is proposed. A series of geometric and spatial constraints are applied to each skeleton to accurately identify and segment vine rows. The algorithm presented here has been applied to a high resolution aerial orthomosaic and has proven its efficiency in unsupervised detection and delineation of vine rows.

The increased use of UAS in Precision Agriculture and Viticulture applications requires automated algorithms for fast, robust and cost effective analysis of remote sensing images to assess target crops. The proposed method aims to address these requirements and has the potential to be applied to other horticultural systems with distinct row and canopy configurations (e.g. fruit orchards and vegetable crops).

2. MATERIALS AND METHODS

2.1. Overview

The algorithm presented in this study uses ‘single band’ imagery to determine the spatial structure of aerial imagery. The imagery can be a linear combination of bands, a vegetation index such as Normalised Difference Vegetation Index (NDVI) (Rouse Jr, et al. 1974), thermal or a single image band of multispectral imagery. The only requirement for the input imagery is sufficient spatial resolution to achieve a contrast between vine row and background pixels.

The image processing algorithm, shown in Figure 1, consists of three main steps; i) histogram filtering, ii) skeletonisation and iii) vine row identification. The histogram filtering step aims to create a ‘rough’ binary mask of possible vine row pixels using a local histogram filter, which is sensitive to heterogeneous regions. The skeletonisation steps aims to create a geometric descriptor of each object by deconstructing them into a collection of interconnected branches. Finally, the vine row identification step uses geometric and spatial constraints in a local neighborhood to identify vine row clusters. Each processing step of the algorithm is applied sequentially, without user intervention, producing an image mask containing all detected vine rows and a quantitative description of the crop structure (planting pattern, spacing and orientation).

The image processing algorithm has been implemented in C++ using the software framework Open CV (Bradski 2000). The algorithm was executed on a 3.4GHz Intel i7 desktop computer with 8 GB of RAM memory.

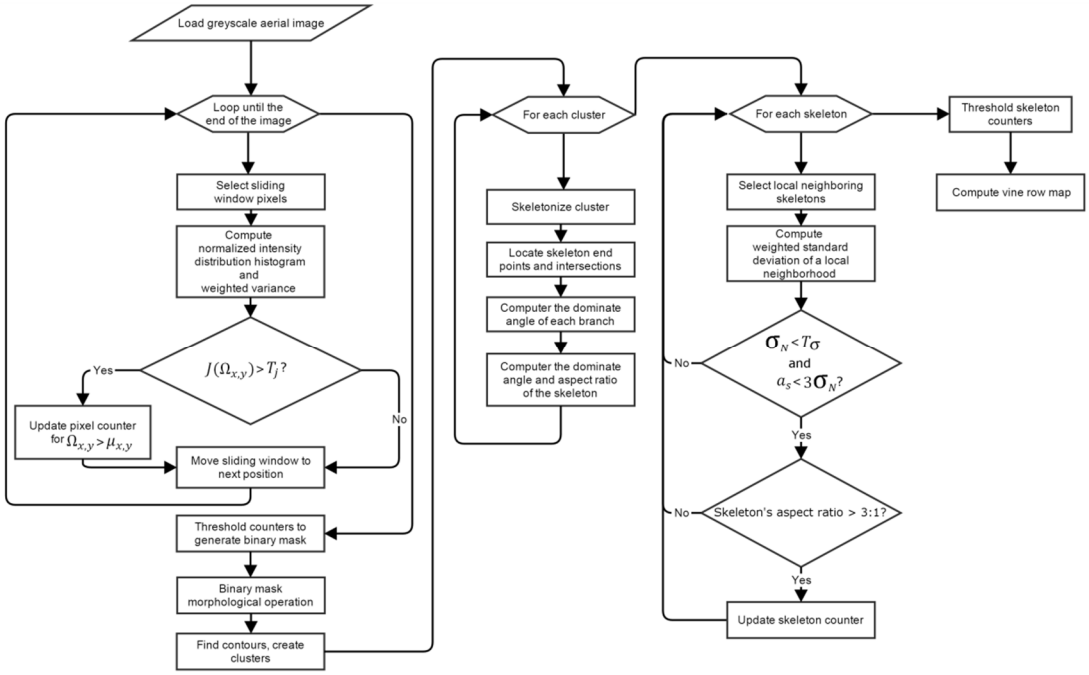

Figure 1. Flow chart of the vine row skeletonisation algorithm.

2.1. Study area and data acquisition

The study area was located at Curly Flat Vineyard (37°17′40″S, 144°42′24″E, 520 m.a.s.l.), in Lancefield, Victoria, Australia. The imagery was acquired using a near infrared (NIR) camera installed in a senseFly eBee UAS with a ground sample distance (GSD) of 4 cm. The images (n=280) were captured pre-harvest (DOY 88) with vine canopies at maximum cover (mid-season), on a cloud free day at solar noon to minimise the effect of shadowing. To aid in georeferencing, the precise positions of 14 ground control points (GCP) were measured using a Leica Viva GNSS-GS15 DGPS, providing a centimetre positional accuracy and 2-4 cm vertical accuracy. The aerial images were geometrically corrected and post processed using Pix4D photogrammetry software to generate a 48 ha georectified orthomosaic. The resulting orthomosaic, shown in Figure 2 contains approximately 14 ha of vine fields, buildings, roads, a water reservoir and non-vine vegetation. The orthomosaic was converted into a single band image for input into the algorithm by a linear combination of bands.

Figure 2. False color image of Curly Flat Vineyard Lancefield, Victoria, Australia.

2.2. Histogram Filtering

An initial image filtering step, based on a histogram slicing approach of Comba (2015), aims to create a 'rough' binary mask of possible vine row pixels using a local histogram based filter, sensitive to heterogeneous regions (Comba, et al. 2015). The normalised intensity distribution histogram of a local region containing vine rows is generally bi-modal, whereas homogeneous regions are bell-shaped. A subset of pixels in a square sliding window $\Omega_{x,y}$ of size l_w, approximately 2 times the vine row spacing l_r, is used to compute the weighted variance of the region's histogram as an index of its heterogeneity, defined by the function (Comba, et al. 2015):

$$J(\Omega_{x,y}) = \frac{\sum_{v=1}^{255} n(v) \cdot (v - \mu_{x,y})^2}{N} \tag{1}$$

where $n(v)$ is the number of pixel characterized by DN intensity value v, $\mu_{x,y}$ is the mean of the intensity histogram for $\Omega_{x,y}$ and N is the number of sizes in $\Omega_{x,y}$. If a local window's $J(\Omega_{x,y})$ index value is greater than a threshold T_j, then all pixels belonging to the sliding wind ($\Omega_{x,y}$), greater than the mean of the intensity distribution histogram $\mu_{x,y}$, are classified as having a high probability of belonging to heterogeneous regions and their counters are increased. Otherwise, if $J(\Omega_{x,y})$ is lower than T_j (bell shaped curve) the region is classified as homogeneous and therefore no counters are incremented. Once the entire image has been processed, pixels with a counter value greater than a threshold are used to generate a binary mask. A series of morphological operations are performed on the binary mask to remove noise and create interconnected clusters. The resulting binary mask, shown in Figure 3b, forms the input for all subsequent processing steps.

2.3. Contour Recognition

The binary mask is processed into a collection of clusters through the topological analysis of interconnected pixels (Suzuki 1985). By following the outer most pixels of an interconnected region, the algorithm constructs a contour around each cluster. The resulting collection of contours defines a shape descriptor for all objects in the scene for further geometric processing. However, inter row vegetation, such as dense grass, and interconnected vine rows often result in complex geometric shapes (Figure 3d & 3e). Identifying vine rows in a collection of complex shapes is achieved by skeletonising each shape into a collection of interconnected branches.

2.4. Skeletonisation

Skeletonisation is the process of iteratively eroding an object into a thin line drawing known as a skeleton. The thinned process must preserve the basic structure of the original object and its connectedness (Wang, et al. 1989). The fast parallel thinning algorithm proposed by Wang et al. (1989) interactively evaluates the contours of the object, evaluating each pixel using a deletion criterion until there are no pixels remaining to delete (Wang, et al. 1989) (Figure 3d & e). After skeletonisation, the thin line representation of the object is deconstructed into a collection of interconnected branches by determining the location end points and intersections. These end points and intersections are detected by counting the number of connected pixels for each skeleton pixel. End points are defined as a pixel location with only one connected pixel and intersections are defined as a pixel location with three or more connected pixels. Intersections often consist of a cluster of pixels with three or more connections, in which case the pixel with the highest number of connections is selected as the intersection location. If the highest connection count is shared by multiple pixels, the pixel closest to the center of the cluster is selected as the intersection location. For instance, the complex skeletal structure of regions of trees (Figure 3d) or vine rows (Figure 3e) are reduced by the algorithm to a collection of connected branches by detecting end point and intersection locations.

An approximate angle of each branch is calculated in degrees, using the formula:

$$a = \tan^{-1}\left(\frac{x_1 - x_2}{y_1 - y_2}\right) \cdot \left(\frac{180}{\pi}\right) \tag{1}$$

where x_1 and x_2 are the horizontal components and y_1 and y_2 the vertical component of the branch end points. The dominant angle of the entire skeleton is calculated using the weighted mean a_s, defined as:

$$a_s = \frac{\sum_{i=1}^{n} l_i a_i}{\sum_{i=1}^{n} l_i} \tag{2}$$

where a_i is the angle and l_i the length of branch i. Weighting the mean by branch lengths, favors longer branches and reduces the influence of interconnected vine rows branches, improving the estimation of a skeleton's dominant angle. Similarly, the weighted standard deviation of a skeleton σ_s defined as:

$$\sigma_s = \sqrt{\frac{\sum_{i=1}^{n} l_i (a_i - a_s)^2}{\sum_{i=1}^{n} l_i}} \tag{3}$$

provides a useful index for the complexity of the skeletal structures. For instance, the large collection of interconnected trees, shown in Figure 3d, has a weighted mean $a_s = 46°$ and a $\sigma_s = 30°$. However, the 50 vine rows (an example shown in Figure 3e) planted parallel to the trees have an average $a_s = 92° \pm 1°$ and a $\sigma_s = 0.5° \pm 0.1°$. A visual representation of each skeleton's dominant angle is shown in Figure 3c.

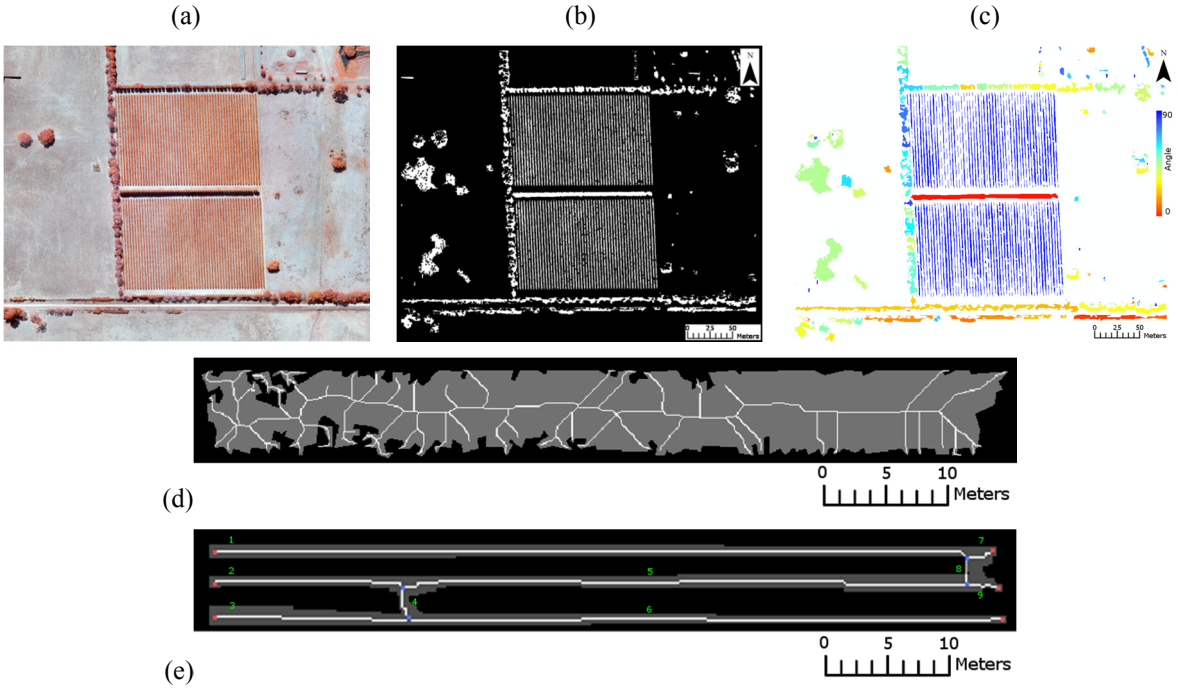

Figure 3. (a) A section of the original image. (b) Binary mask after histogram slicing. (c) Visual representation of dominant angles. (d) A complex skeleton from a region of trees. (e) A skeleton of three interconnected vine rows with junctions (blue), end points (red) and branch labels shown (green).

2.5. Vine row identification

In order to differentiate vine rows from non-vine row objects each skeleton was assessed against other skeletons in the local neighborhood. It was assumed that vine rows are predominantly straight, evenly space and planted in parallel. Therefore, the angle standard deviation σ_N of all skeletons in a local neighbourhood will be lower in a neighborhood with a large percentage of vine rows than in a neighbourhood containing non-repeating/unstructured skeletons of vegetation objects. Weighted standard deviation of a local neighborhood σ_N was defined as:

$$\sigma_N = \sqrt{\frac{\sum_{i=1}^{n} L_i (a_s - \bar{a}_n)^2}{\sum_{i=1}^{n} L_i}} \tag{4}$$

where L_i is a skeleton's length, a_s a skeleton's mean angle and $\bar{a}_n$ the weighting mean of all skeletons in the local neighbourhood. The local neighbourhood N is defined as all skeletons that intersect with a square window of size $l_N = 6 \cdot l_r$ pixels, centered on the middle of the skeleton being evaluated. Skeleton counters are incremented if (1) the local neighbourhood's standard deviation σ_N is below the threshold $T_\sigma = 5°$ (2) the skeleton's weighted mean a_s is less than $3\sigma_N$ and (3) the skeleton's aspect ratio is greater than 3:1. Skeletons are classified as being part of a vine row if its counter exceeds the threshold T_c. The skeleton evaluation criterion makes the vine row identification algorithm sensitive to objects with similar characteristics in the local neighbourhood and removes unstructured and noisy objects from the segmentation mask.

2.6. Evaluation

The source orthomosaic contains seven vine fields (14 ha) spread over a 30 ha vineyard. The entire image was processed by the algorithm to produce an image mask containing the location of all detected vine row pixels. To assess the invariance to row orientation by the algorithm, the orthomosaic was also processed with 45° and 90° counter clockwise rotation (CCW). To enable the evaluation of our automated method, all vine rows were manually annotated using the online tool LabelMe (Russell, et al. 2008) using a boundary evaluation threshold distance of one pixel allowed for small manual segmentation errors. The algorithm was evaluated by assessing the binary classification of each image pixel against the manually annotated image. For evaluation purposes, True and False Positives (TP/FP) refer to the number of correct/incorrect pixels classified as vine row and similarly True and False Negatives (TN/FN) for non-vine row pixels. Three measures were calculated for evaluation; i) Precision (TP/TP+FP), a fraction of detections that are true positives rather than false positives, ii) Sensitivity (TP/TP+FN), a fraction of true positives that are detected rather than missed and iii) False Negative Rate (1-(TP/TP+FN)), percentage of vine row pixels falsely classified as being non-vine row pixels (Powers 2011).

3. RESULTS

The algorithm's detection mask was applied to the original NIR imagery to accurately isolate all of the vine rows for further processing, as seen in Figure 6.

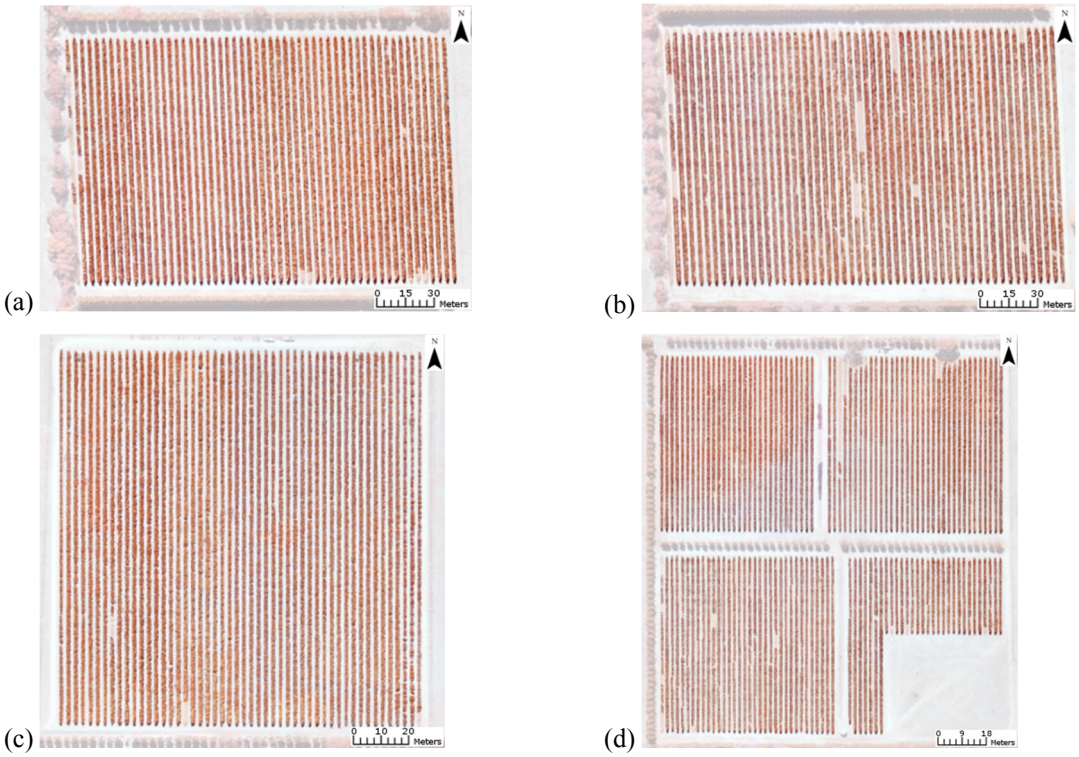

Figure 6. Detected vine rows overlay on faded original imagery.

Table 1. Vine row detection evaluation.

#	Image Section (depicted in Figure 6)	Sensitivity	Precision	False Negative Rate
1	Section a & b	0.976	0.987	0.024
2	Section a & b (rotated CCW 45°)	0.959	0.978	0.041
3	Section a & b (rotated CCW 90°)	0.973	0.957	0.027
4	Section c & d	0.987	0.973	0.013
5	Section c & d (rotated CCW 45°)	0.946	0.965	0.054
6	Section c & d (rotated CCW 90°)	0.986	0.969	0.014
	Average	**0.971**	**0.971**	**0.029**

The vine row detection algorithm achieved average precision and sensitivity results of 0.971 and 0.971, respectively, as shown in Table 1. Some sections of vine rows have been falsely classified as being non-vine row pixels (average 0.029), due to overhanging trees, shadows or initial binary segmentation discontinuities. Figure 6d contains the largest section of misclassified pixels due to neighbouring trees obscuring a section of vine row. The entire image (14 ha) was classified in less than three minutes using the computer capabilities mentioned previously. By comparison, manual approaches typically take 3 hours for 14 ha of vineyard.

4. DISCUSSION AND CONCLUSIONS

In this paper, a method for the automated delineation of vine-rows from high resolution aerial imagery was presented. The proposed method reduces the complexity of agricultural scenes into a collection of skeletal descriptors that enable the application of geometric and spatial constrains to accurately identify vine rows. The results obtained from high resolution aerial images of a 30 ha vineyard demonstrate the effectiveness of the proposed method with high precision (0.971) and sensitivity (0.971) results. Pixel misclassifications were generally due to neighbouring trees obscuring vine rows, shadows or initial binary segmentation discontinuities (average 0.029). In future work these issues will be resolved by using additional skeletal geometric conditions.

The automatically extracted vine row maps can be used as a PV tool for multi-temporal monitoring of vineyard spatial variability, shape and vigor to aid in the application of variable-rate treatments and irrigation scheduling. The algorithm has been designed to minimize the number of parameters required and to automatically adapt to various spatial resolutions. The method has potential applications to other horticultural systems with distinct row and canopy configurations (e.g. fruit orchards and vegetable crops), various sensor types (e.g. thermal, multispectral) and vegetation indices (NDVI, Leaf Area Index (LAI) and Crop Water Stress Index (CWSI)).

Future work will focus on the development of faster algorithms, reducing the number of pixel misclassifications and the integration of our vineyard maps with path planning tools for autonomous navigation of ground and aerial vehicles travelling between vine rows.

ACKNOWLEDGMENTS

This research was funded by a Seed Fund for Horticulture Development grant (602948) from the University of Melbourne and the Department of Economic Development, Jobs, Transport and Resources, Victoria, ARC LIEF grant (LE130100040) and Melbourne Networked Society Institute (MNSI) Seed Funding. We thank Curly Flat Vineyard for permission to utilize their vineyard.

REFERENCES

Bradski, G., (2000), The opencv library, *Doctor Dobbs Journal*, 25, 120-126.

Comba, L., Gay, P., Primicerio, J. and Aimonino, D. R., (2015), Vineyard detection from unmanned aerial systems images, *Computers and Electronics in Agriculture*, 114, 78-87.

Hall, A., Louis, J. and Lamb, D., (2003), Characterising and mapping vineyard canopy using high-spatial-resolution aerial multispectral images, *Computers & Geosciences*, 29, 813-822.

Powers, D. M., (2011), Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation,

Rabatel, G., Delenne, C. and Deshayes, M., (2008), A non-supervised approach using Gabor filters for vine-plot detection in aerial images, *Computers and Electronics in Agriculture*, 62, 159-168.

Rouse Jr, J. W., Haas, R., Schell, J. and Deering, D., (1974), Monitoring vegetation systems in the Great Plains with ERTS, *NASA special publication*, 351, 309.

Russell, B. C., Torralba, A., Murphy, K. P. and Freeman, W. T., (2008), LabelMe: a database and web-based tool for image annotation, *International journal of computer vision*, 77, 157-173.

Suzuki, S., (1985), Topological structural analysis of digitized binary images by border following, *Computer Vision, Graphics, and Image Processing*, 30, 32-46.

Wang, P. S.-P. and Zhang, Y., (1989), A fast and flexible thinning algorithm, *Computers, IEEE Transactions on*, 38, 741-745.

Wassenaar, T., Robbez-Masson, J.-M., Andrieux, P. and Baret, F., (2002), Vineyard identification and description of spatial crop structure by per-field frequency analysis, *International Journal of Remote Sensing*, 23, 3311-3325.