

# Towards a generalized colour image segmentation for kiwifruit detection

P. Wijethunga, S. Samarasinghe, D. Kulasiri,  
Centre for Advanced Computational Solutions,  
Lincoln University 7647,  
Canterbury, New Zealand  
Email: pavithra.wijethunga@lincolnuni.ac.nz

I. Woodhead  
Lincoln Ventures Limited,  
P. O. Box. 133, Lincoln 7640,  
Christchurch, New Zealand

**Abstract—** Developing robust computer vision algorithms to detect fruit in trees is challenging due to less controllable conditions, including variation in illumination within an image as well as between image sets. There are two classes of techniques: local-feature-based techniques and shape-based techniques, which have been used extensively in this application domain. Out of the two classes, the local-feature-based techniques have shown higher accuracies over shape-based techniques, but are less desirable due to the requirement of repeated calibration. In this paper, we investigate the potential of developing a generalized colour pixel classifier that can be employed to detect kiwifruit on vines, under variable fruit maturity levels and imaging conditions. First, we observed the colour data patterns of fruit and nonfruit regions from different image sets. With consistent data patterns it was found that a suitable normalization could produce an invariant colour descriptor. Then, a neural network Self-Organizing Map (SOM) model, which has a hierarchical clustering ability was used to investigate the potential of developing a generalized neural network model to classify pixels under variable conditions. Models were built for colour features extracted in CIELab space for both absolute colour values and relative colour descriptors. The paper presents the positive results of the preliminary investigations. The conditions for a successful application of the approach as well as the potential for extending it for automatic calibration will also be discussed.

**Keywords-** *pixel classifier, colour image segmentation, SOM, automatic calibration*

## I. INTRODUCTION

Vision based detection of fruit in trees was first applied in robotic fruit harvesting systems in the late 1970s [1]. Since then, fruit detection algorithms have been researched and applied in different fruit harvest robotics. Application of vision based techniques for pre-harvest estimation purposes has also become popular in recent years [2, 3]. A survey of vision techniques used to detect fruit in trees can be found in [4]. Jimenez et al [4] divide vision based techniques used in detecting fruit in trees into two categories, based on the analysis method. They are local feature-based techniques and shape-based techniques. Techniques based on local features consider pixel based properties such as intensity, colour, texture and other spectral band values. Shape based methods use morphological properties of a fruit object as a whole and the shape measurements are obtained from edge images or convexity analysis. Out of the two categories the local-feature-based methods have shown higher levels of accuracy, despite

the disadvantage of their need for calibration. On the other hand, when the background contains complex variations, shape based techniques do not perform well. Therefore, local features like colour become an essential feature in detecting objects as well as in isolating relevant information needed for other techniques to perform well.

When using local features, a requirement of calibration arises due to the variability of local features (mainly the intensities of spectral bands) with different imaging conditions. Furthermore, spectral reflectance properties of fruit and other background can vary slightly from season to season. When we deploy a vision system for pre-harvest estimation at several stages of fruit growth, the fruit colour differs slightly at different ages requiring calibrations in order to detect fruit accurately. However, the need of calibration makes a system less user friendly. Therefore, a generalized model which can detect fruit despite slight variations of the spectral reflectance properties and imaging conditions is of interest to the industry. In this paper, we investigate the potential of developing such a model for detecting kiwifruit within orchards, as an attempt to extend the vision system developed in [5] to be more generalized.

The problem of comparing images taken under different lighting conditions and by different cameras is often addressed by colour constancy techniques [6]. Such techniques include gray world assumption, white world assumption and gamut mapping techniques [7]. The images in our work consist of several complexities; the background of the images is uncontrollable and variable and different fruit maturity levels make the fruit colour also variable. Consequently, the surface reflectance can also be variable. Furthermore, we are interested in comparing images captured by different cameras and therefore the sensitivity functions also change. If the gray world assumption is considered, it removes the colour cast caused by an illuminant which is biased towards one colour, by adjusting each colour band's mean value to be equal to the mean value of the gray image [8]. This assumption is suitable for images which have lot of colour variation within the image. In the system we discuss, images are taken at night and images mostly contain fruit and canopy; hence there is little colour variation. Furthermore, only white light illuminants are often used in this application domain, so the gray world assumptions are not valid for the images in our application. The white world algorithm is also another variation of gray world assumption and further it assumes to have a white point, which can not be assumed to exist in every image of our application. Gamut mapping methods map gamut of an image under an unknown

illuminant into the gamut of the image under a known illuminant. If mapping is based on diagonal model, two gamuts can be mapped to each other by multiplying each band by a scalar value. If the shape of the gamut is complex, then the mapping has to consider the constraints; hence the convex hulls of gamuts are mapped [9].

Our approach is data driven and can also be grouped under the gamut mapping method. The approach is a convenient alternative based on the patterns of colour ranges in CIELab colour space for different image sets. After preprocessing, we used a convenient gamut shape in CIELab colour model. We present the effect of a suitable normalizing for gamut mapping towards achieving a generalized pixel classification model, using a Self Organizing Map (SOM) neural network.

The paper is organized in 5 Sections. Section 2 briefly discusses SOM neural network models and the model parameters used for this work. The data used for this work, pre-processing and hypothesis for normalizing is described in Section 3. Then, the results comparing original colour features and normalized colour features are presented in Section 4. Finally the conclusion and further work is discussed in Section 5.

## II. SOM CLUSTERING

Artificial Neural networks are massively parallel networks of processing elements (neurons) that can learn patterns of data via training. SOM neural networks use an unsupervised learning method to organize unknown data into groups of similar pattern based on distance measures. The most important feature of SOM is that it preserves the topological relationship between data so that the neighbouring inputs in the input space are mapped into neighbouring neurons; hence SOM is an ideal tool to visualize the clusters in large input data in a condensed form through the neurons arranged in a two dimensional lattice [10].

As the inputs are represented by neurons, input clustering can be achieved by clustering neurons, i.e. weight vectors. Neurons group the data into subgroups, each of which is represented by a neuron. Then, as in hierarchical clustering, the subgroups are grouped again to form the desired number of clusters. This way, a classifier can be trained to take in to account variations within one cluster and still be classified under one final cluster. SOM clustering techniques have proven superior performance in classifying data with imperfections, such as data dispersion, outliers, irrelevant variables, and nonuniform cluster densities [11]. In this study, we aim to achieve a SOM neural network model such that the feature vectors of fruit pixels are responded by neurons in one cluster where as the feature vectors of background pixels are responded by another cluster of neurons. In order to achieve a general pixel classifier, the model should respond similarly for images captured under different imaging conditions.

Matlab 2009 software and its Neural Network toolbox were used to implement SOM network. A SOM network model with 100 neurons arranged in a hexagonal topology was used for this work. Weights were randomly initialized and the batch training approach with 200 iterations was used to train the network. The factor determining whether the model is properly

trained or not, is the cumulative distances (D) between input vectors and corresponding winning neurons. The D is calculated as in (1), and the model is optimized when it is trained until the D is minimized so that all the neurons have spread themselves among input vectors. In addition, a sufficiently large number of neurons enables a lower value for D. We use 100 neurons which are expected to be sufficient to represent a simple linear shape of data. Furthermore, appropriate training can be verified by visually observing the plot of weight vectors on top of input vectors.

$$D = \sum_{i=1}^C \left\{ \sum_{n_i=1}^{N_i} (v_{n_i} - w_i)^2 \right\} \quad (1)$$

where  $C$  is the number of neurons,  $N_i$  is the number of input vectors responded by  $i^{\text{th}}$  neuron, and  $W_i$  is the winning weight vector corresponds to the  $i^{\text{th}}$  neuron.

## III. DATA AND HYPOTHESIS

### A. Data and Pre-processing

Three sets of three band (R, G and B) images captured at a New Zealand kiwifruit orchard were used for this study. Each set contains five representative images. Images were captured using a custom built acquisition setup used in [5]. The details of imaging conditions for each image set are shown in Table 1. An example image from each image set is shown in Fig. 1.

TABLE I. DETAILS OF IMAGE SETS

	Set 1	Set 2	Set 3
Camera	Lumenera Le256- high resolution	Lumenera Le256- high resolution	AVT Guppy F-080/C
Lighting	Array of fluorescent	Array of fluorescent	Halogen
Image size	1080x1920x3	1080x1920x3	768x1024x3
Month/year captured	December/ 2006 (3months prior to harvesting)	April/ 2007 (2 weeks prior to harvesting)	November/ 2008 (4 months prior to harvesting)

Images were pre-processed using the steps used in [5]. First, each colour band values were contrast stretched using a nonlinear function,  $y = kx^{0.5}$ , where  $k$  is a scalar and  $x$  is the input intensity value. The nonlinearity helps correcting illumination variations within the image. Secondly, the green colour band was omitted and the images were made false colour, replacing R, G and B with B, R and B respectively to

achieve better visual contrast. This false colour combination [5] creates an approximately linear relationship between  $a$  and  $b$  colour components (Fig. 2) in CIELab space, providing a convenient gamut shape to map between images. This approximately linear relationship was mathematically proved and can be collected from authors upon request. It places the fruit region data further towards one corner of the gamut and the non-fruit region data towards the other corner making only one separation line is sufficient.

colour space has several favourable properties, such as device independence and perceptual uniformity, that make it suitable for colour image segmentation [12]. The  $a$  and  $b$  chromaticity components were used as features for clustering image pixels.



(a)



(b)



(c)

Figure 1. Original image (a) set 1, (b) set 2, and (c) set 3

Then, the outliers of each colour band were removed by replacing the colour values of pixels which are six standard deviations away from the mean, with the mean value of the colour band. This outlier removal enables removing undesirable bright points present in some images, thereby making the range of colour values comparable between images. An average filter was then applied to smooth the image. Then we converted the image into CIELab colour space. CIELab

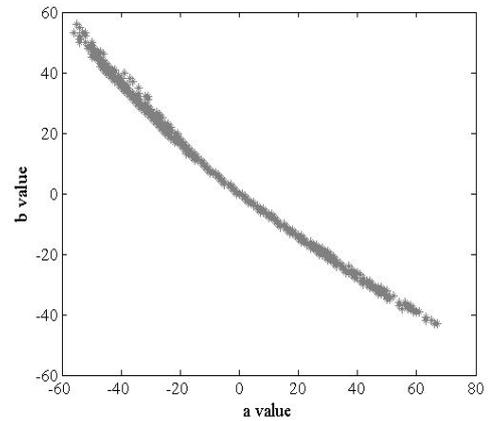
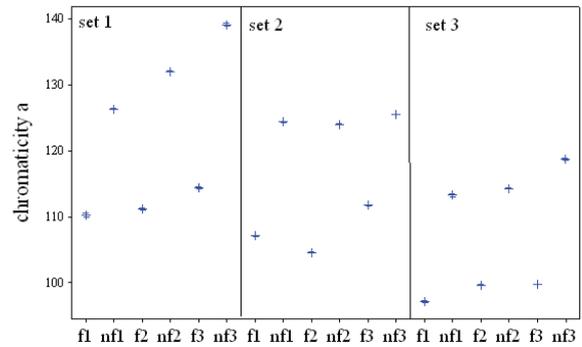


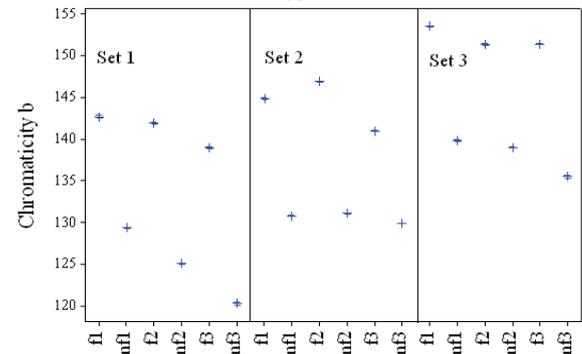
Figure 2: Gamut of  $a$  and  $b$  colour components of CIELab model for a preprocessed image

*B. Data analysis and hypothesis*

Two masks were created for each image in order to extract colour values of fruit regions and background. Regions for masks were selected in a way that the fruit and nonfruit regions are well represented, containing over 5000 sample pixels per region. Then, the patterns of data ranges for fruit and nonfruit regions were observed using interval plots (Fig. 3).



(a)



(b)

Figure 3. The 95% confident interval plot showing the pattern of colour data ranges (for (a) chromaticity  $a$ , (b) chromaticity  $b$  for fruit and non fruit regions for three images from each image set ('fi' and 'nfi' refer to fruit and non-fruit regions of  $i^{th}$  image respectively).

Even though, the variations of data are not clear due to the large number of data points, it can be seen from Fig. 2 that, both  $a$  and  $b$  colour values shows a consistent pattern. Fruit regions have higher  $b$  values and lower  $a$  values within the image compared to non-fruit regions. However, the colour values are not consistent. Particularly, the range of colour values for fruit from set 1 and 2 coincide with that for non-fruit regions from set 3.

Considering the pattern consistency, it can be argued that normalizing each colour band will bring the fruit and non-fruit colour values in to a common range. Different normalizing methods were studied. Normalizing the  $a$  and  $b$  values of the entire image using (2) was desirable as it does not distort the relationship between  $a$  and  $b$  while preserving the distances between similar valued clusters. The idea was to map  $a$ ,  $b$  values into illumination independent descriptors rather than using absolute colour values.

$$(a_n, b_n) = \left\{ \left( \frac{a}{(a+b+L)} \right), \left( \frac{b}{(a+b+L)} \right) \right\} \quad (2)$$

$a_n, b_n$  are the normalized  $a$  and  $b$  components respectively, and  $L$  is the illumination component.

However, as interval plots are not sufficient to properly visualize the variations and overlaps within fruit and non-fruit regions, a SOM visualization tool was used for further observations. Then, two Neural network SOM models were trained: (i) using original ( $a, b$ ) data from image set 1, and (ii) using normalized ( $a_n, b_n$ ) data from image set 1. The models were then tested using images from other two image sets. As we used the same masks to extract both the original and normalized colour data, the effect of normalizing can be compared by looking at the models' response to the data extracted from same regions.

#### IV. RESULTS

Figure 4 and Figure 5 shows the trained neural network SOM models and model responses for the three image sets. The neighbour distance map shows the neuron clusters, which are separated by larger distances (dark boundaries). The weight planes indicate each input variables' contribution in clustering data.

It can be seen from Figure 4 that neurons which respond to fruit data are located towards the right side of neuron lattice whereas the rest of neurons respond to non-fruit data. The fruit captured with the same conditions but at a different stage of fruit growth have still been responded to by the same neurons. However, when the camera and lighting condition changes, the model trained with original data has failed, misclassifying non-fruit regions as fruit (Fig. 3(i)).

The model trained with normalized input data (Fig. 4 (a)) has comparatively more clean cluster boundaries. The neurons representing fruit regions are located towards the top-right corner of the neuron lattice. Even though there are some overlaps, the model has successfully recognized the fruit region

despite the different imaging conditions (Fig. 4(h),(i)), demonstrating the possibility of developing an unsupervised colour image segmentation algorithm to detect kiwifruit within an orchard under variable imaging conditions and fruit age.

#### V. DISCUSSION AND FUTURE WORK

Considering the consistent patterns in fruit colour values with reference to the background colour values, the results show the potential for developing a hierarchical clustering based generalized pixel classifier. A Learning Vector Quantization network can be trained to cluster neurons (weight vectors) of the competitive layer (similar to SOM) into two regions: fruit and nonfruit.

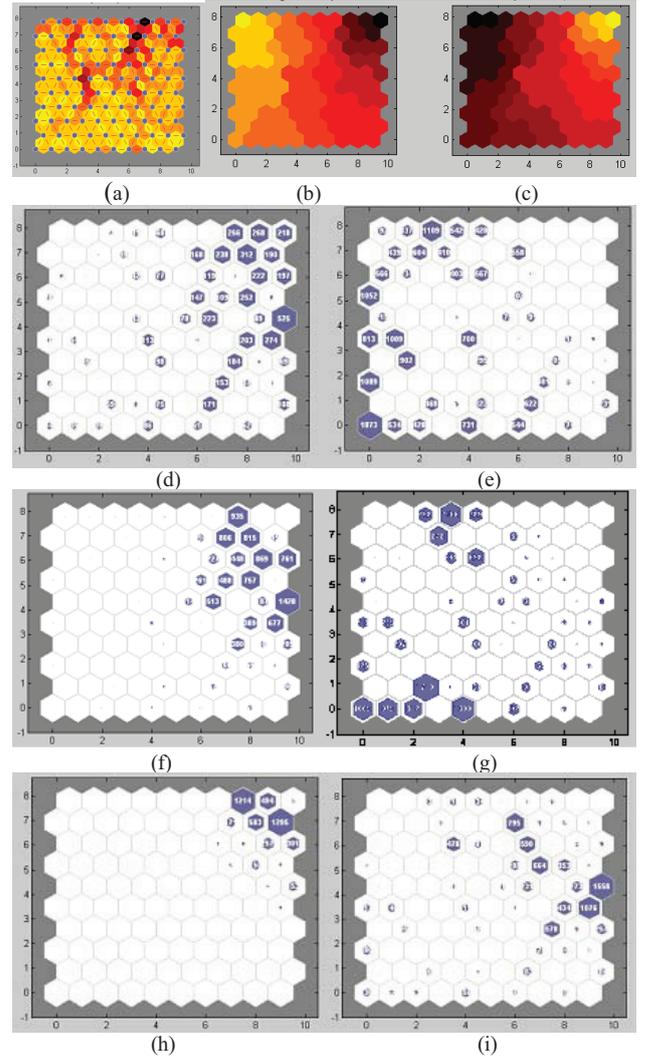


Figure 4: Neural network SOM model for original input data and model responses for fruit and nonfruit regions for three image sets. (a) neighbor distance between neurons representing the clusters of input data (lighter the colour lower the distance between neighbor neurons), (b) weight plane for input variable  $b$ , (c) weight plane for input variable  $a$ , (d),(e) sample hits for data of fruit and nonfruit regions respectively for image set 1. Similarity (f),(g) and (h), (i) are for image set 2 and 3 respectively

### A. Automatic calibration

Considering the conditions, under which automatic calibration could work, it can be expected that the variety of regions (fruit, leaves, dark background etc) present in the images should be similar to that of images used to train the model. Most of the images taken at the orchard contain a common set of regions, but there are also images with no fruit. In such cases, the classifier may falsely detect fruit regions.

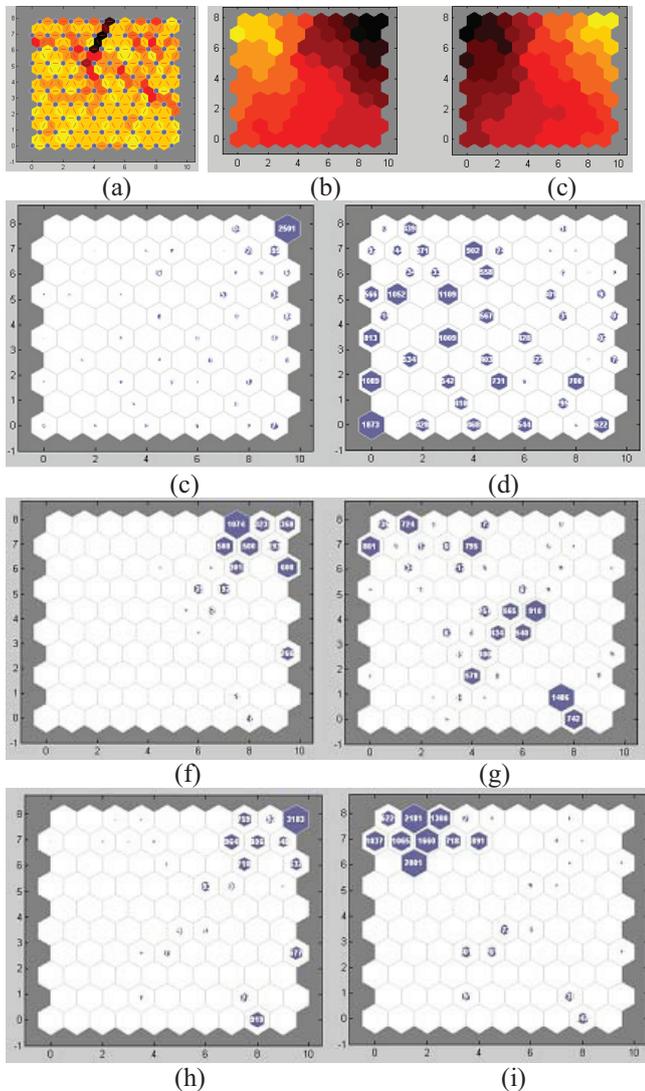


Figure 5: Neural network SOM model for normalized input data and model responses for fruit and nonfruit regions for three image sets. (a) neighbor distance between neurons representing the clusters of input data (lighter the colour lower the distance between neighbour neurons), (b) weight plane for input variable b, (c) weight plane for input variable a, (d),(e) sample hits for dat (f),(g) sample hits for dat (h),(i) sample hits for dat

Therefore, the model can be used first to automatically calibrate the classifier before using it for the whole image set. For that, a model trained by normalized data obtained from a sample image with a typical set of regions can be used. To calibrate the classifier for a new set of images, first the sample image is loaded and the fruit regions located. Then, the colour values of the image can be de-normalized and the range of

original colour values for fruit in the new image set can be recognized using the detected fruit regions. The classifier can then be adjusted to use original colour values, rather than normalized values. This calibration can be automated as it does not require user intervention. However, a suitable sample image from the image set should be provided by the user. With this calibration, the problem of false classification in images with no fruit regions, which occur when using normalized data, can be overcome. Hence, the model trained with normalized data can be better used as an automatic calibration technique.

### B. Future work

Further research will be carried out in order to test the performance of practical use of the classifier in terms of correct and false fruit detection under variable imaging conditions. Next, it is of much interest to use the model to automatically calibrate the system and to assess the performance.

### ACKNOWLEDGMENT

The authors thank Lincoln Ventures Ltd for the supply of image data, support and funding for this research by the New Zealand Foundation for Research, Science and Technology.

### REFERENCES

- [1] E. Parrish, and A. Goksel, "Pictorial pattern recognition applied to fruit harvesting," Transactions of the ASAE, vol. 20, no. 5, pp. 822-827, 1977.
- [2] D. Stajanko, M. Lakota, and M. Hocevar, "Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging," Computers and Electronics in Agriculture, vol. 42, no. 1, pp. 31-42, 2004.
- [3] Q. U. Zaman, A. W. Schumann, D. C. Percival et al., "Estimation OF Wild Blueberry Fruit Yield Using Digital Color Photography," Transactions of the ASABE, vol. 51, no. 5, pp. 1539-1544, 2008.
- [4] A. R. Jimenez, R. Ceres, and J. L. Pons, "A survey of computer vision methods for locating fruit on trees," Transactions of the ASAE-American Society of Agricultural Engineers, vol. 43, no. 6, pp. 1911-1920, 2000.
- [5] P. Wijethunga, S. Samarasinghe, D. Kulasiri et al., "Digital image analysis based automated kiwifruit counting technique." Paper presented at the Image and Vision Computing New Zealand, 23rd International Conference, Christchurch, 2008.
- [6] K. Barnard, G. Finlayson, B. Funt et al., "Colour constancy for scenes with varying illumination.", Paper presented at the Fourth European Conference on Computer Vision, Cambridge, UK, 1996.
- [7] K. Barnard, "Computational color constancy: Taking theory into practice," M.Sc., Computer Science, Simon Fraser University, 1995.
- [8] E. H. Land, and J. J. McCann, "Lightness and retinex theory," Journal of the Optical society of America, vol. 61, no. 1, pp. 1-11, 1971.
- [9] R. Schettini, G. Ciocca, and S. Zuffi, "A survey of methods for colour image indexing and retrieval in image databases," *Color Imaging Science: Exploiting Digital Media*, pp. 183-211, 2001.
- [10] J. Moreira, and L. F. Costa, "Neural-based color image segmentation and classification using self-organizing maps," Anais do IX SIBGRAPI, vol. 12, no. 6, pp. 47-54, 1996.
- [11] P. Mangiameli, S. K. Chen, and D. West, "A comparison of SOM neural network and hierarchical clustering methods," European Journal of Operational Research, vol. 93, no. 2, pp. 402-417, 1996.
- [12] S. Dai, and Y. Zhang, "Color image segmentation with watershed on color histogram and Markov random fields.", Paper presented at the Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on, Beijing, 2003.