Detection of Star-shaped Objects in Aerial Images: Use Case Oil Palm Tree Detection

ABHISHEK MANANDHAR[1], LUDWIG HOEGNER[1] & UWE STILLA[1]

Abstract: Object detection is one of the major methods of remote sensing. Generally in object detection features are extracted from the image to represent the object in the image numerically with lesser number of elements than the image itself. Out of multiple features, shape feature is one of the widely used features. In this study, new rotation invariant shape features are introduced and comparative studies are made among them by using them to detect a naturally occurring star-shaped object, palm trees, in aerial images. Detection of the palm trees helps in creating inventory which is an initial step towards monitoring.

This study proposes shape features, the circular autocorrelation of polar shape matrix (CAPS), Mean of CAPS and Entropy of CAPS to be used with object detection framework. The framework used in the study to evaluate features uses a well-known machine learning algorithm, support vector machine (SVM), to detect palm trees in aerial images.

The implementation on aerial images, taken with unmanned air vehicle (UAV) in five plantation regions in Indonesia, Malaysia, and Thailand, shows promising results. The CAPS produced the best results among the features studied, with an average accuracy of 84% over 8 images, chosen considering different challenges from the five plantation region.

1 Introduction

There are multiple examples of star-shaped objects in nature. Especially, in aerial images, objects like palm tree canopy, banana tree canopy, etc. have star shapes. The shape information of those objects can be used in detection and monitoring of those objects from aerial imageries.

Palm trees produce a wide variety of products like vegetable oil, biofuel, furniture etc. Palm vegetable oil is currently the most traded oil. In 2011, 53 million tons of palm vegetable oil was produced, with trees cultivated in 15 million hectares. The demand for the product is increasing and hence, the production is estimated to rise by 30% in Malaysia and Indonesia, the countries that contribute to more than 90% of trade worldwide (POTTS et al. 2014). This demand will lead to larger cultivation and, hence, the necessity for smart monitoring of the plantations. Detection of the palm trees in aerial imageries and keeping inventory can serve as a start.

Object detection and localization in aerial image is one of the major methods of remote sensing. Researchers, government, and non-government institutions use this method to detect and localize objects of interest like buildings, cars, ships, trees etc., which further help in creating inventories and monitoring the status of the objects. The approach has also been used for identifying hazards like wildfires in inaccessible areas.

Generally, in object detection, different features are extracted from an image to represent the object present in the image and those features are used with detection framework. Features are the numerical representation of the objects with relatively fewer elements than in the image itself.

---

[1] Technische Universität München, Photogrammetry and Remote Sensing, Arcisstrasse 21, D-80333 München, Germany, E-Mail: [Abhishek.Manandhar, Ludwig.Hoegner, stilla]@tum.de

There have been myriad of studies to derive different types of features which can be classified into different categories based on their use and characteristics.

Shape features are among the widely used features in object detection. As the name suggests, shape features are a representation of the shape of the object. As all other features, shape features also have to cope with different challenges, especially induced by different saliencies, rotation, scale, shift, illumination etc. Different features try to address different challenges according to need.

In this study, we derive a new rotation invariant shape feature, as we consider to detect the star-shaped object. Besides, scale and shift invariance can be achieved by using the moving window detection scheme at different scales, therefore, rotation invariance will give feature extra strength. The shape features namely, Circular Autocorrelation of Polar Shape Matrix (CAPS), Mean of CAPS, Entropy of CAPS, etc. will be derived from the star-shaped model and will be used with object detection framework to detect and count palm trees in different plantation regions.

2 Method

The primary objective of the study is to identify plausible shape feature which can be used in object detection framework for detection of the star-shaped object. The framework used in the study (Figure 1) consists of two major processes: Standard feature extraction and the Object detection. Standard feature extraction further comprises of two sub-processes, namely Generic shape feature extraction and Dimensionality reduction and standardization.

Fig. 1:   Object detection framework

In this framework, the process of generic shape feature extraction is kept as a varying process while keeping all other components fixed, in order to provide an identical platform for the evaluation of different generic features. The process of generic shape feature extraction is described in more detail in Section 2.1.

The generic shape features used in this study are of different dimensions, which pose a challenge in comparing them. Therefore, all features are standardized and the dimensions are reduced to one-dimensional scalar value using Support Vector Machine (SVM).

SVM is a standard tool for machine learning technology and widely used for solving problems in classification, regression and novelty detection (CHANG & LIN 2011; BISHOP 2006) because of its good generalization property for even small samples (LIU et. al 2012; VAPNIK & VAPNIK 1998; CORTES & VAPNIK 1995). In this method, the data is mapped into a higher dimensional input space and an optimal separating hyperplane, a straight line in case of two-dimensional input space, is constructed in this space (SUYKENS & VANDEWALLE 1999; CORTES, C. & VAPNIK 1995).

SVM uses training examples of objects and non-objects to determine the decision boundary between them. Generally, the trained SVM are used with sliding window detector where a window translates over the image, calculates the features and based on that feature decide directly the class for the window. But, in this study, it is used not to detect class but to calculate the distance from the decision boundary in feature space and assign it to the central pixel of the window. The process

is illustrated in Figure 2. The function $f(x)$, distance from the decision boundary in feature space, calculated at all pixels of the image result in surface, mentioned as the *feature map*, with peaks at the objects.
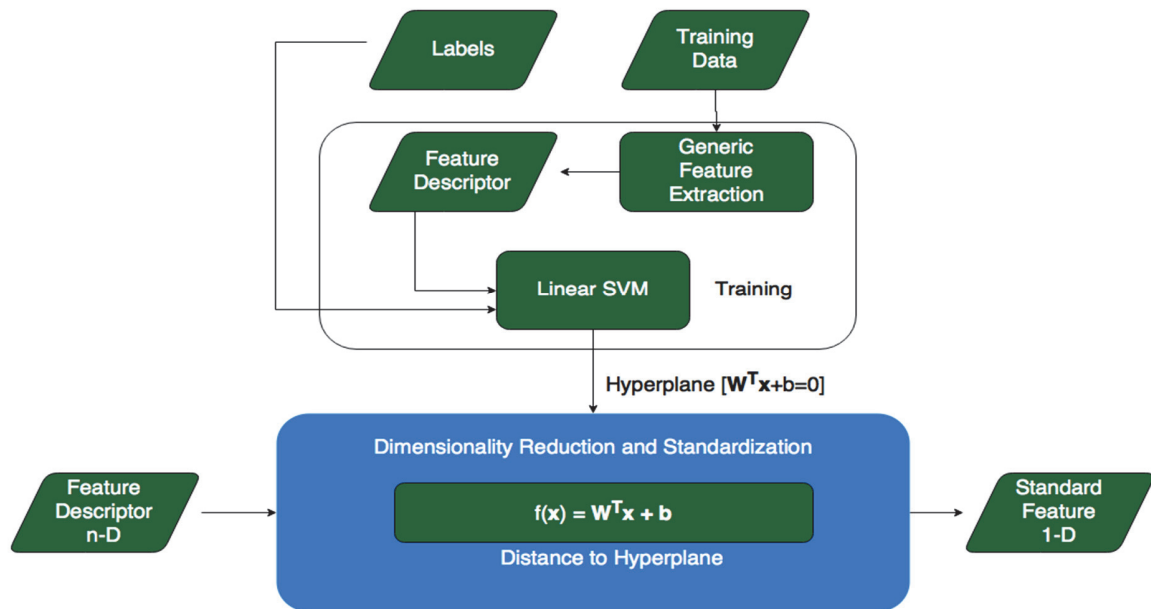

Fig. 2:   Dimensionality reduction and standardization of generic feature

A local maximum detection algorithm is applied to the feature map to achieve the spatial distribution of object centers in the image, called *object map*, as the final result. The results for different features are evaluated by comparing the results with the ground truths. The average of precision and recall are used as the accuracy measure. Precision or user's accuracy is the proportion of detections that are true in reality while recall or producer's accuracy is the proportion of the true objects detected.

2.1   Generic Feature Extraction
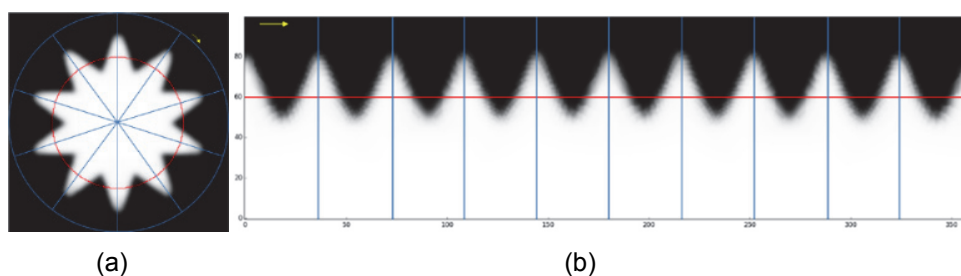

| (a) | (b) |

Fig. 3:   (a) Model of star-shaped object. (b) Polar shape matrix representation of model

Ideally, star-shaped objects (Figure 3a) are radially symmetric object having constant angular differences between consecutive leaves. If the star-shaped object is rotated about its center with the angle equivalent to the angle difference between two consecutive leaves, the resulting object is identical to the original non-rotated object.

In the polar shape matrix representation, the star-shaped object form triangular waveform like structure (Figure 3b). The polar shape matrix is calculated by sampling the object along the circumference of different concentric circles, centered at the center of the object, in a constant angular interval α (Equation 1).

$$\boldsymbol{Z_{r\theta}} = \boldsymbol{Z_{xy}}(x_m + \boldsymbol{r}(\cos\boldsymbol{\theta})^T, y_m + \boldsymbol{r}(sin\boldsymbol{\theta})^T) \qquad (1)$$

where $\boldsymbol{Z_{xy}}$ is the image, $\boldsymbol{\theta} = (0, \alpha, 2, \dots, 360 - \alpha)^T$, $\boldsymbol{r} = \left(0, \beta, 2\beta, \dots, \frac{d_m}{2}\right)^T$. d_m is minimum diameter that encloses the shape and x_m and y_m are the coordinates of the center of the shape.

In the case of star-shaped objects, the circular shift along the angular axis of the polar shape matrix results in the repetition of the same image when the shift is equivalent to the angular difference between leaves. The similarity decreases until it reaches the least when the shift is half the angular difference between leaves, and gradually increases to maximum on the other half. This scenario repeats periodically with the period equivalent to the number of leaves. In order to measure the similarity between original polar shape matrix and the circularly shifted version of it, correlation coefficients are used. Therefore, comes the name of the feature: circular autocorrelation of polar shape matrix (CAPS).

Circularly shifting along the angular axis and calculating the correlation for each shift takes long computation time. However, using Wiener-Khinchin Theorem (COHEN 1998) to compute it in Fourier domain makes the process efficient. The theorem is given by,

$$\boldsymbol{ACR} = real(F^{-1}\{F\{\boldsymbol{Z_{r\theta}}\}.F^*\{\boldsymbol{Z_{r\theta}}\}\}) \qquad (2)$$

where F is the Fourier Transform.

The output of the Wiener-Khinchin function is a circular correlation on both dimensions of the matrix, and the values depend on the distribution of gray level in the matrix i.e. overall lighting condition. As the range of the gray level distribution varies for different images, the output of the Wiener-Khinchin function also varies between images. This is standardized by converting the output to the Pearson correlation coefficient, where the values range from -1 for exact inverse correlation to +1 for exact positive correlation. The conversion can be achieved as following:

$$\boldsymbol{R} = \frac{\boldsymbol{ACR} - N\mu_z^2}{N\sigma_z^2} \qquad (3)$$

where N is the total number of elements in the polar shape matrix. μ_z and σ_z are mean and standard deviation of the polar shape matrix respectively.

As the interest of the study is only along angular shift, the vector with the values at no radial shift from **R** is considered. Furthermore, the circular autocorrelation function is symmetric, therefore, only half of this vector i.e. $\frac{360}{2\alpha}$ elements are enough to describe the object and is used as the feature vector.

CAPS, as shown in Figure 4, is a standardized feature vector where the elements range from -1 to +1 in value. The maximum value can appear multiple times for rotationally periodic objects while it only appears at no angular shift for objects that are not rotationally periodic. Circular autocorrelation by nature is unaffected by the linear shift. But, in the case of CAPS, the shift is

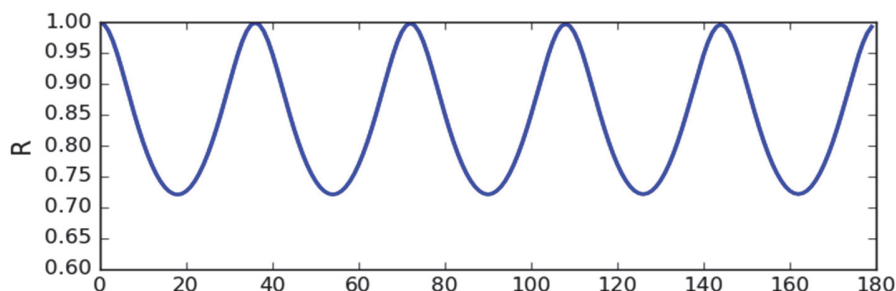carried out along the angular axis. Therefore, CAPS is unaffected by rotation.

Fig. 4: CAPS of the model

Besides the multiple maximum values and fluctuating curve, CAPS of star-shaped objects also have on average high values of correlation throughout all elements. Taking the mean, as the feature, can serve as the minimal representation of the shape. The average of CAPS elements is highest for the circular object. Therefore, considering it as the feature will be a measure of the object's similarity with the circular object. The average of the CAPS is given by

$$\mu_{caps} = \frac{1}{n}\sum CAPS \tag{4}$$

The randomness of CAPS is low due to its periodic nature. Entropy gives the measure of randomness (TSAI et. al 2008). Entropy is small for the less random signal and large for the more random signal. Therefore, the CAPS of star-shaped objects should have lower entropy compared to other objects. However, the entropy of CAPS of circular shape has the lowest, i.e. negative infinity. We consider entropy of CAPS as other feature of study. The entropy of a signal is computed with a probability of the signal. If $S_1, S_2, S_3, ..$ are the signal that occurs with probabilities $P(S_1), P(S_2), P(S_3), ...$ respectively then, the entropy H(**S**) of the signal **S** with Ns as the total number of unique signal occurrences is given by (GRAY 2011, TSAI et. al 2008):

$$H(\boldsymbol{S}) = -\sum_{t=1}^{N_s} P(S_i)\log P(S_i) \tag{5}$$

Both mean of CAPS (M-CAPS) and entropy of CAPS (E-CAPS) are rotation invariant and invariant to the overall illumination change. Besides that, both of them are a scalar value. However, for higher object likelihood, M-CAPS tends to be large while E-CAPS tends to be small.

2.2   Results and Discussion

The framework was implemented for detection of a naturally occurring star-shaped object, palm trees. It was applied to 8 test images from 5 cultivation regions in Indonesia, Malaysia, and Thailand. The images were acquired with Trimble UX5 unmanned air system (UAS). The ground sampling distances of image ranged from 3.2 cm to 10 cm. However, for the study, the pixel sizes were resampled to approximately 5cm. The images were projected to Universal Transverse Mercator projection system with WGS84 as an ellipsoid of reference.
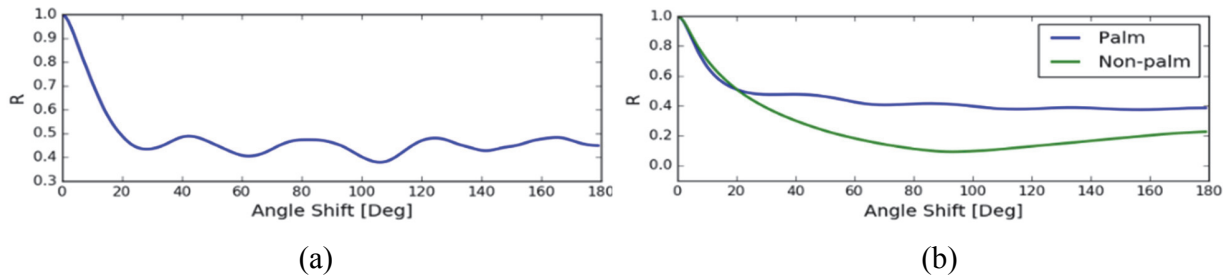
(a)   (b)

Fig. 5:   (a) CAPS of a palm tree training sample. (b) Mean of all elements across all training samples for palm trees and non-palm objects

2.3   Feature characteristics on training data

644 samples of positive and 597 samples of negative training data were randomly selected from one cultivation region only. The inspection of the palm tree canopies found that a window covering 201×201 pixels, in general, covers the entire canopy. Therefore, all the training samples were of the size 201×201 pixels. Same size of the window was also used in the moving window scheme
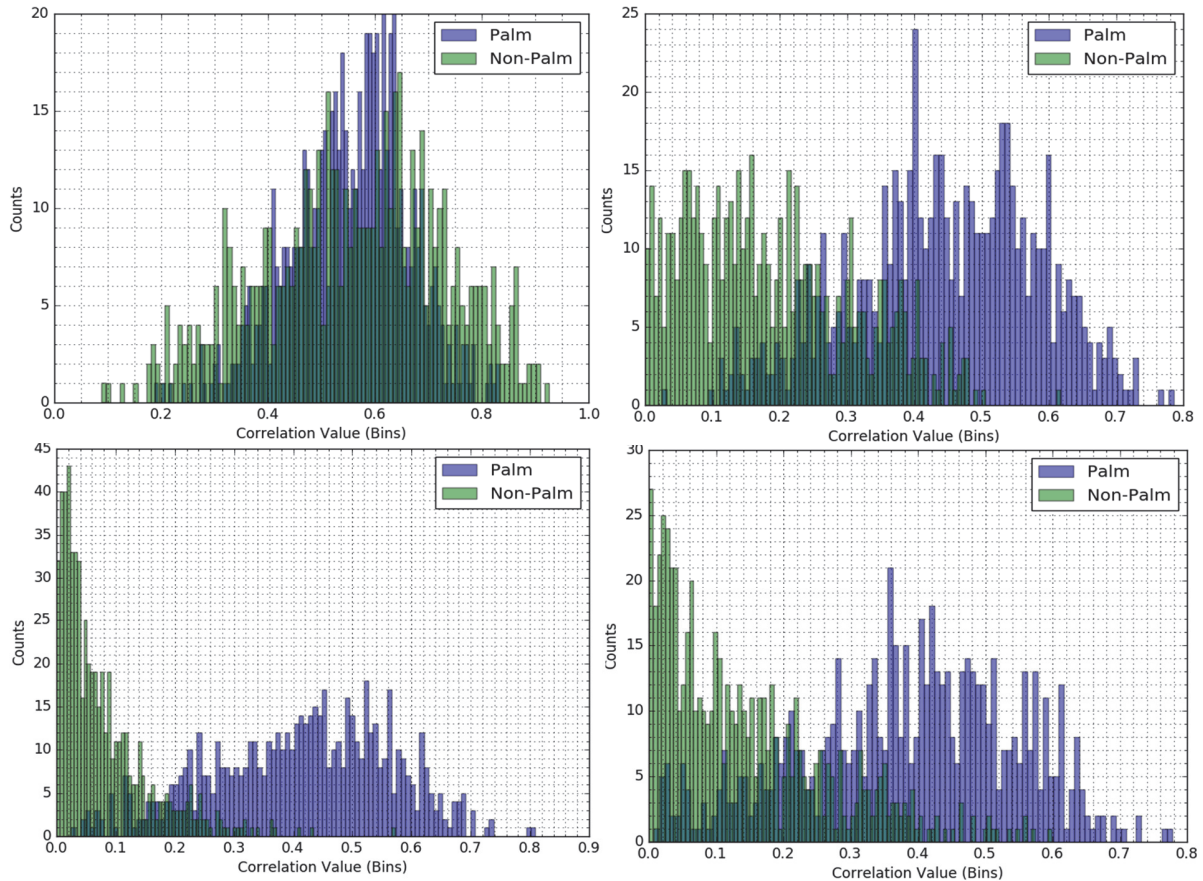
Fig. 6:   Histogram of few elements of CAPS calculate across all training samples.

for feature extraction. During the transformation of the sample to the polar shape matrix, $\alpha = 1px$ and $\beta = 1px$ were used, which resulted in CAPS with 180 elements.

As can be seen in Figure 5a, CAPS of the training samples are not a periodic curve implying that the palm trees are not perfectly star-shaped. Furthermore, the presence of background noise and non-palm objects also contribute to such a characteristics. However, the CAPS is still a fluctuating curve with relatively higher values at the position of leaves of palm trees in the polar shape matrix representation. This is, however, in general, not the case for the non-palm objects. The average of each element across all palm trees and non-palm object shows the significant elementwise
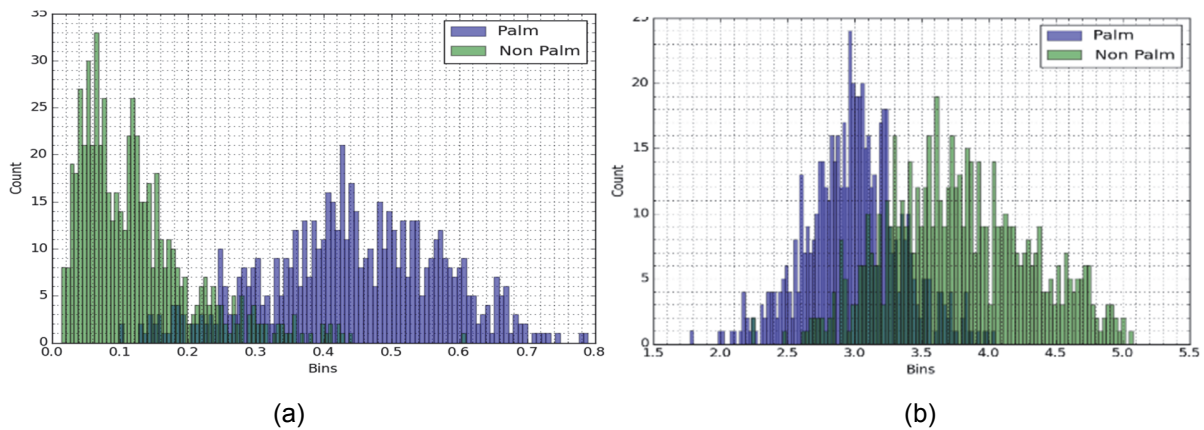
Fig. 7: (a) Histogram of M-CAPS across all training samples. (b) Histogram of E-CAPS across all training samples.

difference between the palm-trees and non-palm objects in Figure 5b.

Histograms of each CAPS elements derived from all positive and negative examples give the preliminary evaluation of the contribution of the element in distinguishing the palm trees and non-palm objects. From the histogram of elements calculate from all positive and negative training samples (Figure 6), it can be perceived that some elements play a significant role in distinguishing the palm trees and non-palm object while some contribute insignificantly, which is demonstrated by a large overlap between the distributions of the palm trees and the non-palm objects. As all elements of CAPS are used, a combined effect of all the elements can expect to be significant.

The histogram of M-CAPS, Figure 7a, shows small overlap between the palm trees and the non-palm objects. As expected, the means are greater for palm trees compared to non-palm objects. It signifies that the palm tree samples tends to be more circular compared to the non-palm training objects. On the contrary, E-CAPS (Figure 7b) have higher overlap. However, the values for palm trees are lower compared to non-palm objects as expected.

The combination of both M-CAPS and E-CAPS, Figure 8, shows highly separable clusters for palm trees and non-palm objects.
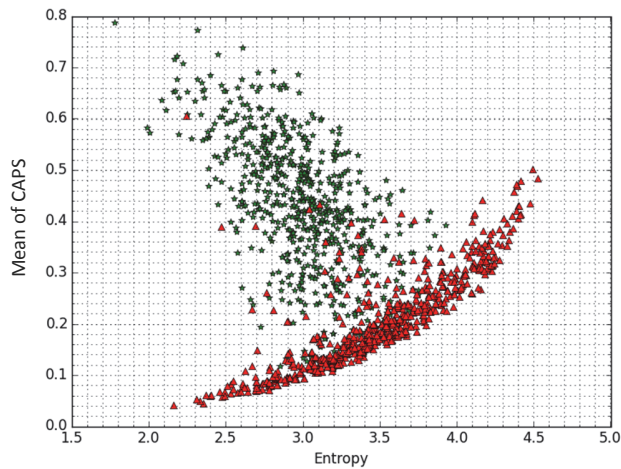
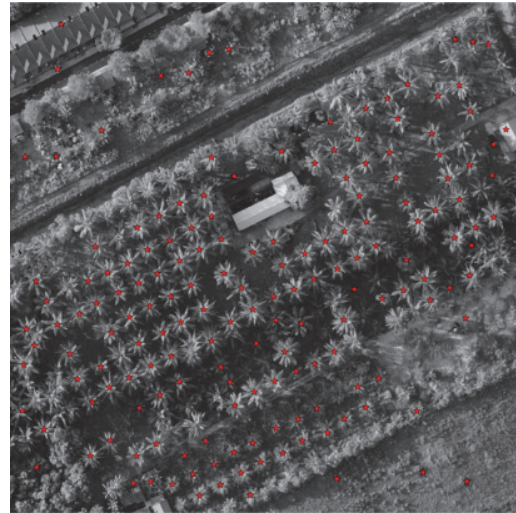Fig. 8:   Distribution of M- & E-CAPS across all training samples

Fig. 9:   Object map of Image 01 using CAPS

2.4   Implementation on test images

A moving window of 201×201 pixels was shifted along the rows and columns of the test images at an interval of 5 pixels on both sides, in order to achieve higher efficiency. This would, however, decreases the locational accuracy of detection.

At each window, the features were calculated and trained SVM was used to standardize the feature, which results in the feature maps as shown in Figure 10. The feature maps have peaks at the center of palm trees. But there are also multiple false peaks in non-palm objects, especially, in the artificial objects like houses and the roads. In the case of the feature map derived with M-CAPS, high values can be seen in straight lines like structures such as road and plantation line while in the case of E-CAPS, the extensive regions have high values. The combination of M-CAPS and E-CAPS shows a very clean feature map, however, it also consists the peaks on the houses. Using local maximum detection algorithm, the object map was derived from the feature map as in Figure 9. Figure 9 is the object map derived using CAPS as the feature. The detections are mostly at the centers of the palm trees. There are few false detections on the house and misses on the palm trees along the road. Similar object maps are derived from all test images and all features and the accuracy measures are calculated and are shown in Figure 11.
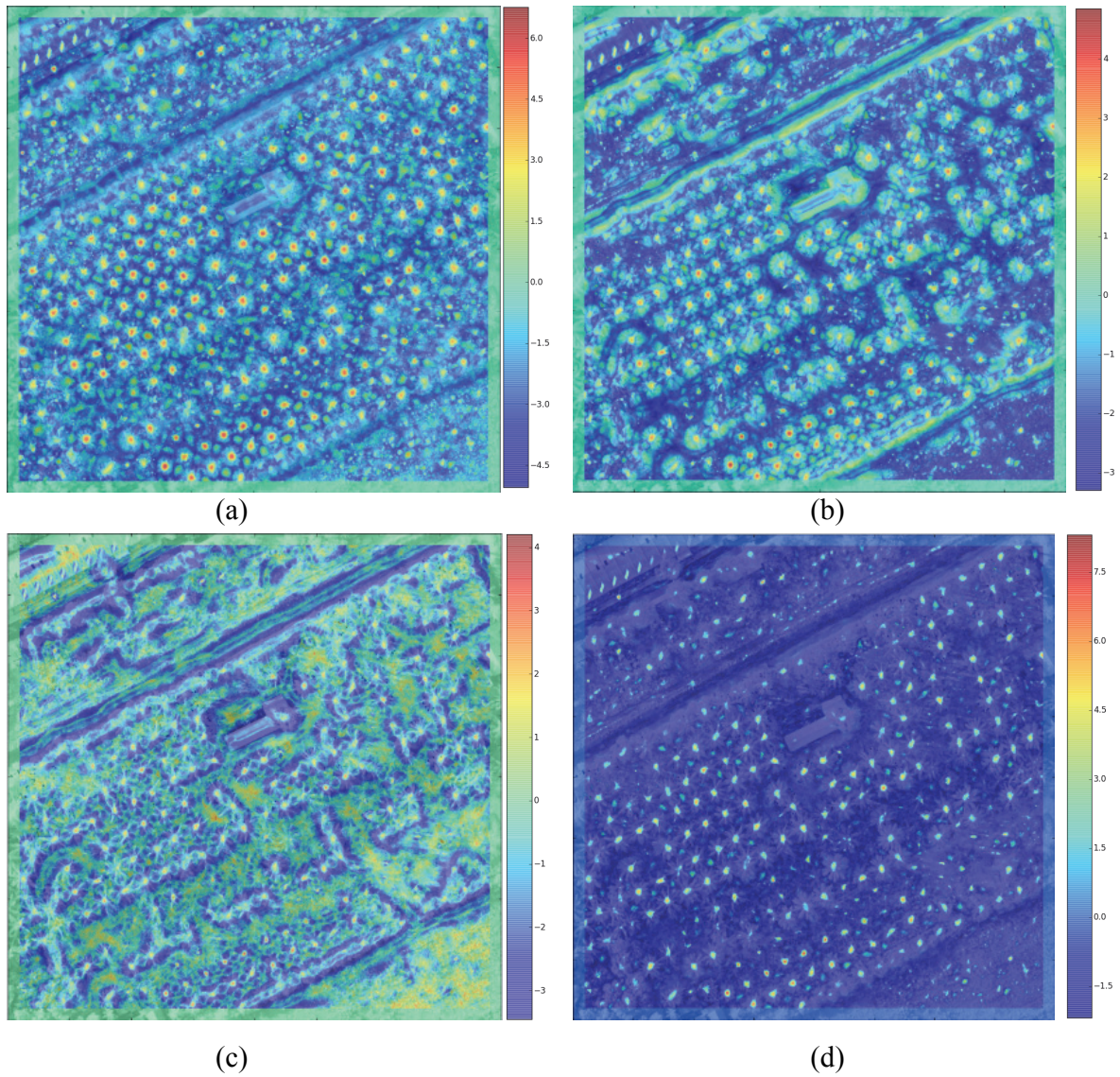
589

Fig. 10: Feature maps extracted from Image 01 using (a) CAPS (b) M-CAPS (c) E-CAPS (d) Combination of M- & E-CAPS
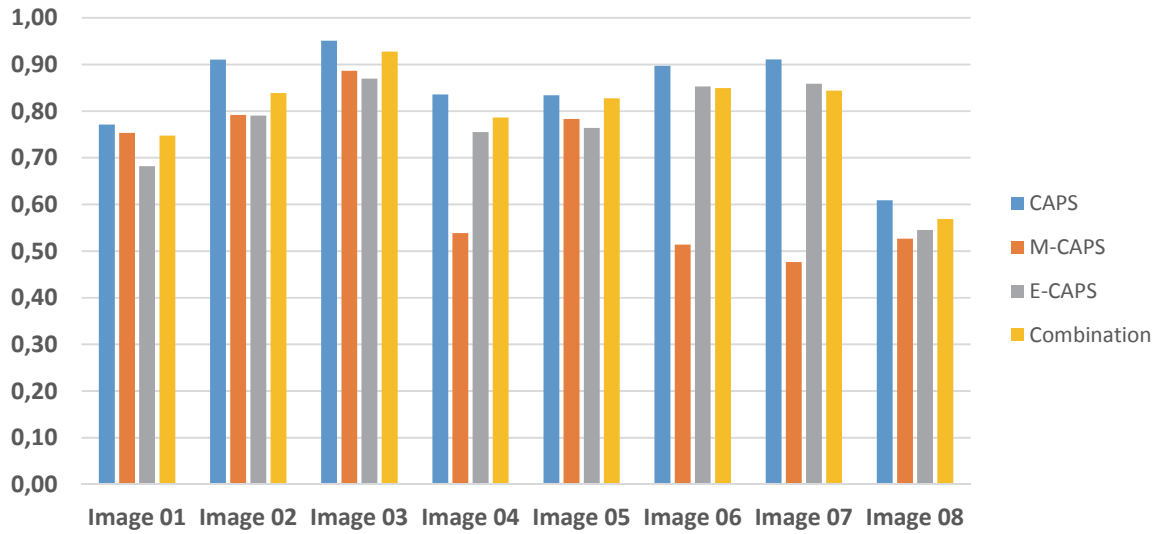
Fig. 11: Accuracies of each feature for each images

The detection using CAPS as the feature produced a good result in detecting palm trees. For six out of eight images, the results were achieved with an accuracy higher than 80% with the highest accuracy up to 95%. The detection using M-CAPS produced on average worst results. The results were especially poor for image 04, 06, and 07, where the palm trees were more resembling the star-shaped objects. The detection using E-CAPS produced better results than M-CAPS. However, the maximum accuracy, it was able to achieve, was 87% and only three out of eight images were detected with accuracy higher than 80%. The combination of M-CAPS and E-CAPS produced better result compared to M- and E-CAPS independently, with the maximum accuracy of 92%, and five out of eight images were detected with accuracy over 80%. The average accuracies of the detection using CAPS, M-CAPS, E-CAPS and the combination of M- & E-CAPS are, as shown in Figure 12, 84%, 66%, 76%, and 80% respectively.

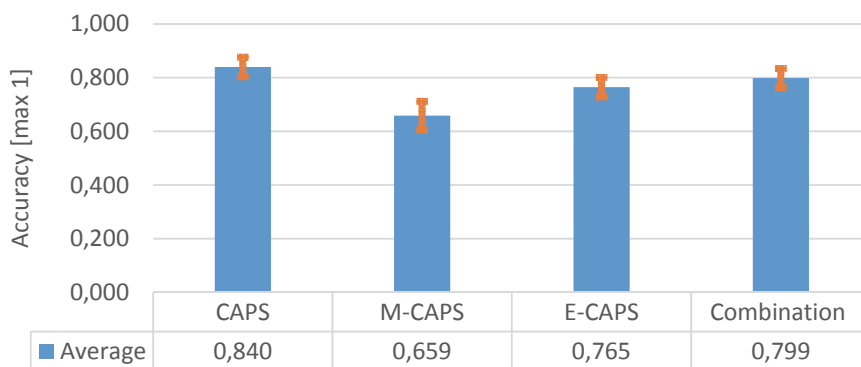|	CAPS	M-CAPS	E-CAPS	Combination
Average	0,840	0,659	0,765	0,799

Fig. 12: Average accuracies of each feature across all test images

In all the images, there were few false detections on the houses, roads, and trees with circular canopies, which is driven by the fact that the objects having regular shapes like square and circle also have high CAPS and thus corresponding M- and E-CAPS.

591

In general, the framework performed quite well in the scenes with multiple challenges like varying sizes, contrasts and cultivation densities. However, artificial objects like houses had several false detections. The false detections were generally at the corners and the centers of buildings. Such false detections can be eliminated using masks or filters with normalized difference vegetation indices. The detection of other regular shapes like square, circles etc. can be reduced by using a learning algorithm, which determines the non-linear decision boundary.

3   Acknowledgements

4   References

BEN-HUR, A. & WESTON, J., 2010: A user's guide to support vector machines. Data mining techniques for the life sciences, Springer, 223-239.

BISHOP, C.M., 2006: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.

CHANG, C.-C. & LIN, C.-J., 2011: Libsvm: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), **2** (3), 27.

COHEN, L., 1998. Generalization of the Wiener-Khinchin theorem. Signal Processing Letters, IEEE, 5(11), pp.292-294.

CORTES, C. & VAPNIK, V., 1995: Support-vector networks. Machine learning, **20** (3), 273-297.

GRAY, R.M., 2011: Entropy and information theory. Springer Science & Business Media.

LIU, X., GAO, C. & LI, P., 2012: A comparative analysis of support vector machines and extreme learning machines. Neural Networks, **33**, 58-66.

POTTS, J., LYNCH, M., WILKINGS, A., HUPPE, G., CUNNINGHAM, M. & VOORA, V., 2014: The state of sustainability initiatives review 2014: Standards and the green economy. International Institute for Sustainable Development (IISD) and the International Institute for Environment and Development (IIED), 332.

SUYKENS, J.A., & VANDEWALLE, J., 1999: Least squares support vector machine classifiers. Neural processing letters, **9** (3), 293-300.

TSAI, D.Y., LEE, Y. & MATSUYAMA, E., 2008: Information entropy measure for evaluation of image quality. Journal of digital imaging, **21** (3), 338-347.

VAPNIK, V.N. & VAPNIK, V. 1998: Statistical learning theory, Volume 1. Wiley, New York.