Spatial Distribution of Honeybee Forage based on Color Satellite Image Segmenting using K-Mean Clustering

Rachid Sammouda[1], Ameur Touir[1], Fahman Saeed[1], Nuru Mohammed[2], Ahmed Al-Ghammidi[2]
[1]Department of Computer Science
[2]Department of Science of Food and Agriculture
[1, 2]King Saud University
Riyadh, Saudi Arabia
{rsammouda, fsaeed, nurmohammed}@ksu.edu.sa, touir@ccis.ksu.edu.sa, aalkhazim@gmail.com

ABSTRACT: *Beekeeping plays an important role in increasing and diversifying the incomes of many rural communities in Kingdom of Saudi Arabia. However, despite the region's relatively good rainfall, which result in better forage conditions, bees and beekeepers are greatly affected by seasonal shortages of bee forage. Because of these shortages, beekeepers must continually move their colonies in search of better forage. The aim of this paper is to determine the actual bee forage areas with specific characters like population density, ecological distribution, flowering phenology based on colour satellite image segmentation using K-mean clustering. K-mean segment region satellite images into five segments, following that we search in a sample of acacia trees against this image clusters to specify the best region for better forage.*

Keywords: Beekeeping, Forage, Satellite Images, Segmentation

Received: 11 September 2012, Revised 3 October 2012, Accepted 10 October 2013

1. Introduction

Apiculture is one of the most important economic activities for rural communities in Saudi Arabia, while approximately 5,000 beekeepers maintain more than million honeybee colonies and produce approximately 900,000 tons of honey annually. In most cases, success in beekeeping depends on the availability of sufficient bee forage in terms of both quality and quantity of nectar and pollen grains. Beekeeping is more dependent on the existing environmental conditions of an area compared with livestock production practices. Unlike other livestock production, it is not possible to keep bees by carrying feeds from other areas, therefore bee forage is considered to be one of the most important elements in the beekeeping industry. In many areas in the Kingdom, bees and beekeepers suffer from seasonal drought, which causes a shortage of bee forage. These conditions drive many beekeepers to move their colonies from one area to another in search of better nectar and pollen sources. Beekeepers are concentrated in few areas in search of special species of plants that provide the most desired and expensive types of honey. Foraging areas have remained approximately stable in contrast to the constant increase in the number of bee colonies reaching fourfold during the last decade from 270,000 to more than one million. The result of which is an overpopulation in bee colonies zones against available foraging areas. Nevertheless, and despite the area's relatively good availability of bee forage, because of seasonal shortages, beekeepers are forced to continually move in search of new bee forage in order to maintain their bee colonies and to obtain a greater honey harvest. The movement of the beekeepers is not systematic or guided with respect to the carrying capacity of the resources and phenology of major bee plants. This leads to serious competition for forage and the subsequent declining of productivity of beekeeping in the region. It is very important to identify and characterize the bee forage of the area in terms of species diversity; population density; ecological distribution; and quality and quantity of nectar and pollen produced by the plants to guide beekeepers.

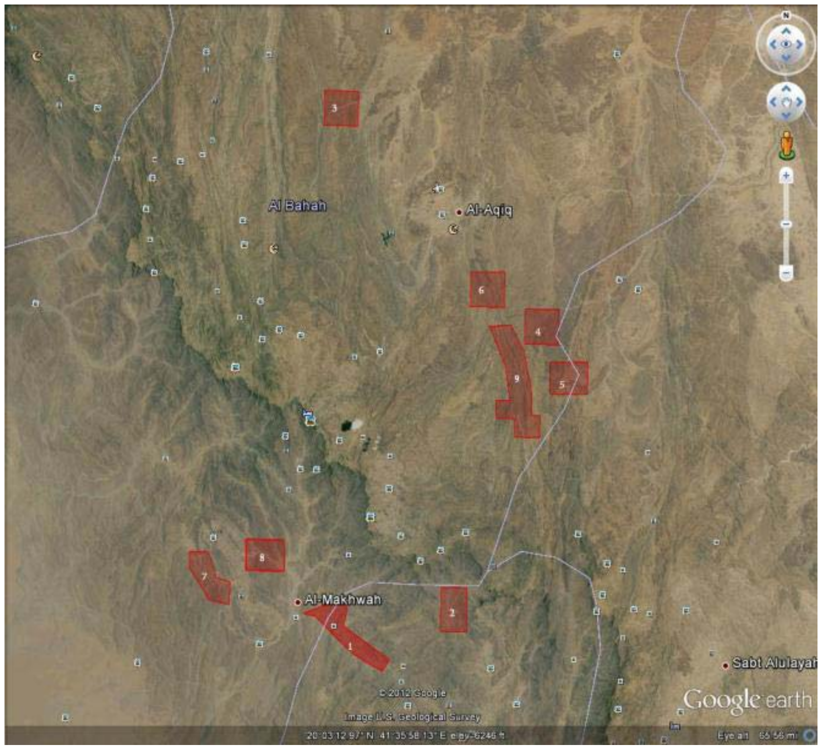


Figure 1. Google map: localization of the valleys used in this study

It is essential to inventory the honeybee populations of the region and determine the optimum bee colony carrying capacity of the different valley areas of Al baha region as an example. Fig.1 shows a Google map view of the valleys that will be used in our studies. This is used to define a development strategy for beekeeping in the Kingdom of Saudi Arabia using satellite image segmentation. In this paper we present an approach to determine the actual presences of a special type of trees like, the acacia tree, in satellite images of Al Baha region. First we present the segmentation technique used in this study, then the analysis process of the segmented images.

2. K-Mean Segmentation Algorithm

Image segmentation is the process of image division into regions with similar attributes [1]. It is an important step in image analysis chain with applications to pattern recognition, object detection, etc. Until recently, most of the segmentation methods and approaches were supervised such as Maximum Posteriori (MAP) [2] or Maximum Likelihood (ML) [3] with an average efficiency rate of about 85% [4, 5]. In the supervised methods a priori knowledge is needed to get a successful segmentation process and sometimes the required earlier information may not be available.

Satellite images are an important source of information which is used in many environmental assessments and monitoring of agriculture, meteorology...etc. When considering the early days of 1972, when Landsat Multi Spectral Scanner System (MSS) was launched, current satellite image systems exhibit extraordinary diversity. There are operational satellite systems that sample all available parts of the electromagnetic spectrum with spatial resolution from 0.61 to 1000 m. Satellite image segmentation continues to be an important area in various research fields, which led to the publication of several methods in this domain.

K-Means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space, and tries to find natural clustering in them. The points are clustered around centroids $\mu_i \; \forall i = 1... k$ which are obtained by minimizing the objective

$$V = \sum_{i-1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2 \tag{1}$$

where there are k clusters S_i, $i = 1, 2,..., k$ and μ_i is the centroid or mean point of all the points $x_j \in S_i$. As a part of this project, an iterative version of the algorithm was implemented. The algorithm takes a 2-dimensional image as input. Various steps in the algorithm are as follows:

1. Compute the intensity distribution (also called the histogram) of the intensities.
2. Initialize the centroids with k random intensities.
3. Repeat the following steps until the cluster labels of the image do not change anymore.
4. Cluster the points based on distance of their intensities from the centroid intensities.

$$c^{(i)} := \arg\min_j \| x^{(i)} - \mu_j \|^2 \tag{2}$$

5. Compute the new centroid for each of the clusters.

$$\mu_i := \frac{\sum_{i=1}^{m} 1\{c_{(i)} = j\}\, x^{(i)}}{\sum_{i=1}^{m} 1\{c_{(i)} = j\}} \tag{3}$$

Figure 2. Part of a valley in the region of study

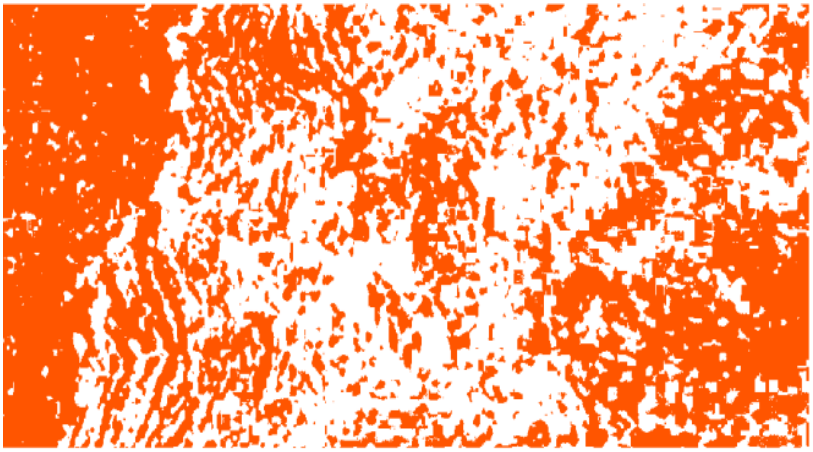


Figure 3. Segmentation result of the sample in Figure 2 using K-mean with $k = 2$

where k is a parameter of the algorithm (the number of clusters to be found), i iterates over all the intensities, j iterates over all the centroids and μ_i are the centroid intensities.

We applied the above algorithm to satellite color images with varying value of k as shown below in the result section.

3. Segmentation Results

We implemented the K-means clustering algorithm to satellite color images. The segmentation results were generated on different images; we present here a case of them.

Figure 2 shows the input image to the K-mean clustering algorithm, which is represented in the computer by its three components in the RGB color space.

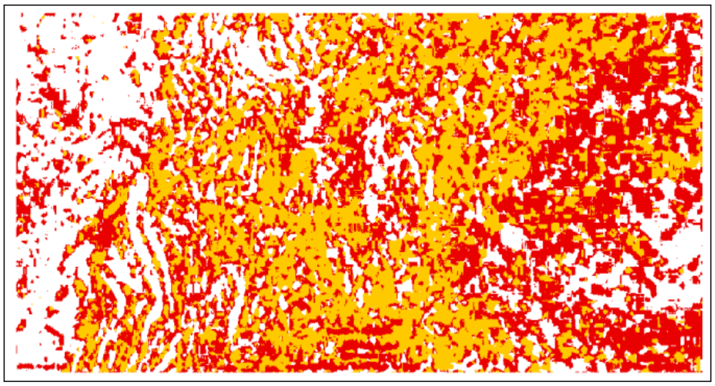


Figure 4. Segmentation result of the sample in Figure 2 using K-mean with $k = 3$

Figure 5. Picture of an Acacia tree taken from a valley of study in Al Baha region in the Kingdom of Saudi Arabia

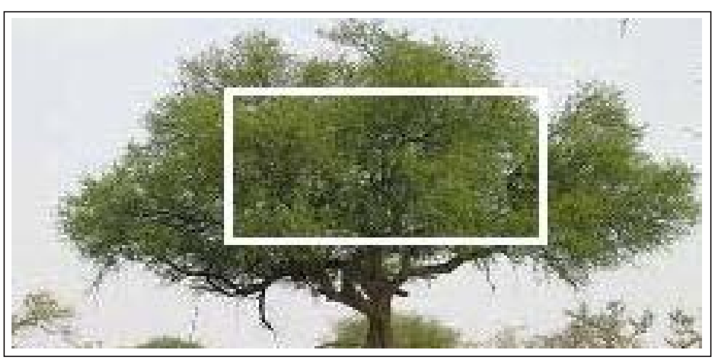
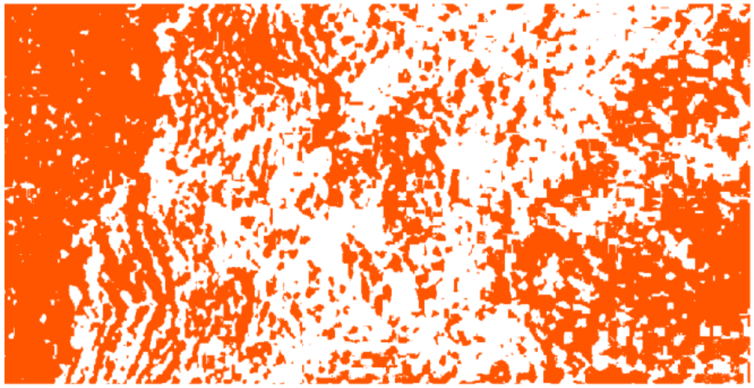


Figure 6. Cropped areas from acacia trees used to calculate a color vector representing the type of the tree



A. Segmentation result of the sample in Figure 2 using K-mean with $k = 2$

B. First cluster in Figure (A) represented with its raw values in the satellite color image

C. Regions containing acacia trees in the regions represented by the First cluster of Figure B

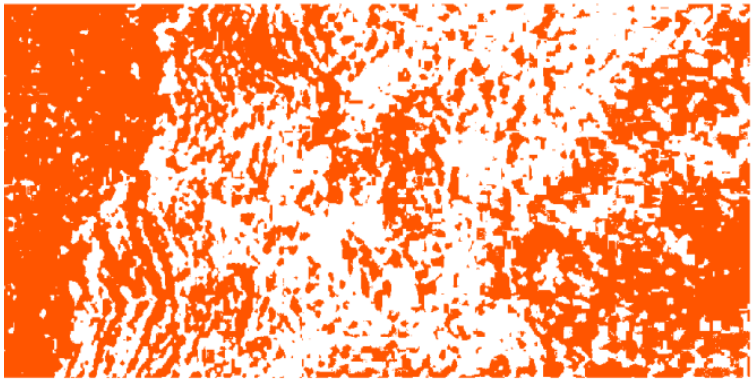
Table 1. Segmentation Results I

We have used different k-values in order to see the accuracy of the algorithm and the sensitivity of the method.

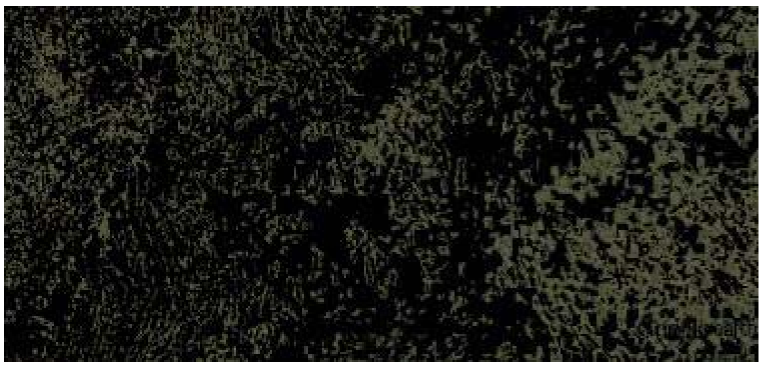
For every pixel in the input image, K-mean returns an index corresponding to a cluster. At the last step this method uses the pixel label to separate object in image based on the color mean value of the corresponding cluster [1]. Figure 3 and Figure 4 show the segmentation results of the satellite color image in Figure 2 obtained using K-mean classifier with $k = 2$ and $k = 3$, respectively.

4. Clusters Analysis

The objective of this research is to determine the actual bee forage areas with specific characteristics like population density, ecological distribution, flowering phenology. Here we start by determining the regions of interests (ROIs), with respect to the



D. Segmentation result of the sample in Figure 2 using K-mean with $k = 2$



E. Second cluster in Figure (A) represented with its raw values in the satellite color image

C. Regions containing acacia trees in the regions represented by the Second cluster of Figure E

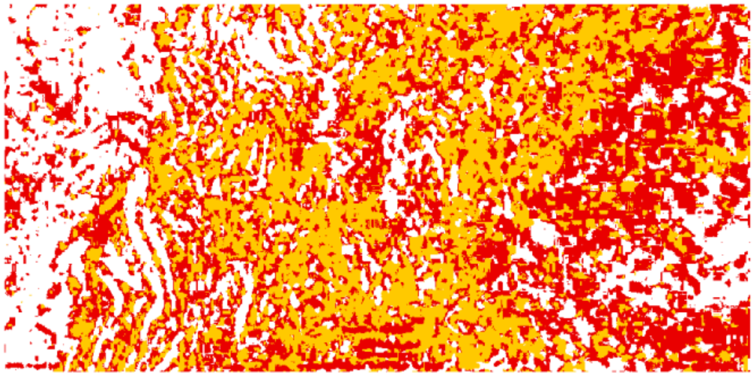
Table 2. Segmentation Results II

beekeepers, based on the type of trees the clusters contains, like **acacia** tree shown in Figure 5.

The clusters analysis will be based on the mean value of each cluster, in the segmented image, and the average color value that may be reflected by acacia trees in the regions of study. For the later value, we have collected different pictures of acacia tree like the one shown in Figure 5, crop part of it, shown Figure 6, and calculate the average of each color component in the cropped area.

Table1 shows the analysis result of the first cluster in the segmented image shown in Figure 3 reflecting the distribution of acacia trees.

Table 2 shows the analysis result of the second cluster in Figure 3, reflecting the distribution of acacia trees.



G. Segmentation result of the sample in Figure 2 using K-mean with $k = 2$

H. Third cluster in Figure (A) represented with its raw values in the satellite color image

I. Regions containing acacia trees in the regions represented by the Third cluster of Figure H

Table 3. Segmentation Results III

Table 3 shows the analysis result of each segmented image reflecting the distribution of acacia trees in the Second cluster (k = 2).

As can be seen from the pictures (*F*) in Table 2 and the picture (*I*) in Table 3, the candidate regions to be visited for acacia trees are almost similar in the two different segmented images (k = 2, and k = 3). These results will be used to more and fine analysis. In a more extended paper we will give more analysis of the situation, as we will have the chance to use high resolution multispectral satellite images of Al Baha region.

5. Conclusion

In this paper we presented segmentation result of Al Baha region satellite images obtained using K-mean clustering algorithm with two, and three, clusters along RGB channels data. The analysis of the segmented images produces candidate regions of the AL Baha area in the Kingdom of Saudi Arabia that contains Acacia trees. This result will help beekeepers to be well-guided about the actual bee forage areas with specific characteristics like population density, ecological distribution and flowering phenology.

6. Acknowledgments

This work is supported by the National Plan for Science and Technology, King Saud University, Riyadh, Saudi Arabia under project number 11-AGR1750-02.

References

[1] CHITADE, A. (2010). Colour based image segmentation using k-means clustering, *International Journal of Engineering Science and Technology*, 2 (10) 5319-5325.

[2] Awad, M. (2010). An Unsupervised Artificial Neural Network Method For Satellite Image Segmentation, *The International Arab Journal Of Information Technology*, 7 (2).

[3] Benediktsson, J., Swain, P., Ersoy, O., Hong, D. (1990). Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing Data, *Computer Journal of Institute of Electrical and Electronics Engineers*, 28 (4) 540-551.

[4] Perkins, S., Theiler, J., Brumby, S., Harvey, N., Porter, R., Szymansk, J., Bloch, J. (2000).GENIE: A Hybrid Genetic Algorithm for Feature Classification in Multi Spectral Images, *In*: Proceedings of SPIE Applications and Science of Neural Networks, *Fuzzy Systems and Evolutionary Computation* III 4120, USA, p. 52-62.

[5] Zhang, P., Verma, B., Kumar, K. (2003). Neural vs Statistical Classifier in Conjunction with Genetic Algorithm Feature Selection in Digital Mammography, *In*: Proceedings of IEEE Congress on Evolutionary Computation CEC, Australia, p. 634-639.