

# Fusion Approach for Dates Fruit Classification

Khaled Marji Alresheedi

College of Computer  
Qassim University, KSA

## ABSTRACT

Depending on the type of the date fruit, dates have many features that should help in recognizing and classifying a date. The main features of the date fruit are the color, texture and the shape of date fruit. This article thus analyzes image based date fruit classification and recognition. In this regards, this paper has two contributions, firstly, as there is no standard dataset available, a dataset of 9 date's classes is constructed. This dataset presents an interesting challenge for computer vision algorithms. Secondly, Gabor features and Color Layout features are used. These features are then fused to increase the classification performance. The proposed Color Layout and Gabor approach achieves an acceptable performance of more than 88% correct detections.

## General Terms

Classification, Computer Vision,

## Keywords

Date fruit recognition, Dates classification, Gabor filtering, Color layout features

## 1. INTRODUCTION

The Kingdom of Saudi Arabia is producing more than 450 different varieties of date's fruit. An estimated 25.1 million trees (and increasing) are available for this lucrative business [1]. An analysis done in 2012 shows that the date's production increased from 350 Kilo tons to about 1.03 Million Tons. This increase is equivalent to three folds. One of the main factors is the increase in cultivated areas, government support and application of new technologies and practices in dates farming [1]. Sukkari date being the most famous variety in Al-Qassim province, the Khalas in Al-Hasa region and the Ajwa is the pride of Al-Madinah Al-Munawara region and general KSA.

Middle East dates are famous for their good taste. Date fruit is not only delicious but contains many nutritional components good for children and adults alike. The fresh dates provide rich amount of the vitamin C. However, the dried dates and vacuumed dates lose this useful vitamin. But the other nutrients are still available after processing dates. The dates contain plenty of Sugar, Fiber, Iron, Calcium, Carbohydrates, Potassium, Thiamine, Riboflavin, Niacin, Folate, Vitamin A and Vitamin K [2]. The dates lack the fat or cholesterol and are therefore useful for many individual as diseases.

As there are almost 450 different types of dates. Interestingly, every type has some features that are different than others. In recent years, not only other fruits [3][4], but also the date business, automated industries, the sorting, separating and arranging of dates has inspired many research dimensions [5][6]. This process can be extremely tedious and requires many man hours, wasting human resources for the huge number of dates to be processed in food and medicine industries. In this paper, the interest is in the domain of the automatic detection, recognition, and sorting of dates. The narrow research direction is thus the image based autonomous

detection of different types of dates. This information can thus be integrated with the robotic actuators/hand for removal/sorting of the required dates type.

Haidar et al. [5] classified dates using linear discriminant analysis. 140 images of dates were used belonging to seven classes, namely, Sukkery, Mabroom, Khedri, Safawi Al Madina, Madina Ajwa, Amber Al Madina, and Safree. Images had a size of 480 X 640 and were in JPEG format. They were able to get a top accuracy of 99% using ANN. Similarly, Zawbaa et al. [7] classified fruits using random forest algorithm. However, they used fruits such as apples, oranges and strawberry. They used features such as shape, color along with Scale Invariant Feature Transform (SIFT). They got better results with random forest as compared to KNN and SVM. Furthermore, higher accuracy was achieved while classifying apples than oranges.

## 2. DATASET

As there is no standard dataset available, a dataset of date fruit is constructed consisting of 9 classes. This dataset presents an interesting challenge for computer vision algorithms. These dates types include:

- Ajwa
- Barhy
- Hasheshy
- Khalas
- Majdool
- Munefee
- Rothanah
- Sugee
- Wannana

Figure 1 shows the examples of some of the samples from the dataset of 9 classes of the dates. Each class contains 30 dates. As the dates are similar in the same class, it is assumed that the 30 images are enough for one class. Figure 2 shows the setup for the dataset generation. The Nikon D7100 DSL camera is used for the dataset generation. Following are some of the important specifications of D7100 [8].

- 24.1 Mega pixels
- ISO 100-6400 standard, up to 25600 expanded
- Maximum six frames per second.
- Fifty-one-point Auto Focus systems system
- RGB metering sensor
- Spot white balance
- 1080p recording
- magnification 0.94-X

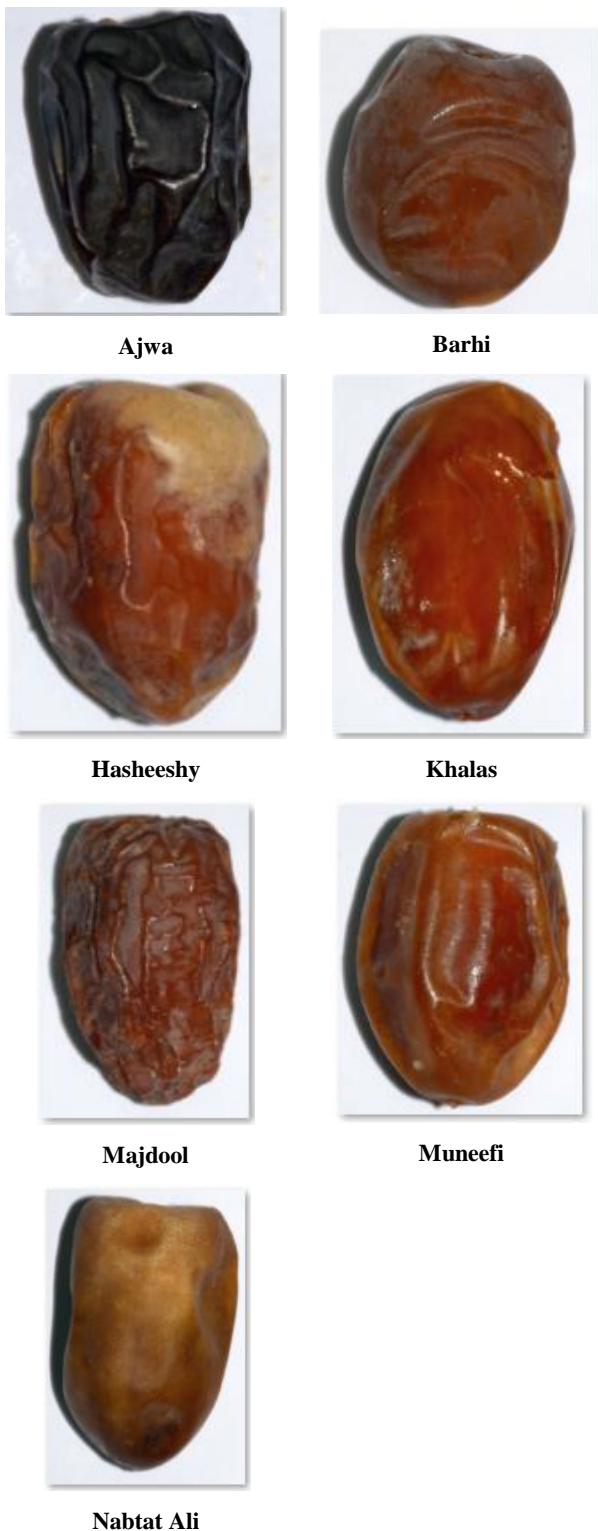


Figure 1 Sample dates from the dataset.



Figure 2: Setup for dataset generation.

### 3. METHODOLOGY

Figure 3 shows the methodology of the dates classification and recognition approach. The images of the dates are selected for analysis. The features of the date's images are extracted using the feature extraction approaches. These approaches are Colour Layout Features and the Gabor approach. Since all the features are not important, the feature selection helps in selecting important features. After the feature extraction and selection, the features are fused. The machine learning is used to learn the features.

The machine learning set includes the Bayesian network, Support Vector Machines (SVM), Random Forest, and the Multi-Layered Perceptron (MLP).

The colour layout features include:

*Dominant Colour* - it comprises of the dominant colors, percentage, variance and the spatial based coherency.

*Colour Structure* - global and the local spatial colour structures in image

*Scalable Colour* - is a Haar transformation applied on the values of a colour histogram

*Homogeneous Texture* - Quantitative texture

*Edge Histogram* - the spatial distribution of edges which represent local distribution in the image.

*Contour-Based Shape* - the contours of the objects

The Gabor filters finds many applications in image processing and computer vision, and are useful in feature extraction for texture analysis and segmentation. Gabor filters are a set of filters that can be tuned to different frequencies and orientations. The filter set is helpful in computer vision and used primarily for feature extraction from an image. The 2D Gabor filters are represented by the Equation 1 and 2:

$$G_c[i, j] = B e^{-\frac{(i^2+j^2)}{2\sigma^2}} \cos(2\pi f(i \cos \theta + j \sin \theta))$$

- (1)

$$G_s[i, j] = C e^{-\frac{(i^2+j^2)}{2\sigma^2}} \cos(2\pi f(i \cos \theta + j \sin \theta))$$

- (2)

In Equation 1 and 2, the symbols B and C represent the normalizing factors.  $f$  is the target frequency being searched in an image. The parameters  $\theta$  and the  $\sigma$  represent the size of the image region being analysed with varying frequencies.

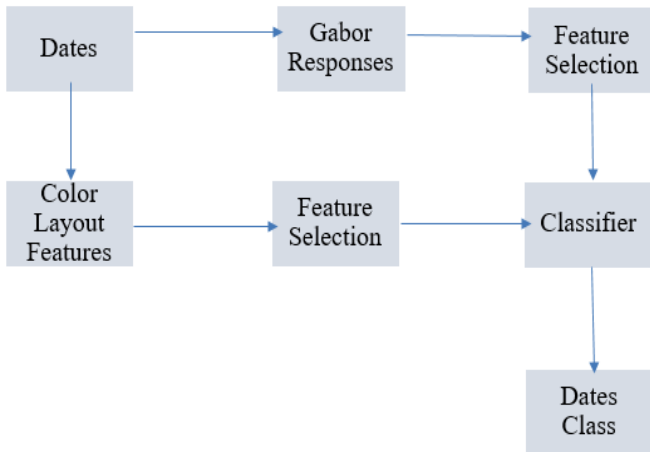


Figure 3: Dates classification based on fusion of features

#### 4. EXPERIMENTAL RESULTS

The colour layout and the Gabor features are learnt by the four machine learning algorithms. The machine learning set includes the Bayesian network, Support Vector Machines (SVM), Random Forest, and the Multi-Layered Perceptron (MLP). These are selected based on their good overall performance in the state of the art.

Bayesian Networks (BN) is a representation of the variables and the relationship between the variables using Direct Acyclic Graph (DAG) [9]. BN uses the probability distribution and thus known as the parametric classifier [9]. The SVMs are generally called the Support Vector Networks [10]. SVMs are used for classification and learning features in computer vision. The SVM in training phase develops a model that maximizes the separating distance between the classes.

Leo Breiman [11] introduced the Random forest, and takes advantage of the combination of many trees from the given training dataset. A Multilayer perceptron (MLP) is a feed forward neural network that functions by mapping the input variable of a dataset onto the output labels [12].

For the performance evaluation, we use the four evaluation parameters. These are the Accuracy, Precision, Recall, and the F-measure.

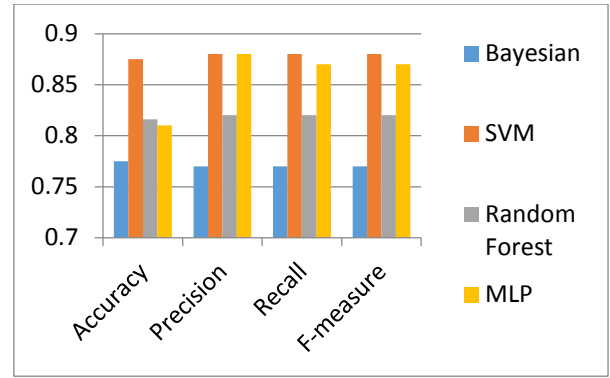
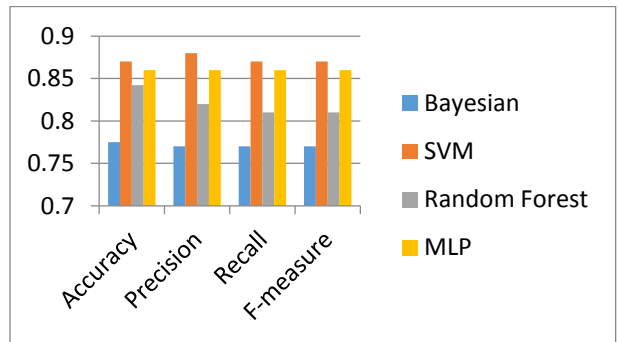
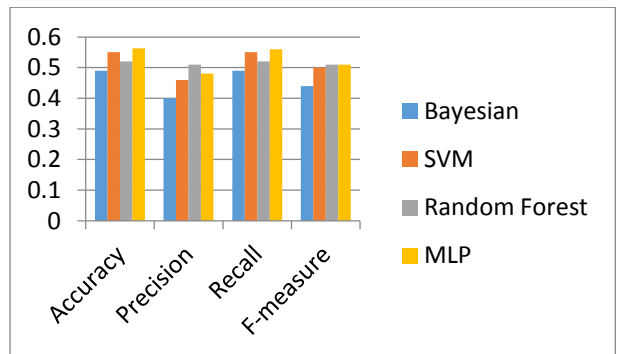


Figure 4: The performance of the fusion of the color layout features and the Gabor filter.

Figure 4 shows the performance of the fusion approach. The Figure 5 shows the performance analysis of the colour layout features and the Gabor filter. Figure 5(a) shows the Colour Layout features and the Figure 5(b) shows the performance of the Gabor features.



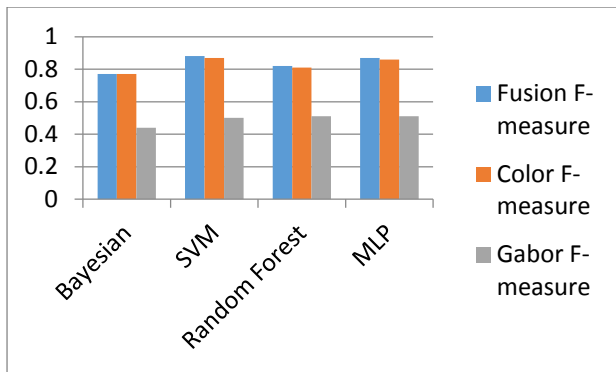
(a)



(b)

Figure 5: Figure (a) shows the color layout result and Figure (b) shows the Gabor filter results.

The Figure 5 is presented here as a reference for comparison with the Figure 4. In Figure 5, the Gabor has lower performance compared to the Colour Layout features. The Figure 6 shows the summarized form of the comparison of the fusion approach, the color layout features and the Gabor filters approach. Figure 6 shows the F-measure of the Fusion approach, the Colour Layout approach, and the Gabor approach.



**Figure 6: Our fusion of the two features (color layout and the Gabor filters) gives increase classification performance in all the cases.**

In the Figure 6, based on the F-measure, the fusion approach outperforms the Gabor approach in almost all the cases with high margin. In the case of the Bayesian network, the fusion approach has an F-measure of 0.77 which is same as the F-measure of the color layout feature. In all other cases, the F-measure of the fusion approach is higher than both the color layout feature and the Gabor feature. The highest F-measure of 0.88 is obtained for the SVM. From this analysis, we can infer that the model of the date's detection is accurate for more than 88%, and thus delivers an acceptable performance. The analysis also shows that the Gabor has overall lower performance compared to the Colour Layout features. Since the Gabor covers only the filtered responses, the performance is not as good as the Colour Layout features. The Colour Layout features not only take colour distribution into account, but considers Contour based Shape features. Therefore, we get good overall performance for Colour Layout features compared to the Gabor features. The fusion further enhances this performance achieving more than 88% correct detections.

## 5. CONCLUSION

This paper discussed dates classification using the two features of color layout and the Gabor filtering. In this regards, the paper achieved two objectives and provides two contributions. The paper introduced a dataset of 9 classes of dates fruit. Secondly, the approach uses the Gabor features and the color layout features. These features are then fused to increase the classification performance. For the performance evaluation, four parameters are used. These are the Accuracy, Precision, Recall, and the F-measure. The proposed approach with the SVM achieved an F-measure of 0.88. This means that the model provides an acceptable accuracy of more than 88%. In future, we aim to increase this classification accuracy by deep learning.

## 6. ACKNOWLEDGMENTS

We are thankful to the Dean (CoC) and the Head of the CS Department for all the help during the MS Thesis.

## 7. REFERENCES

- [1] S. H. A. A. M. A.-A. S.M. Ismaiel, "Marketing volume transaction analysis of dates in Saudi Arabia." 2016.
- [2] M. A. Al-Farsi and C. Y. Lee, "Nutritional and Functional Properties of Dates: A Review," *Crit. Rev. Food Sci. Nutr.*, vol. 48, no. 10, pp. 877–887, Oct. 2008.
- [3] Y. Zhang, P. Phillips, S. Wang, G. Ji, J. Yang, and J. Wu, "Fruit classification by biogeography-based optimization and feedforward neural network," *Expert Syst.*, vol. 33, no. 3, pp. 239–253, 2016.
- [4] Y. Zhang, S. Wang, G. Ji, and P. Phillips, "Fruit classification using computer vision and feedforward neural network," *J. Food Eng.*, vol. 143, pp. 167–177, 2014.
- [5] A. Haidar, H. Dong, and N. Mavridis, "Image-based date fruit classification," in *International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*, 2012, pp. 357–363.
- [6] G. Muhammad, "Date fruits classification using texture descriptors and shape-size features," *Eng. Appl. Artif. Intell.*, vol. 37, pp. 361–367, 2015.
- [7] H. M. Zawbaa, M. Hazman, M. Abbass, and A. E. Hassanien, "Automatic fruit classification using random forest algorithm," in *2014 14th International Conference on Hybrid Intelligent Systems, HIS 2014*, 2014, pp. 164–168.
- [8] "Nikon D7100 In-Depth Review: Digital Photography Review."
- [9] K. Hedman and S. Hinz, "The application and potential of Bayesian network fusion for automatic cartographic mapping," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012, pp. 6848–6851.
- [10] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [11] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [12] C. M. Bishop and C. M., *Neural networks for pattern recognition*. Clarendon Press, 1995.