AN AUTOMATED SYSTEM FOR DATE FRUIT RECOGNITION THROUGH IMAGES

AIADI O. , KHALDI B. and KHERFI M. L.

Université Kasdi Mebah. Ouargla, Fac, Information Technology and communication

Abstract:Date is a fruit with a great health and economic benefits. However, farmers plant a little number of varieties and too little are known by people. Therefore, there is an urgent need to preserve such an important cultural heritage for the next generations.In this paper, we present an automated system for date fruit recognition from their images. Specifically, we collect fifty (50) samples from seven (7) varieties, and then we take images for those samples. Afterwards, and in order to identify the visual characteristics of samples belonging to each variety, we extract shape and color features from the images.Then, we use the Support Vector Machine (SVM) classifier to optimally separate the visual characteristics of the different varieties. Later on, SVM is used to decide for a test sample the variety it belongs to. Our system presents a multitude of advantages: 1) it is able to accurately recognize dates in spite of the large variation within some varieties (intra-variation) and the small variation between some varieties (inter-variation); 2) no physical measurements are needed, and only visual characteristics of sample images are sufficient; and 3) it doesn't require any human intervention. Experimental results, carried out on the samples we collected, show a high recognition rate of 97.14%.

Keywords: date fruit, image, recognition, Support Vector Machine (SVM).

UN SYSTEME AUTOMATIQUE POUR LA RECONNAISSANCE DES DATTES A PARTIR DES IMAGES

Résumé : Les dattes sont des fruits qui ont des bienfaits pour la santé et l'économie. Cependant, les agriculteurs plantent un nombre limité de variétés et peu d'entre elles sont connus par le grand public. Par conséquent, il existe un besoin urgent de préserver un tel patrimoine culturel important pour les prochaines générations. Dans cet article, nous présentons un système automatique pour la reconnaissance des dattes à partir de leurs images. Plus précisément, nous recueillons cinquante (50) échantillons provenant de sept (7) variétés, puis nous prenons l'image de ces échantillons. Ensuite, et dans le but d'identifier les caractéristiques visuelles des échantillons appartenant à chaque variété, nous extrayons les caractéristiques de forme et de la couleur. Nous utilisons les séparateurs à vastes marges (SVM) pour séparer d'une façon optimale les caractéristiques visuelles des différentes variétés. Plus tard, SVM est utilisé pour décider à quelle variété un échantillon d'essai appartient. Notre système présente plusieurs avantages: 1) il est en mesure de connaître avec précision les dattes en dépit de la grande variation au sein de certaines variétés (intra-variation) et la faible variation entre certaines variétés (inter-variation); 2) aucune mesure physiques est nécessaire, seules les caractéristiques visuelles des échantillons d'images suffisent; et 3) il ne nécessite aucune intervention humaine. Les résultats expérimentaux effectués, sur les échantillons que nous avons recueillies, montre un taux de reconnaissance élevé de 97,14%.

Mot-clés: dattes, image, reconnaissance, Séparateurs à vastes marges (SVM)

Auteur correspondant AIADI O. email : o_aiadi@hotmail.com

Introduction

Date palm is growing in many countries including Americas, southern Europe, Asia, Africa, and Australia [1], where the top producers are the Middle East and North Africacountries [2]. Date fruit is consumed at their different maturity levels and served in different manner,suchas,with water or milk, as cakes or integrated within other meals.Date is a highly appreciated fruit due to its sweet taste and the interesting health benefits it provides [3].Beside the significant amount of vitamins, mineralsit contains [4], date fruit is a rich source of carbohydrates (about 70%) [2].

Farmsusually produce more than one variety, according to [8], thereexist more than 900 date varieties,and eachhas specific characteristics. They differ in terms of color, size, shape, texture as well as flavor and price (i.e., inter-variation).However, it is sometimes very difficult to distinguish one variety from another because of the high visual resemblance between them.In addition, dates within the same variety can differ in terms of maturity level, hardness degree and shape (i.e., intra-variation).

Unfortunately, the nouns of some varieties become unknown from a large portion of people. Besides, some of those varietieshave started to effectively disappear from markets andthereby fromthe collective people memory. Therefore, development of an automated system for date recognitioncouldsignificantly helps inpreservingsuch an important cultural heritage for the coming generations.

In the literature, too little attention has been devoted to date recognition from images i.e., specifying for a given datethe variety it belongs to using the date image.For instance, color, shape and texture descriptors were extracted so as to classify dates into seven varieties [5].In [6], used color properties in RGB space with probabilistic neural networks (PNN) to classify five date varieties.However**,** it should be noticed that all the previous studies are limited on recognizing fully mature date samples.

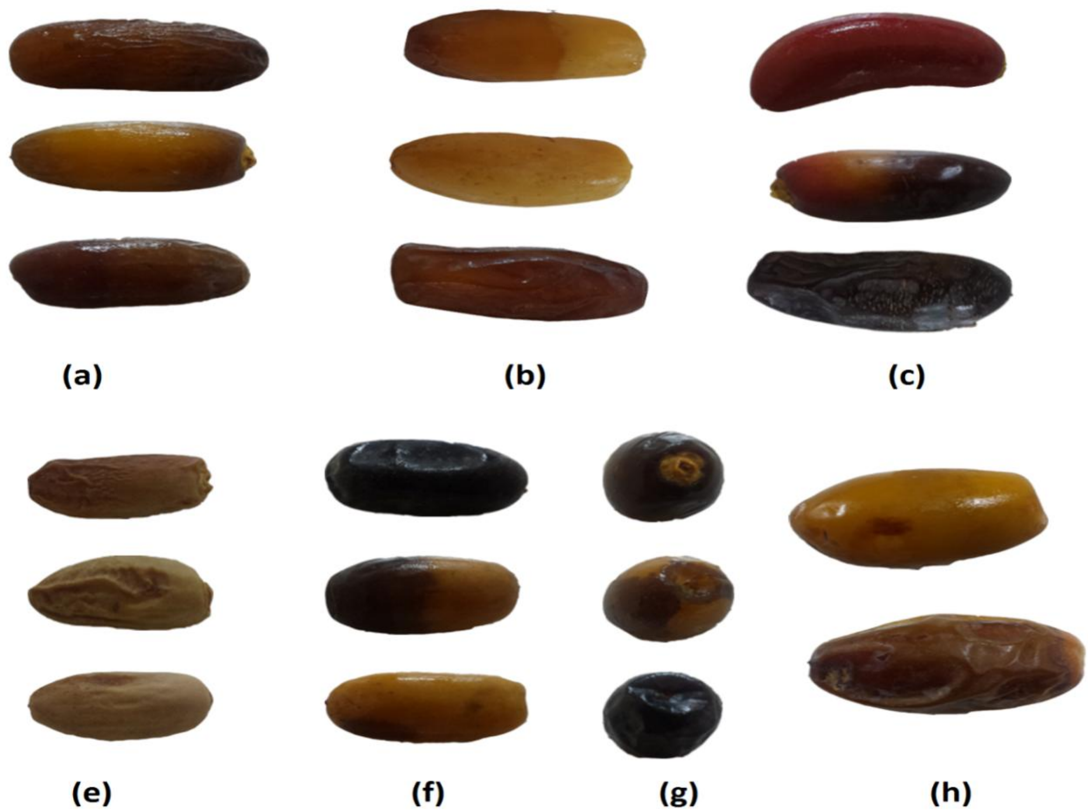

Figure 1: *Representative samples from each variety in our collection(a) Bouaarous; (b) Tafezwin; (c) Hamraya; (e) Degla bayda; (f) Tinisin; (j) Tantbucht and (h) Ajina;*

Example of a large intra-variation includes (b) (c) (f). Example of a small intra-variation includes (e).

In this paper, we propose an automatic system that recognizes dates from their images.Specifically, we collect date samples from seven (7) varieties, where dates in some of them are in different maturity stages.We take images for all the samples we collected,and then we split them into two parts. Images belong to the first part are called training images and are used to learn the visual characteristics of each variety. Whereas, images of the second part are intended for testing purposes.In order to recognize date samples, we conduct a learning process which starts by extracting shape and color features from the training images of each variety. Those features are then used with the Support Vector Machine (SVM) classifier to recognize for each test samples the variety it belongs to. Our system is capable to recognize dates through their images, and without the need for any physical measurements. In addition, our systemis able to recognize dates though the intra-variation, in terms of maturity level, within some varieties. Experimental results show that our system achieved a high recognition rate of 97.14%.

1. Material and Methods

We group the images related to each variety separately, and we try then to learn the visual characteristics that characterize each variety. The detailed steps of the learning process are given in the sub-sections below:

1.1 Samples Collection

Date samples were purchased from the local markets of Touggourt region (south Algeria), as it is amongst the well-known regions that produce distinct date varieties. Our collection is made up of 350 samples from 7 varieties namely, Ajina, Bouaarous, Deglabayda, Hamraya,Tafezwin,Tantbucht and Tinisin. It is worth noting that dates in some varieties are in different maturity levels. Fig. 1 shows representative samples from each variety.

1.2 Image Acquisition

We take an image for each of the date samples using a camera at resolution of 4128*3096, resulting in a total of 350 (50 per variety) images. Those images are then splitted into two parts; the first includes 280 images (40 per variety), it is used to learn the visual characteristics of each variety. The second part assembles 70 images (10 per variety), it is intended for testing purposes.

1.3 Features Extraction

1.3.1 Color Features

The samples, wecollected, have different colors, and thus color features can be used for recognition purposes. In order to represent color distribution within image, standard deviation and mean of three RGB channels were calculated and used as features.

The mean is given by:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} V_i, \qquad (1)$$

Where$N$ is the number of pixels in the image, $V_i$is the value of the pixel in one channel (either R or G or B).

The standard deviation is given by:

$$S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(V_i - \mu)^2}, \qquad (2)$$

1.3.2 Shape Features

Shape is also a quite essential feature because some varieties differ significantly in terms of shape. Four geometrical features that describe date fruit shape were extracted,

- Area: It is the number of pixels forming the date fruit.
- Major axis length: It is the longest diameter within the date fruit shape.

- Minor axis length: Is the shortest diameter within the date fruit shape. It is orthogonal to the major axis.
- Eccentricity: It is the ratio of the major axis length to the minor axis length.

Note that images were binarized before extracting features.

1.4 Classification of New Date samples

In order to be able to recognize new date samples, we conduct a process composed of two stages:

- Learning stage: as mentioned above, we use 280 images (40 per variety) from the whole collection for the learning purposes. To understand how this process is carried out, we give a simple example: suppose that we want to learn the visual characteristics of "Ajina" variety:
- First, we assign the set of images related to "Ajina" variety with the label "Ajina".At machine level, this is interpreted as: "images in which dates have an area of 300 to 500 pixels (shape) and have a brown color (color) represents something called 'Ajina' ". With the same manner, we continue to present the remaining varieties to the machine.
- Second, we use the SVM classifier [7] to put the optimal boundaries that permit to separate the visual characteristics of the different varieties.
- Recognition: this stage consists in predicting, for a given test date sample, the variety it belongs to, based on the knowledge learned in the previous step. To do so, we assign each variety with a score that represents the probability that the sample belongs to it. Finally, the sample is assigned to the variety having obtained the high score.

2. Experimental Results

In orderto assess the performance of our system, we conduct two experiments. The first experiment was intended to measure the overall recognition rate yielded by our system.The aim of the second experiment is to test the strength of each of shape and color features separately.

2.1First Experiment

As mentioned in the previous section, we devote 70 images from the whole dataset to test the performance of our system. Figure2shows the recognition rate yielded by each variety.

In spite of the considerable diversity, in terms of visual characteristics, within somevarieties and the large visual similarity between some varieties, our system has achieved an overall recognition rate of 97.14%. In addition, we notice that in 5 from the 7varieties, our system have achieved a recognition rate of 100%. These varieties are respectively, Ajina, DeglaBayda, Bouaarous,Tantbucht and Tinisin.For Ajina and Tantbuchtvarieties, this rate may be due to the specific shape of these two varieties. Nevertheless, our system has achieved a recognition rate of (90%) in both Hamraya and Tafezwin.

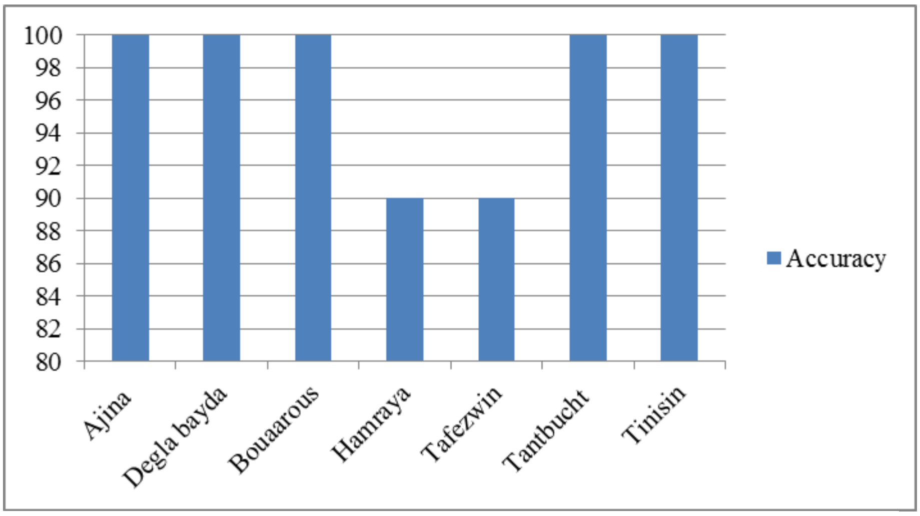

Figure 2: Recognition rate yielded by each variety

2.2Second experiment

This experiment is devoted to investigate the recognition rate that can be achieved by each of shape and color features separately**.**Table 1 shows the accuracy over all varieties yielded by each feature separately and by combining them.

From Table 1, we notice that shape feature has yielded better recognition rate than color features. Shape features have outperformed color features because of the visual resemblance, in terms of color, between several samples that belong to distinct varieties. However, by combining the two features, we gain an overall recognition rate of 97.14%.

Table 1accuracy over all the varieties yielded by each combination

Method	**Accuracy**
Color	90 %
Shape	94.28 %
Combination	97.14 %

Conclusion

Year after year, some date varieties are became unavailable in markets and disappear from the daily discussions of people. Hence, significant efforts have to be made in order to preserve such an important cultural heritage. In this paper, we have presented an automated system for date fruit recognition from their images. We have collected 350 samples from 7 varieties (50 samples per variety), and then we took image for those samples. In order to recognize date samples, we have conducted a process of learning which started by extracting shape and color features from the images. Then, we used the Support Vector Machine (SVM) to learn the visual characteristics of the different varieties, and to recognize new samples. Our system has successfully dealt with the large variation within some varieties (intra-variation) and the small variation between some varieties (inter-variation). Our system can easily be integrated in the industry, as it doesn't require any human intervention. Experimental results have demonstrated the strength of our system and a high recognition rate of 97.14% has been achieved.

References

[1] **AL-YAHYAI R. and MANICKAVASAGAN A.**: An overview of date palm production. 2012 Ed.CRC Press, 3-12.

[2] **AL-FARSI M.,ALASALVAR C.,MORRIS A., BARON M. and SHAHIDI F.** : Compositional and sensory characteristics of three native sun-dried date (*Phoenix dactylifera* L.) varieties grown in Oman. Journal of Agriculture and Food Chemistry ; 2005 ; 53(19). 6586-7591.

[3] **AMANAT A., WALY M., ESSA M. and DEVARAJAN S.**:Nutritional and medicinal value of date fruit. 2012 ; Ed. CRC Press, 361-376.

[4] **AL-HOOTI S., SIDHUJ.S. and QABAZARD H.**:Physicochemical characteristics of five date fruit cultivars grown in the United Arab Emirates.Plant Foods for Human Nutrition ; 1997 ; 50(2).101-113.

[5] **HAIDAR A., HAIWEI D. and MAVRIDIS N.**: Image-based date fruit classification. The 4th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT). 2012 ; 357-363.

[6] **FEDAL M.**:Date fruits classification using probabilistic neural networks.International Commission of Agricultural Engineering. 2007 ; 9.1–6.

[7] **VAPNIK V.**: The Nature of Statistical Learning Theory. Springer ; 1995.

[8] **HANACHIS. and KHITRI D.**: Inventaire variétal de la palmeraie Algérienne.Actes du symposium sur la datte. 1998 ; 44-190.