Texture Features and Decision Trees based Vegetables Classification

Suresha M
Dept. Computer Science
Kuvempu University

Sandeep Kumar K S
Dept. of Computer Science
Kuvempu University

Shiva Kumar G
Dept. of Computer Science
Kuvempu University

ABSTRACT

The proposed work deals with an approach to perform texture extraction of vegetables images for classification. The work has been carried out using watershed for segmentation. The vegetables textures features like red component, green component, skewness, kurtosis, variance, and energy are extracted. The method has been employed to normalize vegetable images and hence eliminating the effects of orientation using image resize technique with proper scaling. Finally, Decision Tree classifier is applied to the above features which return the results of the classification.

General Terms

Image Processing, Pattern Recognition.

Keywords

Decision Tree Classifier, Texture Features, Vegetables Classification.

1. INTRODUCTION

Vegetable means an edible plant or part of a plant, but usually excludes seeds and most sweet fruit. This typically means the leaf, stem, or root of a plant. Some vegetables can be consumed raw, some may be eaten cooked, and some must be cooked in order to be edible. Vegetables are most often cooked in savory or salty dishes. However, a few vegetables are often used in desserts and other sweet dishes, such as rhubarb pie and carrot cake. A number of processed food items available on the market contain vegetable ingredients and can be referred to as "vegetable derived" products. These products may or may not maintain the nutritional integrity of the vegetable used to produce them. Examples of vegetable-derived products are ketchup, tomato sauce, and vegetable oils. The green color of leaf vegetables is due to the presence of the green pigment chlorophyll. Chlorophyll is affected by the pH, and it changes to olive green in acid conditions, and to bright green in alkaline conditions. Some of the acids are released in steam during cooking, particularly if cooked without a cover. The yellow/orange colors of fruits and vegetables are due to the presence of carotenoids, which are also affected by normal cooking processes or changes in pH. The red/blue coloring of some fruits and vegetables (e.g. blackberries and red cabbage) are due to anthocyanins, which are sensitive to changes in pH. When the pH is neutral, the pigments are purple, when acidic, red, and when alkaline, blue. These pigments are quite water-soluble. Eating vegetables regularly in our diet can have many health benefits. Vegetables are one of the most natural foods and contain different vitamins, minerals and thousands of other plant chemicals known to provide health benefits. Eating vegetables can reduce many diseases. Vegetables have generous amounts of vitamins. Vitamins regulate metabolism and vitamins are also used to convert the fats and carbohydrates into energy. Vegetable juices that taste "strong" such as spinach and beet are high in compounds that should be consumed in small quantities. Dilute these with milder tasting juices such as carrot, celery, or apple juice. Green vegetables include varying amounts of phytochemicals such as lutein and indoles, which interest researchers because of their potential antioxidant, health-promoting benefits. Wonderful source of vitamin A are green vegetables, in the form of carotene. The most important source of vitamin A are carrot juice and hot chili peppers, which both includes over 10,000 International Units per hundred grams. Chlorella as a vegetable includes vitamin A content of nearly 50,000 I.U. per hundred grams. Chlorella, which is a green, single-celled edible alga, is also high in chlorophyll, vitamin B12 and nucleic acids, which are sometimes known as "youth factors". Vegetable classification system plays a vital role in Big Bazaars, Super Markets, etc., to separate the incoming vegetable stock.

2. LITERATURE SURVEY

The paper describes an automated visual inspection system (AVIS) for quality control of preserved orange segments, widely applicable to production processes of preserved fruits and vegetables. Main constraints concerning these kinds of inspection applications are addressed: the need of online operation together with a strong requirement of economic profitability. The strong commitment of above circumstances has forced the development of a flexible and low cost AVIS architecture. The data volume to be processed has forced up the development of sophisticated control architecture for high-speed machine vision applications. Special effort has been put in the design of the defect detection algorithms to reach two main objectives: accurate feature extraction and online capabilities, both considering robustness and low processing time. These goals have been achieved combining a local analysis together with data interpretation based on syntactical analysis, which has allowed avoiding morphological analysis. An online implementation to inspect up to ten orange segments by second is reported, Fernandez, C et al. [5]. This paper introduces and discusses several techniques of analyzing the color of produces, based on image processing. An automatic video-inspection system was conceived, designed and realized, and it was employed for testing these techniques. The details regarding its design and the conclusions resulted from these experiments are presented here, as well, Buzera M. et al. [3]. Intensive fruit and vegetable sorting is a common task in productive regions. In order to meet the market standards, produce is classified according to quality levels that depend on maturity degree, weight, size, density, skin defects, etc. Probably the most important of these tasks involve automatic visual inspection. A distributed and scalable system for sorting automation is presented, that addresses all aspects of the quality classification mentioned above. The main characteristics of the system are: it can control from 1 to 10 conveyor belts; its maximum performance is 15 fruits per second per belt; apart from weight sensors, it combines infrared, color and ultraviolet images; some fruit defect modules are available to take account of the presence of a given defect, Pla, F [8]. This paper presents a scalable decision tree algorithm for classifying large dataset with high processing speed, which requires only one scan over the dataset. The algorithm addresses the scalability issue and requires a pass over the dataset in each level of decision tree construction. The proposed algorithm significantly reduces the IO cost and also requires one time sorting for numerical attributes which leads to a better performance in time dimension. According to the experimental results, the algorithm acquires less execution and also adoptable for any attribute selection method by which the accuracy of decision tree is improved, Thangaparvathi B, et al [12]. Drug discoverers need to predict the functions of proteins which are responsible for various diseases in human body. The proposed method is to use priority based packages of SDFs (Sequence Derived Features) so that decision tree may be created by their depth exploration rather than exclusion. This research work develops a new decision tree induction technique in which uncertainty measure is used for best attribute selection. The model creates better decision tree in terms of depth than the existing C4.5 technique. The tree with greater depth ensures more number of tests before functional class assignment and thus results in more accurate predictions than the existing prediction technique. For the same test data, the percentage accuracy of the new HPF (human protein function) predictor is 72% and that of the existing prediction technique is 44%, Singh M., et al [11]. The two most widely used implementations for decision trees are Leo Breiman CART [7] and Ross Quinlan C4.5 [10]. Traditionally, decision trees are used in the common supervised classification setting where the goal is to find an accurate mapping from instance space to label space. The standard performance metric for the supervised ranking setting is the area under the ROC curve, or AUC for short Andrew Bradley [1]. A decision tree, trained in the usual way as a classifier, can be used for ranking by scoring an instance in terms of the frequency of positive examples found in the leaf to which it is assigned. A few papers provide experiments showing that unpruned trees lead to better rankings than standard pruned trees Foster Provost et al. [6], Cesar Ferri et al. [4]. Decision tees have been widely and successfully used in machine learning. Fuzzy representations have been combined with decision trees. The authors propose a comparative study of pruned decision trees and fuzzy decision trees. Further, for continuous inputs, they explore different ways: (1) for selecting the granularity of the fuzzy input variables; and (2) for defining the membership functions of the fuzzy input values. The results show that a fuzzy decision tree constructed using FID3, combined with fuzzy clustering and cluster, is superior to pruned decision trees Benbrahim H[2]

3. SEGMENTATION

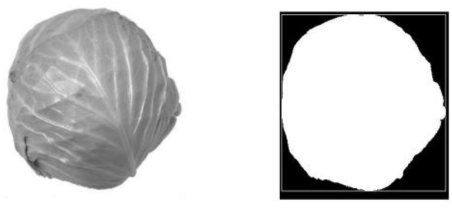
Image segmentation is a process that partitions an image into its constituent regions or objects. Effective segmentation of complex images is one of the most difficult tasks in image processing. Various image segmentation algorithms have been proposed to achieve efficient and accurate results. Among these algorithms, watershed segmentation is a particularly attractive method. The major idea of watershed segmentation is based on the concept of topographic representation of image intensity. Meanwhile, Watershed segmentation also embodies other principal image segmentation methods including discontinuity detection, thresholding and region processing. Because of these factors, watershed segmentation displays more effectiveness and stableness than other segmentation algorithms Rafael C. Gonzalez et al.[9]. Watershed segmentation is an effective method for gray level vegetable image segmentation. To apply watershed segmentation to binary images, we need to preprocess the vegetable binary images with distance transform to convert it to gray level images which are suitable for watershed segmentation. The common Distance Transforms (DTs) include Euclidean, City block and Chessboard. Different DTs produce very different watershed segmentation results for the vegetable binary images. For vegetable images containing components of different shapes, we find that the Chessboard DT can achieve better watershed segmentation results than Euclidean DT and City block DT.

(a) Original image (b) segmented image

Figure 1: Segmentation results for gray beetroot image

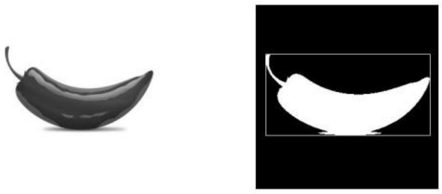
(a) Original image (b) segmented image

Figure 2: Segmentation results for gray carrot image



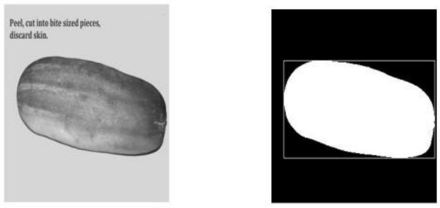
(a) Original image (b) segmented image

Figure 3: Segmentation results for gray cabbage image



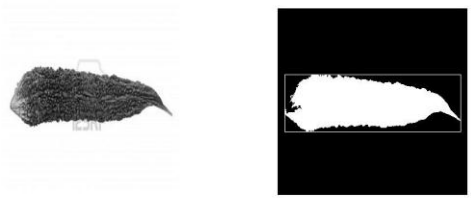
(a) Original image (b) segmented image

Figure 4: Segmentation results for gray chilies image



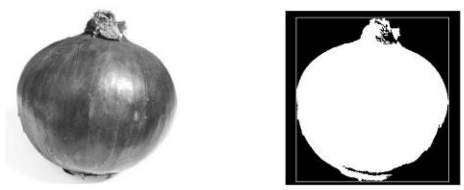
(a) Original image (b) segmented image

Figure 5: Segmentation results for gray cucumber image



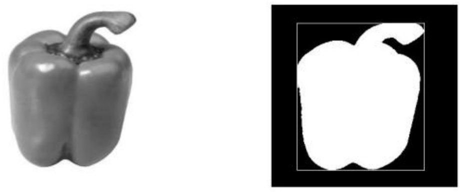
(a) Original image (b) segmented image

Figure 6: Segmentation results for gray bittermelon image



(a) Original image (b) segmented image

Figure 7: Segmentation results for gray onion image



(a) Original image (b) segmented image

Figure 8: Segmentation results for gray capsicum image

4. FEATURE EXTRACTION

In the classification of fruits and vegetables, color is one of the most important parameter that allows for the evaluation of their degree of maturity and freshness, and existence of faults. Also, color along with its level of homogeneity influences the degree of acceptance of consumers, as well as the pricing. Qualitative sorting is usually performed by trained inspectors. This type of evaluation is rather expensive and is determined by operators' inconsistency and subjectivity. Machine vision technology offers objective solutions for all these problems and it is considered to be a promise for replacing the traditional human inspection methods. Color components are one of the features for classification in this work. In this work there are five more features of vegetables like skewness, kurtosis, variance and energy are considered for classification.

Red Component:

The average value of the red component of an RGB image is given by,

$$\mu_R = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} f_R(i,j) \qquad (1)$$

Where f_R is the red component of an RGB image.

Green Component:

The average value of the green component of an RGB image is given by,

$$\mu_G = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} f_G(i,j) \qquad (2)$$

Where f_G is the green component of an RGB image.

Skewness:

Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. The skewness of the normal distribution (or any perfectly symmetric distribution) is zero. The skewness of a distribution is defined as

$$y = \frac{E(x-\mu)^3}{\sigma^3} \qquad (3)$$

Where μ is the mean of x, σ is the standard deviation of x, and $E(x-\mu)$ represents the expected value of the quantity $(x-\mu)$.

Kurtosis:

Kurtosis is a measure of how outlier-prone a distribution is. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3. The kurtosis of a distribution is defined as

$$y = \frac{E(x-\mu)^4}{\sigma^4} \qquad (4)$$

Where μ is the mean of x, σ is the standard deviation of x, and $E(x-\mu)$ represents the expected value of the quantity $(x-\mu)$.

Variance:

It returns the variations in the pixels of an image. The variance of a distribution is defined as,

$$V = \sum_{i=1}^{M}\sum_{j=1}^{N}(f_{(i,j)} - \bar{f}_{(i,j)})^2 \qquad (5)$$

Where f is the original binary image and $\bar{f}$ is the average value of the pixels in a binary image of vegetables.

Energy:

It returns the sum of squared elements in the gray-level co-occurrence matrix. Range = [0 1]. Energy is 1 for a constant image. The energy of a gray image is given by,

$$E = \sum_{i,j} f(i,j)^2 \qquad (6)$$

Where f is the gray-level image of vegetables.

5. DECISION-TREE CLASSIFIER

Decision tree learning used in statistics, data mining and machine learning, uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. More descriptive names for such tree models are classification trees or regression trees. In these decision tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision tree learning is a method commonly used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables. Each interior node corresponds to one of the input variables; there are edges to children for each of the possible values of that input variable. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf. A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. Knowledge discovery is an important tool for the intelligent business to transform data into useful information that will increase the business revenue.

6. RESULTS AND DISCUSSION

Most commonly availabe vegetables in India are considered for this work. The vegetable database contains 8 classes of total 269 vegetable images. The images were downloaded from world wide web and some images taken from Canon Digital camera with natural day light. Images were resized into 300 X 300 pixel resolution to speed up computation. The proposed method effectively classify vegetables into 8 classes. We have used Decision Trees for classification of these Vegetables. The feautre set contains average red component, average green component, skewness, kurtosis, variance and energy. The confusion matrix shows the accuracy of the Decision Tree. When we evaluate the training samples we got the classification accuracy upto 95% and the confustion matrix is shown in table 1.

Table 1. Confusion Matrix

	Cabbage	Beetroot	Capsicum	Carrot	Chillies	Cucumber	Bittermelon	Onion	Succes Rate
Cabbage	47	0	0	1	0	0	0	0	97%
Beetroot	1	49	0	1	0	0	1	2	90%
Capsicum	0	0	57	0	0	0	0	1	98%
Carrot	1	1	0	108	0	0	0	2	96%
Chillies	0	1	0	1	32	0	0	1	91%
Cucumber	0	0	1	0	0	8	0	0	88%
Bittermelon	0	1	1	0	0	1	6	0	66%
Onion	0	0	0	0	0	0	0	71	100%

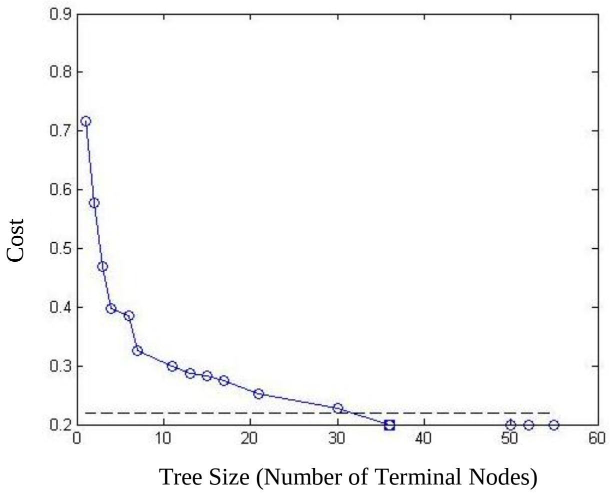


Figure 9: Graph of minimum-cost tree

7. CONCLUSION

In this paper, we have used watershed segmentation to segment the vegetable images form the background by converting RGB image into gray-level images. In the segmented region the average red component, average green component, skewness, kurtosis, variance and energy of individual vegetable objects are determined for classification. The classification is done using decision tree classifier. This method can be extended to other objects such as classification of flowers, fruits and seeds etc. where there is a human intervention is existed for classification.

8. ACKNOWLEDGEMENTS

Our thanks to the experts who have contributed towards development of the paper

8. REFERENCES

[1] Andrew, B. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition.* 30(7), 1145-1159.

[2] Benbrahim H, A comparative study of pruned decision trees and fuzzy decision trees, International Conference of the North American Fuzzy Information Processing Society, 227 – 231, 2000.

[3] Buzera M., Groza V., Prostean G., Prostean O., “Techniques of Analyzing the Color of Produces for Automatic Classification”, Intelligent Engineering Systems, 2008. INES 2008. International Conference on, 209-214, 25-29 Feb. 2008.

[4] Cesar, F. Peter, F. and Jose, H. 2003. Improving the AUC of probabilistic estimation trees. *Proc. of the 14th European Conf. on Mach. Learning*. 121-132, Springer.

[5] Fernandez C., Suardiaz J., Jimenez C., Navarro P.J., Toledo A., Iborra A., "Automated visual inspection system for the classification of preserved vegetables", Industrial Electronics, 2002. ISIE 2002. Proceedings of the 2002 IEEE International Symposium on, 265–269, 2002.

[6] Foster, P. and Venkateswarlu, K. 1999. A survey of methods for scaling up inductive algorithms. *Data Mining and Knowledge.* 3(2), 131-169.

[7] Leo, Breiman,. Jerome, Friedman,. Charles, Stone. and Richard, Olshen. 1984. *Classification and Regression Trees*.

[8] Pla F., Sanchiz J.M., Sanchez J.S., "An integral automation of industrial fruit and vegetable sorting by machine vision", Emerging Technologies and Factory Automation, 2001. Proceedings. 2001 8th IEEE International Conference on, 541 - 546 vol.2, 15-18 Oct. 2001.

[9] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Prentice-Hall, 617-620, 2002.

[10] Ross, Quinlan. 1993. *C4.5: Programs for Machine Learning. Morgan Kaufmann.* San Francisco, CA, USA.

[11] Singh M., Singh P., Hardeep Singh, "Decision Tree Classifier for Human Protein Function Prediction", Advanced Computing and Communications, 2006. ADCOM 2006. International Conference on, 564-568, 20-23 Dec. 2006.

[12] Thangaparvathi B., Anandhavalli D., Mercy Shalinie S., "A high speed decision tree classifier algorithm for huge dataset", Recent Trends in Information Technology (ICRTIT), 2011 International Conference on, 695-700, 3-5 June 2011.