ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 7, July 2016

Color ,Shape and Texture based Fruit Recognition System

Ruaa Adeeb Abdulmunem Al-falluji

University of Babylon, Babylon, Iraq

***Abstract*—The paper presents an automated system for classification of fruits. A dataset containing five different fruits was constructed using an ordinary camera. All the fruits were analyzed on the basis of their color (RGB space), shape and texture and then classified using different classifiers to find the classifier that gives the best accuracy. Gray Level Co- occurance Matrix(GLCM) is used to calculate texture features .Best accuracy was achieved by support vector machine (svm). All the processing was carried out in Matlab.**

Keywords—computer vision, pattern recognition, support vector machine, texture features.

I. Introduction

Advancement in the field of cameras and sensors, in recent years, has led to an increase in intelligent systems. The main objective of these systems is to understand and perceive an image as done by humans i.e. understanding the symbolic meaning of images by the help of statistics, models, geometry e.t.c. Some of the examples of such systems include: Controlling Systems e.g. different types of industrial robots, Navigation Systems e.g. autonomous vehicles and route planner, Automatic Inspection systems e.g. crop disease detection, Event Detection Systems e.g. human action and visual surveillance. Computer vision plays an important role in all such systems.

Agricultural field is increasingly using image processing to automate its processes. Now automated systems are being used for checking the progress of crops[1], diseased crops[2] and to recognize vegetables and fruits[3]. Fruit recognition and classification systems can be used by many real life applications. Such as a supermarket checkout system where it can be used instead of manual barcodes, and as an educational tool to enhance learning, especially for small children and Down syndrome patients[3,4]. It can assist the plant scientists, where shape and color values of the fruit images that have been computed can assist them do further analysis on variation in morphology of fruit shape in order and can help them understand the genetic and molecular mechanisms of the fruits[4]. Also, it can be used as aiding tool for eye weakness people which can aid them in shopping as a mobile application.

Recognizing different types of vegetables and fruits is a repeated chore in supermarkets, where the cashier has to define each item type which will determine its cost. The barcodes usage mostly ended this packaged products difficulty but when consumers want picking their produce; they will not be able to package it, and thus should be weighted. A popular solution to this difficulty is supplying codes for every type of fruit and vegetable; that has problems precondition that the memorization is sticky, leading to errors in pricing. Another solution is as mall book with pictures and codes; the difficulty with this solution is that flipping over the pamphlet is time-consuming[3]. A fruit and vegetable recognition system which automates labeling and computing the price is a good solution for this problem.

The main goal of this work is to automatically recognize fruit image by classifying it according to its features using machine learning techniques. This paper is divided into the following sections. Section II introduces some recent research works related to fruit recognition and classification. Section III explains the proposed methodology. Section IV shows the experimental results and the last section compares the proposed algorithm with other existing algorithms and then presents the conclusion.

II. Literature Review

Different authors, over the years, have proposed different methods for classification of fruits. Bhanu and Navneet[5] used shape (area, perimeter, major axis length and minor axis length), color (mean and variance of HIS, HSV, L*A*B and YbCbCr components) and texture (GLCM texture feature of contrast, dissimilarity, angular second moment, energy and entropy) features and classified them using artificial neural network.

Hashem Tamimi[6] proposed a fruit and vegetable recognition system which used histograms of hue and saturation components and features and Chi-square minimum distance method as classifier. [7]Chromaticity in the RGB space combined with shape features extracted using Fourier descriptors, Hu moments and basic geometric features are used by Farid for the classification using artificial feed forward neural network.

Another approach proposed by Pragati Ninawe[8], uses shape (area, perimeter and roundness value), color(mean of RGB value), and texture(entropy) with nearest neighbor classifier.

Saswati Naskar[9] proposed method represents fruit recognition expanding multiple feature based analysis that includes texture, color and shape. To recognize the texture of a fruit the Log Gabor filter has been used, mean hue has been calculated for color and shape has been analyzed by counting perimeter and area pixel.

III. Propsed System

The proposed system is divided into three main parts: Preprocessing, Feature extraction and classification. Figure 1 shows the flow diagram of the proposed system.

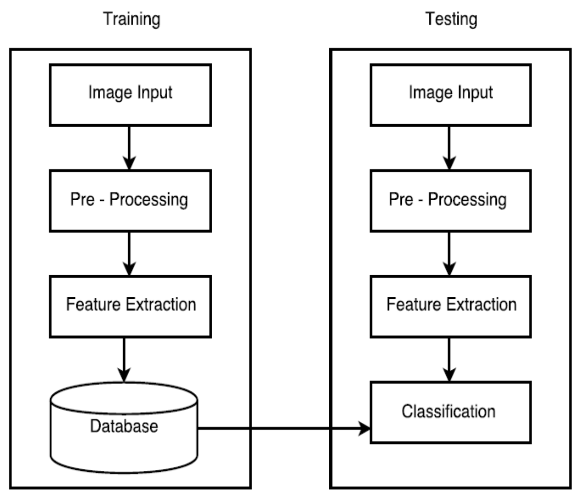

Fig. 1. Flow Diagram

A. Preprocessing

Arithmetic mean filter of 3x3 is applied on each image before extraction of features to remove any uniform and Gaussian type noise that may be present during the imaging process. The fruit is segmented from the background by thresholding the images. Also the image is resized to 256x256 pixels to increase the speed of the processing.

B. Feature Extraction

Color, shape and texture features are extracted and concatenated together to form a complete feature vector.

1) Color Features

Color is considered as a significant feature for image representation due to fact that color is invariance with respect to image translation, scaling, and rotation. Therefore, the first feature extraction method uses color characteristics to generate the feature vector for each fruit image in the dataset. The moments used to describe color are mean and variance.

$$Mean = \frac{1}{n}\sum_{i=1}^{n} x(i)$$

Where 'n' is the total number of values, 'x' is the value at the 'i' position.

$$Variance = \frac{\sum_{i=1}^{n}(x(i)-\mu)}{n}$$

Where μ is the mean, n' is the total number of values, 'x' is the value at the 'i' position.

Mean and variance are calculated separately for each channel i.e R, G and B, and are concatenated together to form color feature vector.

2) Shape Features

The second component that can be used as an attribute for fruit recognition is shape. The main motivation behind the shape analysis is that different fruits may have identical color or shape but possibilities of same value in both attributes (color and shape) are rare. The shape feature used are area, Number of pixels in an image is used for determining the area fruit. Perimeter, specifies the distance around the boundary of the region. Eccentricity, measure of how much the conic section deviates from being circular. EquivDiameter, the diameter of a circle with the same area as the region. Lastly, Major Axis Length and Minor Axes Length.

3) Texture Features

In statistical texture analysis, texture features are computed from the statistical distribution of observed combinations of intensities at specified positions relative to each other in the image. According to the number of intensity points (pixels) in each combination, statistics are classified into first-order, second order and higher-order statistics. The Gray Level Coocurrence Matrix (GLCM)[10] method is a way of extracting second order statistical texture features.

GLCM is created in four directions with the distance between pixels as one. Texture features are extracted from the statistics of this matrix. GLCM is composed of the probability value, it is defined by which expresses the probability of the couple pixels at direction and d interval. When and d is determined, is showed by Pi, j. Distinctly GLCM is a symmetry matrix and its level is determined by the image gray-level. Elements in the matrix are computed by the equation shown below:

$$P(i,j \mid d,\theta) = \frac{P(i,j \mid d,\theta)}{\sum_i \sum_j P(i,j \mid d,\theta)}$$

where P(i, j | d, θ) contains the second order probability values for changes between gray level 'i' and 'j' at distance 'd' a particular angle 'θ' .

GLCM expresses the texture feature according the correlation of the couple pixels gray-level value at different positions. It quantificational describes the texture feature. In this system, four texture features are considered. They include energy, contrast, correlation, Homogeneity.

$$Energy = \sum_{i,j} p(i,j)^2$$

It is a texture measure of gray-scale image represents homogeneity changing, reflecting the distribution of image gray-scale uniformity of weight and texture.

$$Contrast = \sum_{i,j} |i-j|^2 \, p(i,j)$$

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 7, July 2016

Contrast is the main diagonal near the moment of inertia, which measures how the values of the matrix are distributed and number of images of local changes reflecting the image clarity and texture of shadow depth. Large Contrast represents deeper texture.

$$\text{Correlation}=\sum_{i,j}\frac{(i-\mu i)\,p(i,j)}{\sigma_i\sigma_j}$$

Correlation measure of the intensity contrast between a pixel and its neighbor over the whole image.

$$\text{Homogeneity}=\sum_{i,j}\frac{p(i,j)}{1+|i-j|}$$

Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

IV. EXPERIMENTAL RESULTS

A. Dataset

No public dataset was available for fruits so a new database was constructed. The dataset contains images of five fruits namely apple, apricot, lemon, banana and Orange. Fifty different images of each fruits were taken from different positions and angles.

B. Classification

For testing and training 5-fold cross validation is used, the dataset is divided into five parts. Each part is used for testing while the other parts are used for training. Final accuracy is computed by taking the mean of accuracy for each part. Cross-validation combines (averages) measures of fit (prediction error) to correct for the optimistic nature of training error and derive a more accurate estimate of model prediction performance

C. K – Nearest Neighbor(KNN)

KNN[11] with Euclidean distance[12] was used to classify the features. The accuracy of KNN for each testing set while varying the number of neighbors from 5 to 9 is shown in Table 1. The last column of the table shows the average percentage accuracy.

TABLE I. KNN PERCENTAGE ACCURACY

K[1]	T1	T2	T3	T4	T5	Avg(%)
5	95	99	99	99	97	98.4
6	99	97	95	97	97	97.6
7	95	99	95	99	97	97.6
8	97	99	95	93	97	97.6
9	93	97	93	97	99	97.6

1. Number of Neighbors

D. Binary Classification Tree

Binary Classification tree[13] were primarily used for two class classifications. Improvements have been made to the algorithm to use for multi-class classification. In multi-class classification tree a class group is partitioned into two distinct subgroups at each node, and each node has a binary classifier that in turn assigns the sample pattern to one of the two classes.

TABLE II. BINARY TREE PERCENTAGE ACCURACY

Split Criteria	T1	T2	T3	T4	T5	Avg(%)
Gini's Diversity Index	93	97	93	97	99	97.6
Twoing rule	95	99	95	99	97	97.6
Deviance Reduction	95	99	95	99	97	97.6

E. Support Vector Machine

The Support Vector Machine (SVM) classifier is a theoretically superior machine learning methodology that used for classification and regression of high dimensional datasets with great results [14,15,16,17]. SVM tries to find an optimal separating hyperplane which effectively separates between classes for solving the classification problem. SVM aims to maximize the margin around a hyperplane that separates a positive class from a negative class.

TABLE III. SVM PERCENTAGE ACCURACY

Kernal	T1	T2	T3	T4	T5	Avg(%)
Linear	100	100	100	100	100	100
Quadratic	100	100	100	100	100	100
Gaussian	97	93	89	89	85	92.4

F. Discussion

Experimental results show that the best accuracy achieved was 100% using svm with linear and quadratic kernel. Figure 2 below shows the confusion matrix for svm with linear kernel.

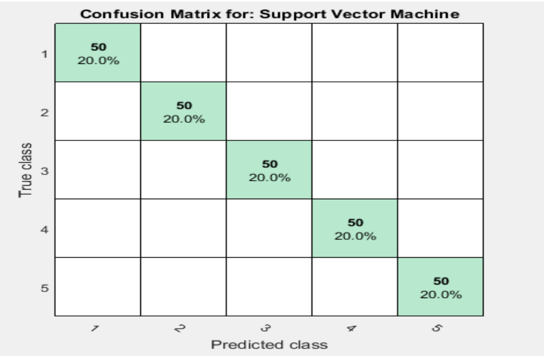

Fig. 2. Confusion Matrix

The use of different classifiers show the choice of classifier and its parameter are as important as the quality of the features extracted. We were able to get better results than any other proposed method.

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 7, July 2016

TABLE IV. COMPARISION TABEL

Author	Accuracy(%)
Bhanu Pratap	96
Hashem Tamimi	75
Farid García-Lamont	93.89
Pragati Ninawe	95
Saswati Naskar	90
Proposed System	100

V. CONCLUSION

The paper purpose a fruit classification system based on the color, shape and texture of the fruits. Many fruits have same color and shape but the chances of have all similar features are negligible. Different classifiers were experimented with to find the classifier which gives best results. SVM provided the best results with an accuracy of 100%. We were able to get better accuracy than all the recently proposed methods. To further test the efficiency of the system more fruit types should be included

REFERENCES

[1] Elhariri, E., El-Bendary, N., Fouad, M.M.M., Platos, J., Hassanien, A.E., Hussein, A.M.M.: Multi-class SVM based classification approach for tomato ripeness. In: Abraham, A., Krömer, P., Snášel, V. (eds.) Innovations in Bio-inspired Computing and Applications. AISC, vol. 237, pp. 175–186. Springer, Heidelberg (2014)

[2] Camargo, A., Smith, S.: An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosystems Engineering 102(1), 9–21 (2009)

[3] Rocha, A., Hauagge, D.C., Wainer, J., Goldenstein, S.: Automatic fruit and vegetable classification from images. Computers and Electronics in Agriculture 70(1), 96–104 (2010)

[4] Seng, W.C., Mirisaee, S.H.: A new method for fruits recognition system. In: International Conference on Electrical Engineering and Informatics, ICEEI 2009, Selangor, Malasia, pp. 130–134 (2009)

[5] Pratap, Bhanu, et al. "Development of ANN based Efficient Fruit Recognition Technique." Global Journal of Computer Science and Technology 14.5-C(2014)

[6] Pujari, Jagadeesh D., Rajesh Yakkundimath, and Abdulmunaf S. Byadgi. "Recognition and Classification of Normal and Affected Agriculture Produce using Reduced Color and Texture Features." International Journal of Computer Applications 93.11 (2014).

[7] García-Lamont, Farid, et al. "Color Characterization Comparison for Machine Vision-Based Fruit Recognition." International Conference on Intelligent Computing. Springer International Publishing, 2015.

[8] Ninawe, Pragati, and Mrs Shikha Pandey. "A Completion on Fruit Recognition System Using K-Nearest Neighbors Algorithm.", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 7, July 2014

[9] Naskar, Saswati, and Tanmay Bhattacharya. "A Fruit Recognition Technique using Multiple Features and Artificial Neural Network." International Journal of Computer Applications 116.20 (2015).

[10] Nitish Zulpe, V.P.: GLCM Textural Features for Brain Tumor Classification. International Journal of Computer Science Issues 9(3), 354–359 (2012)

[11] E.Umbaugh, S., Computer Vision and Image Processing: A Practical Approach using CVIP tools. first ed. 1998: Prentice Hall Professional Technical Reference.

[12] AA.R. Jimenez, R.C., J.L. Pons. Shape-based Methods for Fruit Recognition and Localization using a Laser Range-Finder. in International Workshop on Bio-Robotics. 1997.

[13] Breiman, L., J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Boca Raton, FL: CRC Press, 1984.

[14] Tzotsos, A., Argialas, D.: Support vector machine approach for object-based image analysis. In: Object-Based Image Analysis. Lecture Notes in Geoinformation and Cartography, pp. 663–677 (2008)

[15] Wu, Q., Zhou, D.X.: Analysis of support vector machine classification. Journal of Computational Analysis & Applications 8(2), 99–119 (2006)

[16] Zawbaa, H.M., El-Bendary, N., Hassanien, A.E., Abraham, A.: SVM-based Soccer Video Summarization System. In: Third World Congress Nature and Biologically Inspired Computing, Salamanca, Spain, pp. 7–11 (2011)

[17] Zawbaa, H.M., El-Bendary, N., Hassanien, A.E., Kim, T.-h.: Machine Learning- Based Soccer Video Summarization System. In: Kim, T.-h., Gelogo, Y. (eds.) Mul- GraB 2011, Part II. CCIS, vol. 263, pp. 19–28. Springer, Heidelberg (2011)

APPENDIX 1:DATASET IMAGES

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 7, July 2016