A Comparative Study of Image Classifiers in a Fruit Recognition System.

Seema Paragi, H. Girish.

***Abstract*— This paper deals with the analysis of various image segmentation techniques and classifiers designed for a fruit recognition system. A classifiers efficiency mainly depends on the feature vector and the size of the training samples, the efficiency of the image segmentation algorithm depends on the threshold value selection. We have implemented image segmentation algorithm, namely, intermeans algorithm. The segmented images are used to extract the colour and shape features. Using the 2D feature vector three classifiers, namely, Distance classifier and Bayesian classifier and S.V.M are implemented for the classification of four classes of fruits. The performances of the classifiers have been tabulated for different sizes of training data set. The computations have been carried using MATLAB.**

***Index Terms*—Segmentation, Classifiers, Features, Fruit Recognition.**

I. Introduction

Pattern recognition is the scientific discipline whose goal is classification of objects into a number of categories or classes. Depending on the application, these objects can be images or signal waveforms or any type of measurements that needs to be classified. Pattern recognition is an integral part of most machine intelligence systems built for decision-making. Machine vision is an area in which pattern recognition is of importance. A machine vision system captures images via a camera and analyzes them to produce descriptions of what is imaged. Based on these descriptions, the application and the specifications, the pat- tern recognition system is built.

One of the applications considered here is a fruit recognition system. The fruit recognition system can be applied for educational purpose to enhanced learning of fruits, especially for small kids and Down syndrome patients. It can be used in grocery store, which makes the customers label their purchases using automatic fruit recognition based on computer vision. It can also be used to remove defective products in food industry. The system can be applied for grading the fruits by their quality, size or ripeness based on their appearance, as well as to make decisions on whether they are healthy or diseased. A number of challenges had to be overcome to enable the system to perform the automatic fruit recognition. Fruit images are subject to significant variation in color, size and texture. For example, Bananas range from being uniformly green, to yellow, to patchy and brown [4]. A typical pattern recognition system consists of the processes as shown in Figure 1.

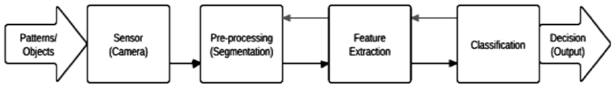


Figure 1. System Architecture.

Figure 1 the present work deals with four different classes of fruits, namely, apple, banana, chickoo and strawberry. The flow is structured as follows. Section 2 discusses the Segmentation Techniques. Section 3 gives the features. Section 4 deals with the classifiers. Finally, Section 5 and 6 gives the results and concluding remarks of this paper.

II. Segmentation

Image segmentation is the division of an image into regions or categories, which correspond to different objects or parts of objects. Every pixel in an image is allocated to one of a number of these categories. It is a pre-processing stage of a pattern recognition system, in which the object is isolated from other objects or from the background. Segmentation is often the critical step in pattern recognition processes: the point at which we move from considering each pixel as a unit of observation to working with objects (or parts of objects) in the image, composed of many pixels. If segmentation is done well then the subsequent processes and analysis are made simpler. Out of the three general approaches to segmentation; termed thresholding, edge-based methods and region-based methods; we have chosen thresholding technique due to its simplicity. The algorithms generated using the thresholding techniques are fully automatic.

Histogram thresholding is one of the popular techniques for image segmentation. Consider an image f(x, y) whose histogram of pixel values is denoted by h0, h1, …, hN ; where hk specifies the number of pixels in an image with greyscale value k and N is the maximum pixel value. One obvious way to extract object from background is to select a threshold t that separates histogram. Then any point (x, y) for which f(x, y) it is called an object point; otherwise the point is called a background point, the threshold image g(x, y) is defined as (1) The histogram-based thresholding methods; inter-means algorithm, are considered which helps to determine the threshold value.

***Manuscript received June, 2014*.**

Seema Paragi, *4th Semester, M. Tech Research Scholar, Dept. of C. S. E., RYMEC, Bellary , Bellary, India, 8123565503.*

H. Girish, *Asst. Professor and Head of Dept, Dept. of C. S. E., RYMEC, Bellary.*

Figure 2. Segmented Results.

A. Inter-Means Algorithm

It was proposed[7] by Ridler and Calvard (1978) and Trussell (1979). The algorithm is as follows:

1. Initially, a guess has to be made at a possible value for the threshold, say ‘t’. For example, median pixel value or mean pixel value can be chosen.
2. From this, the mean values of pixels in the two categories produced using this threshold are calculated.

$$\mu_1 = \frac{\sum_{k=0}^{t} k * h_k}{\sum_{k=0}^{t} h_k}$$

For values less than or equal to t, this is given by: (2)
For values greater than t, it is given by: (3)

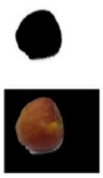
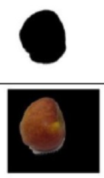
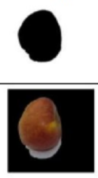
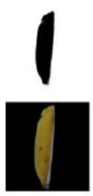
$$\mu_2 = \frac{\sum_{k=t+1}^{N} k * h_k}{\sum_{k=t+1}^{N} h_k}$$

3. The threshold is repositioned to lie exactly half way between the two means. i.e,

$$t = \frac{\mu_1 + \mu_2}{2}$$

4. Mean values are calculated again and a new thresh- old is obtained, and so on until the threshold stops changing value, i. e., steps (2) and (3) are repeated till ‘t’ stops changing value between consecutive evaluations.

The segmentation results are in following Figure 2.

Fruit	*Intermeans*	*Error- Minimum*	*Shannon’s Entropy*
			
Threshold	163	171	164
			
Threshold values	188	205	200

III. Feature Extraction

The segmented images are used to extract the features. The two features used in this paper are color feature and shape feature. [3, 4, 5].

1. *Color Feature* : The mean of the Red component and the Green component of the segmented image is computed and its ratio is taken, i. e., R/G.
2. *Shape Feature* : the area of the segmented image and the perimeter is obtained in terms of pixels and the measure of roundness is calculated using,

$$S = 4 * \pi * \left(\frac{area}{perimeter^2}\right)$$

IV. Classifiers

A. Distance Classifier

The distance classifier [2] that we have implemented employs the Euclidean distance given by, (18) Steps followed for implementing the distance classifier:

1. Extracting the features from each of the tagged training set of images.
2. Representation of each image as a point in the 2D feature vector space.
3. Then locate the corresponding centroids of all the classes of fruits in the feature vector space.
4. The features are extracted for the test image and the respective image is represented in the 2D feature vector space.
5. The Euclidian distances of the test image from the centers of all the classes are calculated. Then the test image is assigned to that class for which the Euclidian distance of the test image is less.

The results are tabulated in Table 1 shows each fruit of 100 training images and Table 2 shows each fruit of 150 training images, and Figure 3 shows the distribution of the samples in the feature space.

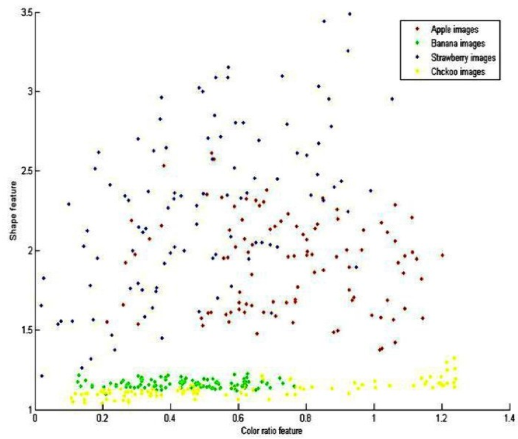


Figure 3. Feature vector of colour and shape.

ISSN: 2278 – 1323

Table 1. Distance Classifier for 100 training images

Test Images	Actual	Expt. Results
Apple	24	21
Banana	30	45
Strawberry	24	20
Chickoo	22	14

Table 2. Distance Classifier for 150 training images

Test Images	Actual	Expt. Results
Apple	40	31
Banana	40	49
Strawberry	35	31
Chickoo	19	24

B. Bayesian Classifier

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. Baye's theorem is given by,

$$p(w_j / x) = \frac{p(x / w_j) * p(w_j)}{p(x)}$$

Where,

$$p(x) = \sum_{j}^{p} p(x / w_j) * p(w_j)$$

The Baye's Formula can also be expressed as:

$$Posteriori = \frac{Likelihood * \Pr oir}{Evidence}$$

Here, p(ωj |x) is the Posteriori Probability or Posteriori p(x|ωj) is the likelihood of ωj with respect to x. p(ωj) is the Prior probability, the true state of nature being ωj . p(x) is a scale factor that guarantees that the posterior probabilities sum to one (evidence). The product of Likelihood and the Prior Probability is most important in determining the Posterior Probability. Steps followed for implementing the Bayesian classifier:

1. Extracting the features from each of the tagged training set of images.
2. Likelihood estimation a. Based on the observation of the feature color, discretize the feature into 'i' ranges, a1, a2, a3,.,ai.. b. Count the number of images falling into each range. c. Count/total image of the respected class gives you the likelihood. d. similarly find the likelihood of the feature shape.
3. Calculate the prior probability.
4. Use the likelihood and the prior to calculate the posteriori.
5. Assign to class A if the posteriori of class A is greater than the other class. Tabulated results are in Table 3 shows each fruit of 100 training images and Table 4 shows each fruit of 100 training images but using both the features.

Table 3. Bayesian Classifier

Test Images	Actual	Expt. Results
Apple	20	12
Banana	20	27
Strawberry	20	28
Chickoo	20	13

Table 4. Bayesian Classifier (using both features)

Test Images	Actual	Expt. Results
Apple	20	15
Banana	20	26
Strawberry	20	24
Chickoo	20	15

C. Support Vector Machine Classifier

SVM are supervised learning models with associated learning algorithm that analyze data and recognize patterns used for classification.

Group = svmclassify(SVMStruct,Sample) Group = svmclassify(SVMStruct,Sample, 'Showplot',true) Group = svmclassify(SVMStruct,Sample) classifies each row of the data in Sample, a matrix of data, using the information in a support vector machine classifier structure SVMStruct, created using the svmtrain function. Like the training data used to create SVMStruct, Sample is a matrix where each row corresponds to an observation or replicate, and each column corresponds to a feature or variable. Therefore, Sample must have the same number of columns as the training data. This is because the number of columns defines the number of features. Group indicates the group to which each row of Sample has been assigned.

CONCLUSION

The segmentation result shows that for the application of fruit recognition, the intermeans algorithm gives a better threshold value than any other algorithms. The inter- means threshold value is observed to be lesser than that of the other two implemented algorithms, which is advantageous because it eliminates the shadow problem, which helps in extracting the features accurately. The response of the classifiers with respect to the feature vector and size of the training data set states that Distance classifier yields quite good results even with the fewer amounts of sam- ples. Our results show that Bayesian classifier works quite well even with a single feature, but the Distance classifier yields poor results. S. V. M. works will efficiently with two classes of fruits.

ACKNOWLEDGMENTS

I would like to show my greatest appreciation to my respected guide, Mr. H. Girish, Asst. Professor. I would like to thank Mr. H. Girish, Asst. Professor and Head, Dept. of CS & E, and The Principal, RYMEC, Bellary for their support and encouragement.

to the journal again.

REFERENCES

[1] Andrew webb, - Statistical Pattern recognition, second edition, Wiley publication.
[2] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, second edition.
[3] Woo Chaw Seng, Seyed Hadi Mirisaee, A New Method for Fruits Recognition System.
[4] Hetal N. Patel, Dr. R. K. Jain, Dr. M. V. Joshi, Fruit Detection using Improved Multiple Features based algorithm, International Journal of Computer Applications (0975 8887), Volume 13 No.2, January 2011.
[5] S. Arivazhagan, R. Newlin Shebiah, S. Selva Nidhyanandan, Fruit recognition Using Color and Texture, Journal of Emerging Trends in Computing and Information Sciences, VOL. 1, NO. 2, Oct 2010.
[6] Gonzalez & Woods, Digital image processing.
[7] John C. Russ, The Image Processing Handbook, Sixth edition.
[8] Marek Wojcikowski, Robert Zaglewski, Bogdan Pankiewicz, "An Intelligent Image Processing Sensor - the Algorithm and the Hardware Implementation", Gdansk University of Technology.
[9] Images, D S Guru, Y. H. Sharath, S. Manjunath, "Texture Features and KNN in Classification of Flower", Department of Studies in Computer Science Manasagangotri, University of Mysore.
[10] Abhishek Gupta, "A Flexible and Efficient Algorithm design for Segmentation in Textural Images", Department of Computer Science and Engineering, Punjab Engineering College, University of Technology, Chandigarh.
[11] Michel Vidal-Naquet, Shimon Ullman, "Object Recognition with Informative Features and Linear Classification", Faculty of Mathematics and Computer Science The Weizmann Institute of Science.

