Recognition and Classification of Normal and Affected Agriculture using Fruit disease detection

Dr. M. Vargheese[1], J. Yamuna Bee[2], P. Kanagalakshmi[3], A. Anitha[4], K. Soniya[5]

[1]Professor, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India
[2]Assistant Professor, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India
[3]Assistant Professor, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India
[4]Assistant Professor, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India
[5]Student, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India

---

Abstract - *Object Recognition is an important study in Computer Science. Object recognition is emerging technology to detect and classify objects based on their characteristics. Fruit recognition and automatic classification of fruits is also a domain of object recognition and it is still a complicated task due to the various properties of numerous types of fruits. Different fruits have different shapes, sizes, colour, textures and other properties. Similarly, some of the fruits like Tangerines and Mandarin Oranges share the same characteristics like colour, texture, size, etc. This project aims to find a better way of a fruit classification method using supervised machine learning algorithms and image processing mechanisms based on multifeature extraction methods. Firstly, we pre-process the training sample of fruits' images. Pre-processing includes separating foreground and background, scaling and cropping the image to reduce the dimension so that the processing is fast. Then, we extract features from the fruit's image, which includes colour, texture and shape of the fruit image. Extracted features are then fitted into the neural classifier machine-learning algorithm. Finally, the results obtained from the machine-learning network are cross-validated with the test sample. The output obtained will give us the prediction accuracy and class of the fruit that it has acknowledged. Experimental results have been collected using a fruit image database consisting of five different classes of fruits and number of fruits images overall. Therefore, average prediction accuracy of more than 55% is obtained with a learning rate of 0.7.*

Key Words: Object Recognition, Pre-processing, Machine Learning Networks, Multifeature Extraction

1. INTRODUCTION

Object Recognition implements pattern recognition of different objects. Pattern recognition builds up from different areas such as statistics and machine learning. To achieve good object detection, classification and recognition, different machine learning algorithms and object's feature extraction algorithms are used. While using machine learning algorithms, it is not a guarantee that every algorithm gives accurate result. The achievement of accuracy can be different for different algorithms. Hence, we need to select the (best) algorithm with the highest classification and prediction accuracy. Also, while training the system, proper learning rate also plays a vital role.

For fruit classification and detection this project implements a portion of computer vision and object recognition with machine learning model. The rapid development of computer vision, image processing and recognition, advancement in computer technology provides the possibility of fruit classification through computer vision. In recent years, fruit recognition using computer vision is being gradually applied in agriculture sector, education sector and supermarkets [1]. Computer vision has been widely used in industries to aid in automatic checking processes [2]. The important problem in computer vision and pattern recognition is shape matching. Shape comparison and shape matching can be carried out by using computer vision and image processing algorithms. Shape matching applications contain image registration, object detection and recognition, and image content based retrieval [3]. Many agricultural applications use image processing to automate their duty. Detecting crop diseases are one of these applications in which the crop images are analyzed in order to discover the affected diseases [4].

1.1 Problem Statement

Despite of advancement in computer vision, image processing, recognition and advancement in computer technology, automatic fruit classification is a challenging task. The primary parameters that play vital role while classifying a fruit include the machine learning algorithm that is being used, quality of images in the fruit database, fruit's images' shape and size and fruit's color. Secondary parameters that affect the classification are similar characters of fruits like color, shape, size, etc. If both primary and secondary parameters are not analyzed properly in the beginning then it may cause problem during classification and may lead to less accuracy and unexpected results. Many related works have been conducted in fruit classification using different classification algorithms but those approaches still lack in some aspects. A research in fruit classification has been carried out by just considering only three fruits with 100% accuracy [5]. However, considering only three fruit in the sample is not enough because the trained model may not recognize the fruit's images' that are out of the training sample.

Similarly, proper implementation of machine learning algorithm should also be taken into consideration while performing classification. Using Multiclass SVM algorithm, gives success rate or accuracy in range of 70% to 75% [6]. However, this model may provide wrong interpretation or recognition for fruits with similar features like shape, size, colour, texture, etc. Some approaches are only focused on one feature while others combine two features, resulting in distinctive methods. However, different fruit images can have same color and shape, which may pose a problem. So, it is required to have more features to make the recognition process more robust and effective.

1.2 Objective of the Project

The main objectives of this project are: a) To extract at least three features from the fruit's image. The features extracted are Haar-Like Features [7], Hue Histogram Feature or Color Histogram [8] and Edge Histogram Feature [9]. b) To implement the neural classifier algorithm [10] for automatic fruit classification. c) To develop a web interface platform for testing the prediction of fruit image.

The organization of this document is as follows. In Section 2 (Methods and Material), I'll give detail of any modifications to equipment or equipment constructed specifically for the study and, if pertinent, provide illustrations of the modifications. In Section 3 (Result and Discussion), present your research findings and your analysis of those findings. Discussed in Section 4(Conclusion) a conclusion is the last part of something, its end or result.

2. METHODS AND MATERIAL

The waterfall development model was followed for this project because it is simple, easy to understand and to use it. Since, it is an individual project, it becomes easy to manage due to the rigidity of the model and each phase has specific deliverables and a review process. It was used because the model phases are processed and completed one at a time and these phases do not overlap. In addition, the requirements are well understood at the beginning of this small project, so, the waterfall model is helpful in this. In testing and validating the project, this model posed some difficulties such as it is very difficult to go back and change something that was not well thought out in the concept stage.

2.1 Scope

The scope of this project is only limited to edible fruits that are available in the fruit data sets which are used to train the system. Leafy and other vegetables like lettuce, cabbage, spinach, etc. may not be in the scope of this project. So, when images of these are provided as input to the system, it may not recognize and may not produce the desired result.

2.2 Limitation

For this project only 5 categories of fruits are used to train the system, hence, there might be false prediction and misclassification of fruits that are out of the training classes. Similarly, only three features from the fruit's image are extracted to reduce the processing complexities, which possibly will limit the prediction accuracy score.

(i) Data Collection

Due to project deadline, manual data collection was not performed. Rather, the fruit image data sets were recycled from the previous research [18]. The data consist of 30 fruits categories and each category of fruit has approximately 30 images. These images were used to train the system using neural classifier machine-learning algorithm. The set of data is definitely not enough to cover wider range and classes of fruit, but due to the project being solely based on the implementation of the machine learning algorithm to automatically classify an image rather than being used as a business product and also because of the tight schedule, these data sets were considered enough for the purpose of the project.

(ii) Data selection

There are around 971 image datasets of 30 fruits' categories in the collected data. Out of 30 fruit category, only five categories are selected for training purpose that has 120 images. Out of 120 images, 90% is used for the training purpose, whereas remaining 10% are used for the testing purpose. The fruit categories for this project are chosen randomly.

(iii) Algorithms

Different algorithms are implemented in this project. Algorithms implemented are image processing, feature extraction algorithm and machine learning algorithm. Image feature extraction algorithms are listed and described below: a) Haar-Like Feature Extraction This is used to generate Haar-Like features from an image. The classifiers of machine learning to help identify objects or things in the picture use these Haar-Like features. a) Hue Histogram Feature Extraction this feature extractor takes in an image, gets the hue channel, bins the number of pixels with a particular Hue, and returns the results. This feature extractor takes in a color image and returns a normalized color histogram of the pixel counts of each hue. b) Edge Histogram Feature Extraction. This method takes in an image, applies an edge detector, and calculates the length and direction of lines in the image. It extracts the line orientation and length histogram. The machine-learning algorithm used in the project is described below: In this project, a popular boosting algorithm Neural, introduced in 1995 by Freund and Schapire [16] has been implemented to classify the fruit images. The core principle of Neural is to fit a sequence of weak learners (i.e., models that are only slightly better than random guessing, such as small decision trees) on repeatedly modified versions of the data. The predictions from all of them are then combined through a weighted majority vote (or sum) to produce the final prediction. An Neural classifier is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases. Neural can be used to boost the performance of any machine-learning algorithm. The most suited and therefore most common algorithm used with Neural is decision trees with one level. Because these trees are so short and only contain one decision for classification, they are often called decision stumps. The Neural Training algorithm is described below: a) the initial weight for each instance in the training dataset weighted as

Weight (xi) = 1/n

Where xi is the i th training instance and n is the number of training instances. b) One weak learning model trained as: i. A weak classifier (decision stump) is prepared on the training data using the weighted samples. Only binary (two-class) classification problems are supported, so each decision stump makes one decision on one input variable and outputs a +1.0 or -1.0 value for the first or second class value. ii. The misclassification rate for weak learner is calculated for the trained model. Traditionally, this is calculated as:

Error = (correct – N) / N

Where, error is the misclassification rates correct are the number of training instance predicted correctly N is the total number of training instances.

3. RESULTS AND DISCUSSION

It can be concluded that the chosen ensemble machine-learning algorithm is not suitable for fruit classification problems. The cross validating process was iterated until the highest score was obtained. The result obtained is shown in the table and results below:

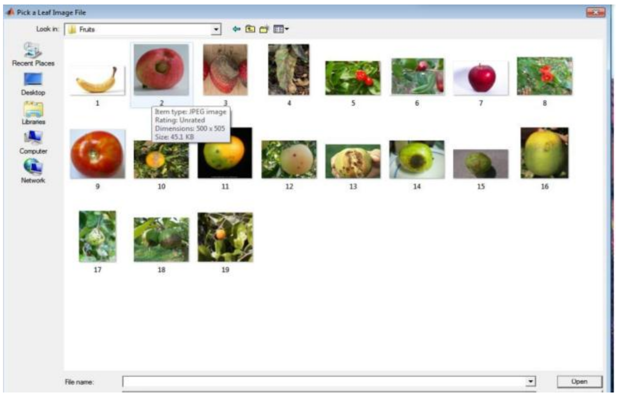

Fig -1: Input Images

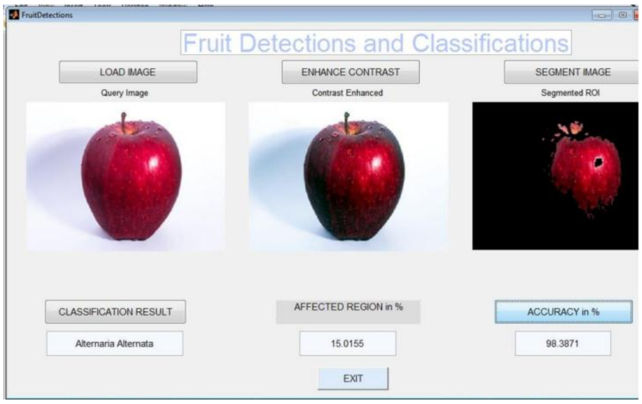

Fig -2: Segmentation Image

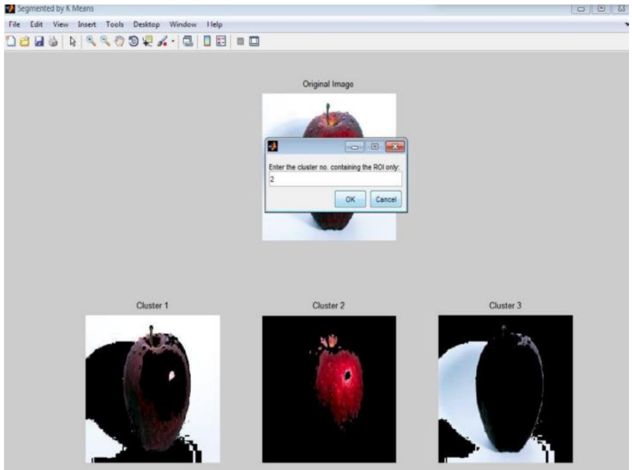

Fig -3: Clustering Image

Table -1: Mean Cross Validation Score (W.R.T) Learning Rate

Learning rate	Mean cross validation score
0.1	0.506
0.2	0.519
0.3	0.543
0.4	0.533
0.5	0.542
0.6	0.545
0.7	0.59

4. CONCLUSIONS

This project aims to classify the fruit images based on its Haar-Like, Hue and Edge histogram features. The project is designed in such way that it reads image, extracts features, pre-processes it, implements machine learning algorithm and generate output based on the input provided. The project has been able to classify the fruit images based on the fruits features. The cross validation score obtained is 54.9% with learning rate of 0.7 and the prediction accuracy of the system is above 55%. This result is not satisfactory since the cross validation score and probability of prediction accuracy is very less than what was expected. In some cases, the system does not predict the fruit images even the provided input falls under the training category. With this result, it can be concluded that the chosen ensemble machine-learning algorithm is not suitable for fruit classification problems.

REFERENCES

[1] Patel H.N et al., "Automatic Segmentation and Yield Measurement of Fruit using Shape Analysis", International Journal of Computer Applications, Volume 45– No.7, May 2012

[2] Aasima Rafiq, et al. "Application of Computer Vision System in Food Processing- A Review", PInt. Journal of Engineering Research and Applications, Vol. 3, Issue 6, Nov-Dec 2013

[3] Oikonomidis, A. A. Argyros and R. Boyle, "Deformable 2D shape matching based on shape contexts and dynamic programming" in ISVC 2009. LNCS, vol. 5876, 2009, Springer-Verlag, pp. 460-469

[4] A. Camargo and J. S. Smith, "An image-processing based algorithm to automatically identify plant disease visual symptoms," Biosystems Engineering, vol. 102, January 2009, pp. 9-21

[5] Cairo University, Faculty of Computers and Information, Cairo, Egypt Scientific Research Group in Egypt (SRGE) "Automatic fruit classification using random forest al2gorithm," in 2014 International Conference on Hybrid Intelligent Systems (HIS)., Kuwait, 2014, pp. 164-168.

[6] Yudong Zhang and Lenan Wu "Classification of Fruits Using Computer Vision and Multiclass Support Vector Machine" School of Information Science and Engineering, Southeast University, Nanjing 210096, CH, Sensors 2012, 12

[7] "Haar-like features", Wikipedia, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Haar-like_features. [Accessed: 19- Aug- 2016].

[8] "Color histogram", Wikipedia, 2016. [Online]. Available:https://en.wikipedia.org/wiki/Color_histogram. [Accessed: 19- Aug- 2016].

[9] "Edge detection", Wikipedia, 2016. [Online]. Available:https://en.wikipedia.org/wiki/Edge_detection. [Accessed: 19- Aug- 2016].

[10] "Neural", Wikipedia, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Neural. [Accessed: 19- Aug- 2016].

[11] D. Marr, "Vision". W. H. Freeman and Company, 1982.

[12] Ming-Hsuan Yang, "Object Recognition". University of California at Merced

[13] J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors. "Toward category-level object recognition". Springer-Verlag, 2006.

[14] Kwak, N. "Principal Component Analysis Based on L1-Norm Maximization". IEEE Trans. Patt. Anal. Mach. Int. 2008 , 30, 1672–1680

[15] D. G. Lowe, "Object recognition from local scale-invariant features," In Computer Vision, 1999. The proceedings of the seventh IEEE international conference, Corfu, Greece. pp. 1150-1157.

[16] Y. Freund, and R. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting" , 1997

[17] "Welcome | Flask (A Python Microframework)", Flask.pocoo.org, 2016. [Online]. Available: http://flask.pocoo.org/. [Accessed: 19- Aug- 2016].

[18] M. Š krjanec. "Automatic fruit recognition using computer vision", Bsc Thesis, (Mentor: Matej Kristan), Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, 2013.

[19] "Scientific Computing Tools for Python", Scipy.org, 2016 [Online] Available: https://www.scipy.org/about.html [Accessed: 20- Aug- 2016].

[20] "scikit-learn: machine learning in Python — scikit-learn 0.17.1 documentation", Scikit-learn.org, 2016. [Online]. Available: http://scikit-learn.org/stable/. [Accessed: 20-Aug- 2016].

BIOGRAPHIES

Dr. M. Vargheese,
Professor, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India

J. Yamuna Bee, M. E.,
Assistant Professor, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India

P. Kanagalakshmi, M. E.,
Assistant Professor, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India

A. Anitha, M. E.,
Assistant Professor, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India

K. Soniya, B. E.,
Student, CSE, PSN College of Engineering & Technology, Tirunelveli, Tamil Nadu, India