Naive Bayes Classifier to Classify Image of Citrus Fruits

Shubhi Jain, Graduate Student, *Computer Science Dept., University of Florida*

***Abstract*—The idea is to analyze citrus fruit images to extract hue and classify using features like surface color, horizontal diameter, and probability distribution function. The performance of the application will be evaluated in terms of classification accuracy relative to human classification and computational complexity. Naive Bayes Algorithm will be implemented to classify citrus fruits based on color , diameter and maturity.**

***Index Terms*—Classification, Single view, Color distance, Image Processing, Naive Bayes Theorem, Feature extraction, Feature selection**

I. Introduction

Fruit detection system is primarily developed for robotic fruit harvesting. However, this technology can easily be narrowed down to other applications such as on tree yield monitoring, Fruits classification, disease detection, maturity detection and other operations which require vision as a sensor. In citrus manufacturing industries, caliper and color are successfully used for the automatic classification of fruits using vision systems. To achieve the proposed goal, the specific objectives are as follows.

1. To build a rugged system for image processing.
2. To develop a pattern recognition algorithm that will successfully classify various citrus fruits.
3. To conduct experiments and evaluate the performance of the developed system.

I spent almost a week researching on how to achieve these objectives successfully and after Literature Survey I came up with following project proposal. Combining many features and classifiers, where all features are concatenated and fed independently to a classification algorithm is the aim of the project. The four basic features which characterize the fruit are: intensity, color, edge and orientation. Among these color and size are the most important features for accurate classification and/or sorting of citrus such as oranges, lemons and tangerines. The Bayesian Classifier is capable of calculating the most probable output depending on the input. It is possible to add new raw data at runtime and have better probabilistic classifier. Naive Bayes is fast to train and handles real and discrete data well. The performance of the classification can be determined by the accuracy of classification of citrus fruits. The higher the accuracy of the application to classify fruits using single view fruit images, more intensive and accurate is the probability based function. The paper is structured as follows: the next section discusses the Project Plan. Section 3 lists references I read that were relevant to my project topic.

II. Project Plan

Fruit recognition algorithm consists of segmentation, filtering, perimeter extraction, color extraction and color and perimeter-based detection.

A. Feature Extraction

Firstly, I plan to extract the features of my image and tend to complete this part of the project by end of March. This step will dimensionally reduce my test image to a compact feature vector that efficiently represents interesting parts of an image. This will help me to quickly complete tasks such as image matching and retrieval. The segmentation procedure will be using a procedure based on a Bayesian discriminate analysis, which will allow fruits to be precisely distinguished from the background. Features will rely on the value of the average channel Red, Green, Blue and Horizontal Diameter. Features like horizontal diameter, color, size shape will be extracted from the image to help classify various citrus fruits accurately.

B. Max-Min Filters

I will be using Max-Min filters in contrast enhancement, texture description and edge detection. Maximum and minimum filters attribute to each pixel in an image a new value equal to the maximum of a minimum value in the neighborhood around that pixel. The neighborhood stands for the shape of the filter.

C. *Image Descriptor*

I have chosen Color Mean Descriptor for my project since Citrus fruits have distinct bright single colors. I aim to complete this part by second last week of March. Image descriptor will handle the logic necessary to quantify an image and represent it as a list of numbers. The output of

image descriptor is a feature vector: the list of numbers used to characterize fruit's image and correlation factor between the light intensity in relation to value of FFB from Red, Green and Blue component of image taken. Looking at the dimensions of the image I can obtain height, width and 3 channels—one for each of the Red, Green, and Blue channels, respectively. This method returns a tuple with four values, the color features of the fruits. The first value is the mean of the blue channel, the second value the mean of the green channel, and the third value is the mean of red channel.

D. Horizontal Diameter

Since all Citrus fruits have similar shape, I will use size as the other feature. So, other feature that will be used apart from Mean color channels is the horizontal diameter. This can be implemented within a day or two by the and of second week of March. For this I will use a method in which pixels are scanned from top to bottom and left to right and will use this process in left to right and right to left manner and this will help me find the maximum value as diameter of a citrus fruit.

E. Feature Analysis

This part of the project I will begin in March ending week. A simple feature can not entirely represent the character of the fruit region. Therefore, multiple features analysis will be used in the proposed method. The computed features can then be integrated according to their weights. After integrating the feature map, we can obtain a final image map. Because of the use of multiple features and efficient feature integration, the proposed method will give good result of detection in clustered region of the single view image (my test image). Two features analysis methods: color-based and size-based will be used in order to increase accuracy of recognition.

F. Naive Bayes Classifier

To achieve the above feature extraction goal, I will implement Naive Bayes Classifier. I expect to complete the implementation of the algorithm by starting week of April. A naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence/absence of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

I expect to implement all the above algorithms and features in Java or Matlab. I expect to complete the project coding and implementation part by first week of April. I have set aside a time buffer at the beginning and at the end of the project so that if the project requires me to learn/research new domains of Java or Matlab, I have ample time.

III. CONCLUSION

I will test and improve the project from first week of March till 10 April 2016. After that I expect to start working on the video and report for the same. I will do more research then and will look forward to enhance the scope of my project.

REFERENCES

[1] H. N. Patel, D. R.K. Jain, and D. M.V. Joshi, "Fruit detection using improved multiple features based algorithm," *International Journal of Computer Applications*, vol. 13, no. 2, pp. 1–5, Dec. 2011.

[2] A. Rosebrock, "How to describe and quantify an image using feature vectors," in *Image Descriptors*, PyImageSearch, 2014. [Online]. Available: http://www.pyimagesearch.com/2014/03/03/charizard-explains-describe-quantify-image-using-feature-vectors/. Accessed: Mar. 11, 2016.C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

[3] R. Venkata Ramana Chary, "Feature extraction methods for color image similarity," *Advanced Computing: An International Journal*, vol. 3, no. 2, pp. 147–157, Mar. 2012.

[4] javafreecode, "Naive Bayes Classifier to classification image of citrus fruits with java code," Java Free Code, 2015. [Online]. Available: https://javafreecode.wordpress.com/2015/02/13/naive-bayes-classifier-to-classification-image-of-citrus-fruits-with-java-code/. Accessed: Mar. 11, 2016.

[5] "Pattern recognition (psychology)," in *Wikipedia*, Wikimedia Foundation, 2015. [Online]. Available: https://en.wikipedia.org/wiki/Pattern_recognition_(psychology)#Feature_analysis. Accessed: Mar. 11, 2016.

[6] R. Pydipati, T. F. Burks, and W. S. Lee, "STATISTICAL AND NEURAL NETWORK CLASSIFIERS FOR CITRUS DISEASE DETECTION USING MACHINE VISION," *Transactions of the ASAE*, vol. 48, no. 5, pp. 2007–2014, 2005.

[7] [Online]. Available: http://www.cs.ucr.edu/~eamonn/CE/Bayesian%20Classification%20withInsect_examples.pdf. Accessed: Mar. 11, 2016.

[8] "A machine vision system for quantification of citrus fruit dropped on the ground under the canopy," *Transactions of the ASABE*, pp. 933–946, Aug. 2015.

[9] P. W. Verbeek, H. A. Vrooman, and L. J. Van Vliet, "Low-level image processing by max-min filters," *Signal Processing*, vol. 15, no. 3, pp. 249–258, Oct. 1988.

[10] R. Pandey, S. Naik, and R. Marfatia, "*Image processing and machine learning for automated fruit grading system: A technical review*," *International Journal of Computer Applications*, vol. 81, no. 16, pp. 29–39, Nov. 2013.