Fruit (Banana)- Shelf Life Prediction

Ashwini Pisal
Department of Computer Science
Georgia State University
Atlanta, GA.
apisal1@student.gsu.edu

When we buy fruits, we cannot determine how many days will fruit stay healthy. Due to that uncertainty large amount fruit wastage happens. Also, sometimes we forget to eat fruit when its healthy and eventually it gets spoiled. If we know in how many days fruit will get spoiled, we can control wastage and have that fruit when its nutritious and healthy. We can predict this information if we identify and measure the changes in fruit over time before it gets spoiled. Some fruits might change their color, have dark spots on fruit skin, shrink in volume and some might not have any change but internally they get spoiled. To study this information and to predict shelf life I have considered fruit Banana which is all season fruit and consumed by large mass of population. I have created my own dataset for this analysis

I. Introduction

Fruit wastage occurs mainly because we cannot control and determine when will fruit be spoiled. Once its spoiled, it's no more eatable. For my study I have considered fruit Banana. Different fruits change differently as they ripe. In case of Banana it changes its color as its ripened and eventually has dark spots and turns black when its spoiled. We can observe these changes over a time. In order to study this information, I had collected images of Banana over a few days for different stages of Banana. Starting from raw stage to mid-ripened, ripened, over ripened and finally spoiled Banana. To predict the shelf life of Banana, I am targeting following features.
1.Color based pixel count: number of yellow pixels, brown pixels and black pixels in given Banana image.
2.Spots identification and number of spots on Banana.
3.Percentage of healthy area in Banana
To gather these features, I am following two step approach:
1.Banana Dataset creation
2.Feature set creation using dataset Banana images
Project material link:
https://github.com/jmandivarapu1/CV_Team5/tree/master/Ashwini-Project%20Files

II. Dataset

I have created my own dataset of Banana Images. I had fixed setup to collect the images of Banana. I had followed the following steps:

1. I calibrated my camera
2. Identified the position for lamp such a way that light will be refelcted in all directions. It should not create a shadow over a fruit.
3. I had put white paper sheet on table and placed Banana on it horizontally.
4. I have taken images of Banana's front and back view
5. Banana Images are collected on daily basis till it gets spoiled.
6. Label the images based on day and date.

After trying different postions and methods I have identified this particular setup is suitable for the study. Initially I was hanging my fruit and I was taking pictures in all direction where I was facing shadow and brightness issues and then I changed my setup to this. For my final experiment I had collected Banana images over 10 days and categorised as per the days.

III. Implementation

Once my setup was fixed and I had few days images then I started working on feature extraction. I have used MATLAB tools and functions to do image processing and feature extraction.

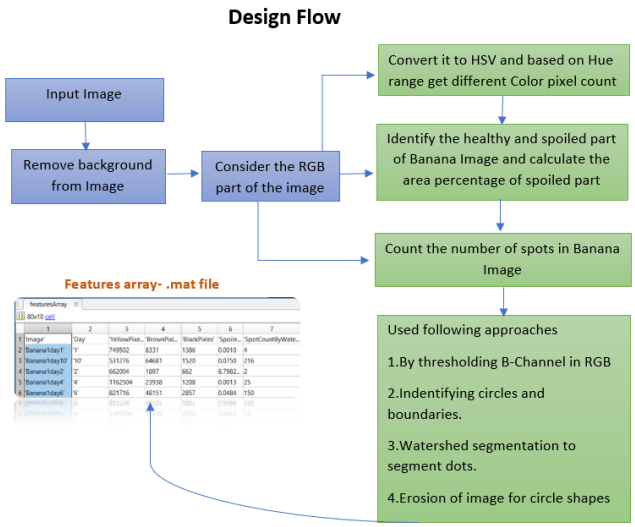

Figure 1: Design flow of algorithm

Background Removal:
As per the flow first I am removing background from images using color thresholder app in MATLAB, I have used LAB and HSV based color thresholding. When I was considering hanging Banana images, object was not getting segmented properly in few directions Banana was partially identified.

Hence, I decided to take horizontal view of Banana. I have set background of my images to gray so that it should not mix with black areas of Banana. All my images have same size and background. This is implemented as separate function in my code.

.

Color feature extraction: I have converted the RGB part of background removed image to HSV. I have thresholde the value of S, V and identified the hue ranges for different colors. Here I am targeting Yellow, Black and Brown colors in image. I have defined this process as function to which I pass RGB part of background extraction process and get each color pixel count.

Spoiled area percentage: I have cleared the boundary regions of RGB image and converted to HSV. I have threshold the V-value (v<0.6) to get black pixels, xor these pixels to get background part of image. Again, I have threshold the H-value to get healthy region of image. By taking a complement of healthy region we get spoiled area. To get the area of actual object I have taken complement of background part of image which will give area of interest, based on that area. I have calculated the amount of area of spoiled part of Banana.

Identify and count the spots: To identify spots I have used various approaches,

1.Threshold B-channel: I have identified the spots by thresholding the B-channel(b<70.5) and clearing the borders, filled the holes to identify the spots and connected hole regions spots are counted.

2. Identifying internal border regions: Convert the RGB image to gray scale. Threshold the gray image assigns black color to all the pixels that have luminosity than a threshold level and others as white. Then complement the image, and it will mark all non-spot region as black and spot region as white based on this only white spot regions are counted.

3 Watershed segmentation: Convert the image to grayscale. Apply top-hat filter with a disk-shaped structuring element to smooth out the uneven illumination. Threshold the image to binary image. compute the distance transform of the complemented binary image, modify it to force the background to be its own catchment basin, and compute the watershed transform. On transformed image spots regions are highlighted and number of spots are counted**.**

4. Erosion: Converted the image to gray scale, identified the structuring element which is sensitive to image I have considered square and applied the erosion to those regions to make those regions prominent and counted the spots in image.

In order to do this processing, I have written different functions targeted to each processing. Results of color extraction, spoiled area percentage and spot identification and count are stored as .mat file which is features array which has additional details such as which Banana image was processed, what day image it is and how many days that Banana lasted.

IV. Testing

To test the input image, I am collecting similar set of features for an input image. Extracted features are compared with generated features array. Features with minimum difference are considered as match. Matched features have the day information and how many days that sample of Banana lasted. Shelf life will be identified as difference between the matched feature, number of days Banana lasted and day-number when sample Banana image was captured. I had considered 100 images of Banana over 10 days for features array creation. I have considered test set of 50 Banana images of different days.

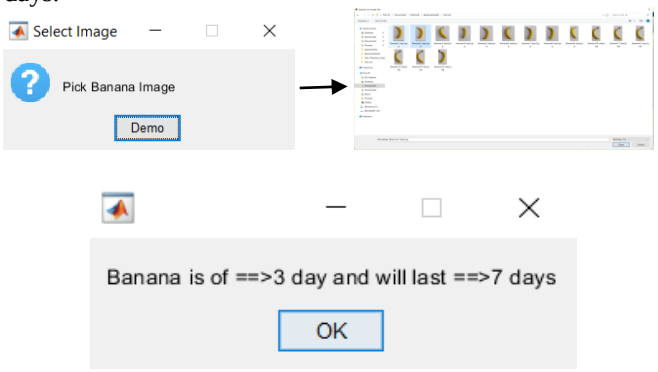

Figure 2: Input selection and output of processing

V. Result

I had performed a test on 50 test images, based on number of correctly predicted images. Overall model accuracy is around 55%. I have considered Mean absolute deviation metric to measure the deviation in actual and predicted day values for an input Banana image. For each image input into the model. Shelf-life value is calculated. The difference between the predicted value and the actual value is calculated. The absolute value of the difference is taken. Finally, the mean of all the calculated absolute deviations is calculated. Smaller MAD value then prediction is better. MAD value varies in the range of 1.05 to 1.25.

VI. Conclusion

Proposed model only works for Banana Images. Its dependent on how images are captured. It is affected by light setting, shadow, brightness, direction in which object is placed. All these factors effect the process at different stages such as background removal. If object has shadow, then it will not be segmented properly. Identifying correct set of Banana samples, because for this study I did not get completely raw Bananas, so I need to choose Banana carefully which are relatively similar in color and shape, if that is not the case then some of the Bananas will be spoiled in 2-3 days. Getting similar set of Bananas was difficult. Identifying spots region and counting spots has errors, it works properly if Banana has larger and prominent spots, for small and initial stage spots it is not able to identify all the spots correctly. This approach can be further enhanced for other fruits which show color or spots as they ripe.

VII. References

[1]TZhoua,A.DHarrison,RMcKellara,J.CYoung,JOdumeru,PPiyasen,XLu,D.GMercer,SKarr "Determination of acceptability and shelf life of ready-to-use lettuce by digital image analysis"

[2]https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=2947&context=the