INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 05-03-2024

Accepted: 07-04-2024

Published: 03-05-2024

Citation: Neethi MV, Raviraj P (2024) Mango Tree Dataset for Yield Estimation: Some Exploratory Analysis. Indian Journal of Science and Technology 17(18): 1906-1913. https://doi.org/10.17485/IJST/v17i18.499

* **Corresponding author**.

neethi.mv@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Mango Tree Dataset for Yield Estimation: Some Exploratory Analysis

M V Neethi[1]*, P Raviraj[2]

1 Assistant Professor, Department of Computer Science & Engineering-Data Science, ATME College of Engineering (Affiliated to VTU, Belagavi), Mysuru, Karnataka, India
2 Professor, Department of Computer Science & Engineering, GSSS Institute of Engineering and Technology for Women (Affiliated to VTU, Belagavi), Mysuru, Karnataka, India

Abstract

Objectives: Employing computer vision methods for yield estimation requires standard datasets. However, unavailability of such a dataset in the literature limits the researchers from evaluating and comparing the performance of their proposed algorithms. In this paper, a benchmark of mango tree dataset is introduced. **Methods:** The dataset is gathered over a 5-month period, starting from the fruit's blossoming stage and ending with its ripening stage. There are 21,000 photos of 4 distinct mango tree varieties in the dataset. From each cultivar, images are captured with different views, distances, and daylight conditions. Further, preprocessing, and exploratory analysis of the dataset are carried out by extracting a few global features such as colors, textures, and histograms for intra-class and inter-class mango trees. **Findings:** The analysis of the collected dataset with different color layers by extracting a few global features and classification of the cultivars of mango trees and, based on the results obtained, the optimal layer of color is attained for the further yield estimation process. **Novelty:** Exploratory analysis of a novel temporal mango crop dataset is executed, and a color analysis classification method is proposed to aid in the early estimation of mango fruit crop yields.

Keywords: Mango Fruit; Yield Estimation; Temporal Dataset; Feature Extraction; Tree Classification

1 Introduction

Precision Agriculture is a notion in farm management that centers on the process of monitoring, quantifying, and adapting to different inputs that introduce variability both within and between fields in contemporary agriculture. The phrase precision agriculture is commonly defined as a technology-enabled farming management technique that monitors, assesses, and evaluates the requirements of individual fields and crops. Precision farming aims to lower input costs, boost output and efficiency, and enhance environmental sustainability[(1)].

One of the main fruit crops in the Karnataka region is mango cultivation. The state grows mangoes on about 1.60 lakh hectares and produces third of all mangoes in India. There is depletion in the yield of mangoes because of unfavourable weather, a decline in soil fertility, and the inability of farmers to plan ahead or expect yields in a timely manner. The primary goals of this work are to encourage precision farming and automate the early estimation of mango crops so that farmers can use it to maximize production while using the least amount of resources. In connection with that, mango crop dataset is built in order to estimate the productivity of the fruit. The images of the mango trees are taken at different times during the course of the season. Data is gathered throughout the fruit's life cycle, from blossoming to maturing. There are 21,000 images of different mango fruit tree cultivars included in the temporal dataset that was created. Mango tree photos are taken with the cameras on mobile phones. Mobile phone cameras are more affordable and practical. Farmers may simply use their phones to take images of trees and consult with specialists to instantly estimate crop production.

Datasets are very important for training the model in order to build the prediction model[2] The lack of temporal data for mangoes, in particular, despite the abundance of ready datasets available in many sources, inspired us to create a mango fruit crop temporal dataset.

Many studies on crop yield estimation have been conducted, however they have made use of a wide variety of datasets with various fruit varieties.[3] Suggested a technique to use past yield data to enhance the grape yield sampling and yield estimation for the current season. To forecast the yield, they have used agricultural data from the past and present. Through their research, it is shown that selecting trustworthy sample locations to enhance grape yield estimation can be facilitated by using localised historical yield data.[4] Collected a citrus fruit dataset to identify orange fruits using a faster region-convolution neural network framework. An olive yield estimation methodology was trained on a set of 37 sample points (74 images), and externally validated on a set of 10 (20 images), manually taken in a super-intensive olive orchard and achieved root-mean-square-error of 0.9914 kg per sample point[5]. Collected Citrus fruit dataset using smartphone field photography and field visible light camera photography, 247 citrus green fruit images were captured for experimentation[6]. Apple fruit detection methodology was tested with a total of 15,335 annotated apples at different growth stages, with diameters varying from 27 mm to 95 mm. Fruit detection results reported an F1-score for apple detection of 0.88 and a mean absolute error of diameter estimation of 5.64 mm[7]. On tree apple fruit yield estimation was carried out using 100 lab imaged apples and was achieved mean absolute error ranging from 1.1 to 4.2 mm for the five-month trial period[8]. The survey indicates that researchers have not used temporal datasets for yield estimation particularly frequently.

In this work we have extracted a few global features to classify the cultivars of mango trees as part of the exploratory analysis of the images of the mango fruit crop dataset. Mango crop cultivars are categorized using a proposed color analysis classification method that involves analyzing images using ten different color channels and extracting eight different attributes.

The organization of this paper is as follows. The methodology is discussed in Section 2. In section 3, the results and discussion are depicted. The last section presents the conclusion and future work.

2 Methodology

In this work, dataset is created from flowering to the ripening stage of the fruit in a temporal manner. This provides an efficient yield estimation rate and further, the introduced dataset can give detailed information about each fruit of the tree, and it also gives the information related to some necessary care to be taken at a certain stage of the fruit growth to improve the overall yield of the mango fruits.

2.1 Dataset

2.1.1 Study area

The current study was conducted in a mango orchard on the college campus (College of Horticulture) located in Yelachachahalli, Yelwala, Mysore, Karnataka, India (139° 07′ 44.29″ east and 35° 12′ 13.01″ north). The region spans more than 200 acres. We have chosen this specific place for our experiments since it is known to have a high yield of mangoes. Uniform fertilizing and management schemes are adopted. The management schemes include timing and amount of fertilizing, watering, weeding, pest, and disease control, harvesting, etc. An average temperature of about 27 °C was recorded over the duration of the investigation.

2.1.2 Dataset collection

Images of mango trees captured in natural daylight using a cell phone's main camera (One plus A5000) comprise the dataset. The resolution of the image is 1920 x 1080 pixels. Four distinct cultivars—Mallika, Alphonso, Arka Arun, and Mulgoa—make up the dataset. For the first three months, images were temporally gathered every seven days for a period of five months. Then, for the following two months, data collection frequency is raised in accordance with fruit maturity. Every cultivar is represented in the 800 photos that are taken each day. Two sessions are used to take the pictures, one in the morning and one in the evening, to correspond with the direction of the sun. Additionally, these trees were photographed at two and three metres away.

The chosen orchid trees range in age from 15 to 20 years old. The trees are about seven and eight feet above the ground. Every image that is obtained has a size of about 7-8 megabytes. The dataset has 21000 photos in total. The dataset is gathered in two sessions per day, one at 9.30 am to 10:00 am and the other from 2.30 pm to 3:00 pm, both in the north-east and south-west directions as shown in Figure 1). As a result, the obtained photos encompass a tree's contour. Top and bottom views, however, are not included in the acquisition. The dataset was gathered over a five-month period.

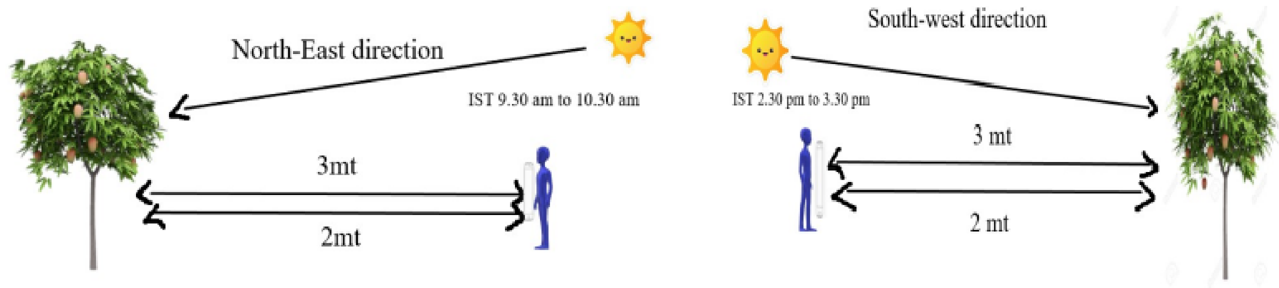


Fig 1. Image Acquisition

2.1.3 Dataset Analysis

The images are processed into various colour channels, namely, Grey, Red, Green, Blue, Yellow, Chroma Blue, Chroma Red, Hue, and Saturation. The colour channels for the tree are shown in Figure 2.

From each channel, features namely mean, variance, contrast, correlation, energy, homogeneity, slope, and offset are extracted, which are correspondingly shown in the Equations (1), (2), (3), (4), (5), (6), (7) and (8).

$$Mean = \frac{1}{N}\sum_{i=1}^{N} Ai \tag{1}$$

$$Standard\ deviation = \sqrt{\sum \frac{|x-\mu| * |x-\mu|}{n}} \tag{2}$$

$$contrast = \sum_{i,j=0}^{N-1} pij\,((i-j)*(i-j)] \tag{3}$$

$$correltion = \sum_{i,j=0}^{N-1} pij\frac{(i-\mu)\,(j-\mu)}{\sigma*\sigma} \tag{4}$$

$$Energy = \sum_{i,j=0}^{N-1} (pij*pij) \tag{5}$$

$$Homogenity = \sum_{i,j=0}^{N-1} \frac{pij}{1+(i-j)*(i-j)} \tag{6}$$

$$Slope\ (Regression) = b+ax \tag{7}$$

$$Offset\ (Regression)\ =\ b+ax+\log\ (t) \tag{8}$$

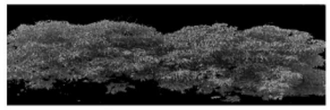
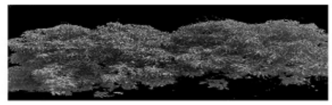
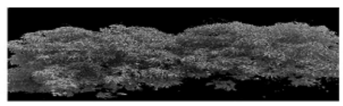
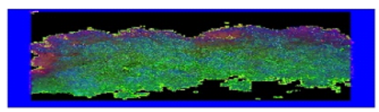
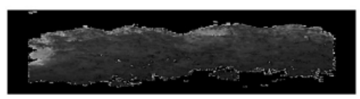
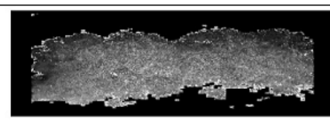
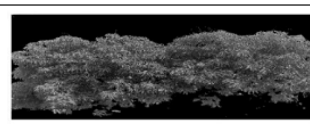
Colour Spaces	Image
Grey	
RGB	
Red	
Green	
Blue	
HSV	
Hue	
Saturation	
Value	

Fig 2. Different Colour Layers

For classifying the trees, using the extracted features is found to be expensive. Therefore, to ease out the process, we have transformed them into three common features, namely minimum, maximum, and mean. As a result, each tree is represented with these three features. The Euclidean distance measure is used to compute the similarity for tree classification. The collected dataset of mango trees can be represented as follows:

DiSjMkTlCm = {Min,Max,Mean}

where,

D= days

i = number of days

S= sessions

j = number of sessions
M= meter
k= the distance from which the image is captured
T=trees
l = number of trees
c= count

3 Results and Discussion

The first step in the process of segmenting the fruits from mango trees is to segment the image of the tree. To obtain a higher accuracy rate, it is crucial to segment the tree because the photos of the tree contain the sky, the ground, and other background elements. To segment trees, a straightforward statistical method based on thresholds is used. Using Gaussian blur filters, the noise in the picture is eliminated. Eight attributes are then taken out of the segmented tree in order to classify the tree cultivars. As mentioned in the previous section, we have selected 4 cultivars of mango trees. The notion is to classify them with the extracted features. Based on the classification results, the best colour layer is selected for further processing of mango yield estimation.

Based on the results given in the experimental results section, hue, saturation and value (HSV) colour spaces are giving good results. The HSV colour space attempts to characterize colours according to their hue, saturation, and value. The hue of a colour identifies what is commonly called "colour". The saturation of a colour identifies how pure or intense the colour is, which helps to distinguish between foreground and background objects present in the image. A fully saturated colour is deep and brilliant. As the saturation decreases, the colour gets paler and more washed out until it eventually fades to neutral. Ranges from 0-100%. Also, sometimes called the "purity" by analogy to the colourimetric quantities excitation purity and colourimetric purity. The lower the saturation of a colour, the more "greyness" is present and the more faded the colour will appear, thus useful in identifying the mango fruits in the image. The brightness of a colour identifies how light or dark the colour is. Any colour whose brightness is zero is black regardless of its hue or saturation. Future work on mango fruit yield estimation will be carried out with HSV colour space. Edge detection is the next step in the pre-processing step of the mango fruit dataset. Texture segmentation using the texture filters method is applied to obtain the edges of a masked image as shown in Figure 3.

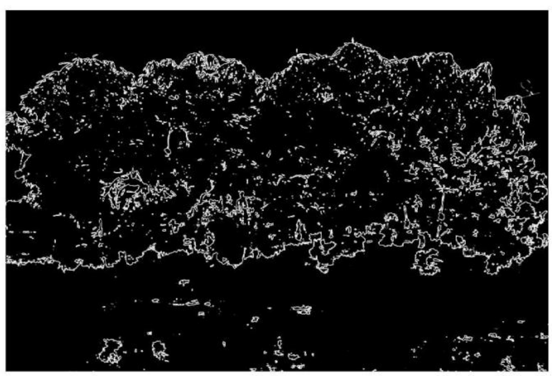


Fig 3. Edge Detection of Masked Image

Considering the fact that the images were captured in open areas, due to variations in the illumination conditions, and the plant canopies are irregular, it is mandatory to adjust the brightness. Histogram equalization is used for this purpose. The most challenging step of digital image processing is segmentation. Segmentation is more difficult when the image acquisition occurs in a bizarre state. It was possible to highlight the fruits in the images of mango trees canopies through a combination of the techniques of colour model conversion, thresholding, Gaussian blur and spatial filtering with Laplace and Sobel masks as shown in the Figure 4. CNN based segmentation methods are widely used in segmentation of mango crops where similar pixels are classified into individual classes[9].

The fruits are highlighted as in bas-relief sculptures. The automatic generation of these representations enables the use of a method of recognition and interpretation for the number of visible fruit estimation. This leads to the first step of the development of the automatic system for mango fruit yield estimation.

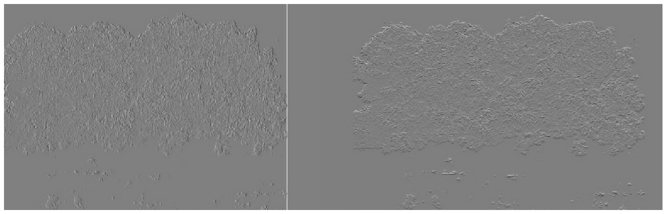


Fig 4. Highlighting the Green Fruits on the Basis of Texture and Brightness Analysis

The intra class analysis of the collected data is given in Table 1. The results are obtained by calculating the distance of each colour layer of one image to another image, Euclidean distance measure is used to find the distance. From the obtained distance matrix, the results are consolidated into minimum, maximum and mean values, which are shown in Table 2.

The inter class analysis results of hue colour channel are given in Table 3. The classification of different cultivars of mango trees can be observed in the mean value. To experiment, tree 2 and tree 3 were taken from the same cultivars and the results show the accurate classification.

The proposed work is compared with DL based semantic segmentation method, in this work total of 123 tomato images were taken annotated and augmented during preprocessing stage though the proposed work uses 21000 images obtained from the mango orchards and detailed exploratory analysis is carried out [10].

Table 1. Eight Features of a Single Image

Layer	Mean	Variance	Contrast	Correlation	Energy	Homogeneity	Slope	Offset
Grey	23.3311	1863.489	0.656922	0.75694	0.517395	0.922579	0.75722	0.00127
Red	24.09756	1981.556	0.702049	0.759355	0.516127	0.920962	0.7401187	0.001307
Green	24.09191	1960.334	0.68298	0.7661	0.511955	0.9214547	0.739219384	0.001314
Blue	17.4112	1307.482	0.570924	0.700779	0.598488	0.9285906	0.558095912	0.000798
Yellow	36.03807	1374.406	0.640473	0.768777	0.507108	0.9237582	0.238990036	0.002336
Cr	125.0772	46.59415	0.040151	0.885174	0.611794	0.9799244	0.254581565	0.005831
Cb	128.4759	4.526696	0.024591	0.843572	0.818807	0.9877043	0.819012271	0.005857
H	0.065706	0.023017	0.524162	0.769725	0.601269	0.9654611	0.750458163	0.000962
S	0.10269	0.043278	0.610761	0.862507	0.56758	0.9265762	0.731027746	0.001504
V	0.099597	0.033258	0.7797	0.757786	0.507773	0.9195295	0.735144	0.001364

Table 2. Results of Intra-class Analysis of Single Image

Layer	Minimum	Maximum	Mean
Grey	0.158119	713.0967	234.7053
Red	0.259248	782.6599	270.0982
Green	0.137552	696.3706	219.1873
Blue	0.37265	641.7406	234.0963
Yellow	0.137872	525.8526	173.0826
Cr	0.024116	11.57299	3.980401
Cb	0.002848	2.486015	0.963413
H	0.000263	0.290008	0.133179
S	0.001558	0.515006	0.22447
V	0.002444	0.513729	0.194604

Table 3. Results of Inter-class Analysis of hue Colour Layer

Tree	Minimum	Maximum	Mean
1	0.158119	713.0967	234.7053
2	0.093056	291.9663	99.10925
3	0.136429	250.2184	95.03804
4	0.368601	421.4321	162.3574
5	0.780026	340.3281	113.0418

3.1 Comparative Analysis

Table 4 depicts suggested and current comparative analysis techniques that are compared to various techniques used for exploratory data analysis.

Table 4. Comparative Analysis

Methods	Fruit	Analysis
Exploratory data analysis for disease prediction [9]	Pomegranate	Classification methods
Exploratory data analysis for yield estimation [10]	Blueberry	Bayes and function models
Proposed Exploratory Data Analysis for Yield Estimation	Mango	Color analysis models

The proposed work is compared with other works of exploratory analysis of fruit crop dataset [11,12]. Based on the comparison, it is stated that yield estimation requires a lot of data for many machine learning methods. The availability of training data in appropriate quality and quantity continues to be a limiting factor in the training and experimentation of machine learning algorithms. To improve training for yield estimation models, the proposed work primarily focuses on creating a large temporal dataset of mango crops. It also suggests a color analysis classification model for classifying mango cultivars, which has the potential to achieve results as high as 98% accuracy. In contrast, other works have analyzed the collected dataset using Bayes and function models, focusing primarily on optimal features and dimensionality reduction of collected features for more accurate classification.

Mango fruit cultivar classification can be done using the suggested color analysis method for mango tree cultivar classification.

4 Conclusion

In conclusion, the work proposes exploratory analysis and a novel color analysis classification method of a temporal mango crop fruit dataset. In order to enhance the images for the ensuing processing of the mango crop yield estimation and analysis, the dataset has been preprocessed using image processing techniques such edge detection and noise reduction. To learn about the color distributions of the items in an image, various color channels are explored. For the four distinct mango tree cultivars included in the dataset, the hue color channel yielded superior results for both intra- and inter-class classification, with inter-class mean values of 234.70, 99.10, 95.03, 162.35, and 113.04.

Estimating the crop yield of mango fruit is made possible by the work that follows. The dataset, which includes data on the mango fruit at its beginning, middle, and end stages, aids researchers in their examination of the production of the mango tree at every step. A lot of new innovations pertaining to mango fruit yield estimation can be made possible using this dataset.

References

1) Rimpika, Anushi, Manasa S, Anusha KN, Sharma S, Thakur A, et al. An Overview of Precision Farming. *International Journal of Environment and Climate Change*. 2023;13(12):441–456. Available from: https://dx.doi.org/10.9734/ijecc/2023/v13i123701.
2) Patil P, Athavale P, Bothara M, Tambolkar S, More A. Crop Selection and Yield Prediction using Machine Learning Approach. *Current Agriculture Research Journal*. 2023;11(3):968–980. Available from: https://dx.doi.org/10.12944/carj.11.3.26.
3) Araya-Alman M, Leroux C, Acevedo-Opazo C, Guillaume S, Valdés-Gómez H, Verdugo-Vásquez N, et al. A new localized sampling method to improve grape yield estimation of the current season using yield historical data. *Precision Agriculture*. 2019;20(2):445–459. Available from: https://dx.doi.org/10.1007/s11119-019-09644-y.
4) Galdames JPM, Milhor CE, Becker M. Citrus fruit detection using Faster R-CNN algorithm under real outdoor conditions. In: Proceedings of the 14th International Conference on Precision Agriculture. ;p. 1–8. Available from: https://ispag.org/proceedings/?action=abstract&id=5161&title=Citrus+Fruit+Detection+Using+Faster+R-CNN+Algorithm+Under+Real+Outdoor+Conditions&search=types.

5) Aquino A, Ponce JM, Noguera M, Andújar JM. Olive-fruit yield estimation by modelling perceptual visual features. *Computers and Electronics in Agriculture*. 2023;214. Available from: https://dx.doi.org/10.1016/j.compag.2023.108361.
6) Lu J, Chen P, Yu C, Lan Y, Yu L, Yang R, et al. Lightweight green citrus fruit detection method for practical environmental applications. *Computers and Electronics in Agriculture*. 2023;215. Available from: https://dx.doi.org/10.1016/j.compag.2023.108205.
7) Ferrer-Ferrer M, Ruiz-Hidalgo J, Gregorio E, Vilaplana V, Morros JR, Gené-Mola J. Simultaneous fruit detection and size estimation using multitask deep neural networks. *Biosystems Engineering*. 2023;233:63–75. Available from: https://dx.doi.org/10.1016/j.biosystemseng.2023.07.010.
8) Mirbod O, Choi D, Heinemann PH, Marini RP, He L. On-tree apple fruit size estimation using stereo vision with deep learning-based occlusion handling. *Biosystems Engineering*. 2023;226:27–42. Available from: https://dx.doi.org/10.1016/j.biosystemseng.2022.12.008.
9) Neupane C, Koirala A, Walsh KB. In-Orchard Sizing of Mango Fruit: 1. Comparison of Machine Vision Based Methods for On-The-Go Estimation. *Horticulturae*. 2022;8(12):1–17. Available from: https://dx.doi.org/10.3390/horticulturae8121223.
10) Maheswari P, Raja P, Hoang VT. Intelligent yield estimation for tomato crop using SegNet with VGG19 architecture. *Scientific Reports*. 2022;12(1):1–11. Available from: https://dx.doi.org/10.1038/s41598-022-17840-6.
11) Raj SB, Shankar G, Murugesan S, Raju MN, Mohan E, Rani PJI. Exploratory Data Analysis on Blueberry yield through Bayes and Function Models. *International Journal on Recent and Innovation Trends in Computing and Communication*. 2023;11(11s):634–641. Available from: https://dx.doi.org/10.17762/ijritcc.v11i11s.8299.
12) Nirgude V, Rathi S. Exploratory Data Analysis and Optimal Feature Selection on Pomegranate Dataset for Enhancing the Performance of Pomegranate Disease Prediction." CERN European Organization for Nuclear Research - Zenodo, 2022. *Annals of Forest Research*. 2022;65(1):746–764. Available from: https://www.e-afr.org/login/pdf/Vaishali%20Nirgude.pdf.