Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 78 (2016) 691 – 697

Procedia Computer Science

International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Nagpur, INDIA

Orange Sorting by Applying Pattern Recognition on Colour Image

Jyoti Jhawar *

M.Tech. Student, J.D. College of Engg. & Mgmt., Nagpur,Maharashtra, India

Abstract

Manual sorting/grading of oranges is done at wholesale markets/ food processing factories based upon its maturity, size, quality and breeds. With an aim to replace the manual sorting system, this paper proposes the research work for automated grading of Oranges using pattern recognition techniques applied on a single color image of the fruit. This research is carried out on 160 Orange fruits collected from varied geographical locations in Vidarbha Region of Maharashtra. System designed can automatically classify an Orange fruit from this region, given its single color image of 640 × 480 pixel resolution, taken inside a special box designed with 430 lux intensity light inside it, by a digital camera. Only 4 features are used to classify oranges into 4 different classes according to the maturity level and 3 different classes as per size of oranges. In this paper two novel techniques based on Pattern Recognition are proposed – Edited Multi Seed Nearest Neighbor Technique and Linear Regression based technique; although Nearest Neighbor Prototype technique is also deployed. Linear Regression based technique can explicitly predict the maturity of the unknown orange fruit, enabling classification into multiple classes with desired lifespan. Experimental results indicate success rate up to 90 % and 98 %.

Peer-review under responsibility of organizing committee of the ICISP2015

Keywords: Orange Sorting; Pattern Recognition; Color Image Processing.

1. Introduction

At present, most farms & food industries use manual experts for sorting of the fruits which is time consuming, laborious, and suffers from the problem of inconsistency and inaccuracy in judgment by different human experts[1].

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
*E-mail address:*author@institute.xxx

Peer-review under responsibility of organizing committee of the ICISP2015
doi:10.1016/j.procs.2016.02.118

"Vidarbha" region in Maharashtra is well known for the production of citrus fruits; especially oranges. Maharashtra owns 2nd rank in the country for production of Sweet Limes[17]. Yet, no any automated sorting, packaging and processing machines are developed for Indian Breeds of Citrus Family. National Research Center for Citrus (NRCC) have developed a mechanical sorting machine which sorts' Oranges mechanically size wise, but it doesn't sort fruits quality/grade wise[22]. Some research work is carried out for automated classification of verities of mangoes, tomatoes, cranberries & dates[1, 14, 11, 15]; but no remarkable research work is carried out for Citrus fruits classification especially Indian varieties.

Color and size are the most important features for accurate classification and sorting of citrus[2]. The maturity of the fruit can be predicted by its color; and based upon the maturity level / life span market for the fruit can be decided[9]. This research is carried out with an aim to design a pattern recognition based automated Orange sorting software. System designed can automatically classify an Orange fruit based on its maturity, given its single color image of 640 × 480 pixel resolution taken inside a special box designed. Only 4 features are used to classify oranges into four different classes according to maturity level and it can also predict the size of the fruit. In this research two novel techniques based on Pattern Recognition are proposed – Edited Multi Seed Nearest Neighbor Technique and Linear Regression based technique; although Nearest Neighbor Prototype technique is also deployed. Linear Regression based technique can explicitly predict the maturity of the unknown orange fruit, enabling classification into multiple classes with desired lifespan. The software developed can further be embedded in an automated machine, like the one designed for the grading of mangoes[1].

1.1. Pattern Recognition System

The design model of a pattern recognition system essentially involves the following three steps[15, 7].

1) *Data acquisition and preprocessing*: Here the data from the surrounding environment is taken as an input and given to the pattern recognition system. The raw data is then preprocessed by either removing noise from the data or extracting pattern of interest from the background so as to make the input readable by the system.

2) *Feature extraction*: Then the relevant features from the processed data are extracted. These relevant features collectively form entity of object to be recognized or classified.

3) *Decision making*: Here the desired operation of classification or recognition is done upon the descriptor of extracted features.

2. Methods and Materials

2.1. Data Acquisition

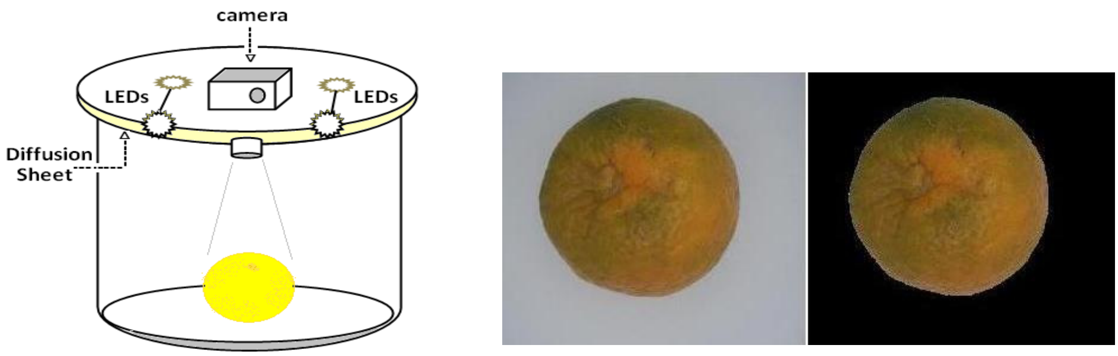
Colour Image of orange fruit is the input to the system designed. It is found that the images of a same object vary with Source of light, Intensity of light, background, distance from Camera, and Camera settings etc[19]. To maintain the consistency in all the fruit samples, an imaging chamber is designed as shown in fig. 1(a). A white colour cylindrical plastic box is used as imaging chamber. Base of the box is coated with white paper, to reduce reflections from the base. The inner surface of the box is coated with light reflexive material. LEDs mounted on the top are used as source of light inside the imaging box; which is then covered with diffusion Sheet. UPS power supply is used to avoid voltage fluctuations resulting in constant light intensity inside the box; which measured 430 Lux using digital Lux meter. Light intensity used follows Hunter Labs standard of diffused day light[20]. Camera used is DSC 2000 Sony Camera in VGA mode (640 x 480), with light setting in auto mode, flash off & imaging angle is 0º. Distance of a fruit from the camera is 18 cms for all the images taken.

2.2. Sample Collection

This research is carried upon 160 Orange fruit samples which were collected from Kalamana Market & farms of Bhandara district, under guidance of ‘Krishi Utpann Bazar Samiti, Kalmana’, Nagpur. These samples are collected in 4 slots with a time difference of 1month for each slot. ‘Mrig Baar’ (production in April-June) and ‘Amiya baar’ (Production in November –December) fruits are collected in order to get all the variety of Oranges.

Oranges selected are from various geographical regions; Amravati, Nagpur, Katol and Bhandara. Different set of samples are used for Training & Testing. Training Sample space consists of 80 ‘amiya baar’ samples, and Test sample space is of 80 samples collected in ‘Mrig baar’ and ‘Amiya Baar’ both. Samples chosen are classified into four classes/grades by two human experts (Horticulture specialists) as per the maturity. The classes are ‘Not Ripe’, ‘Semi Ripe’, ‘Ripe’, and ‘Over Ripe’; almost equal number of samples fell into each manual class mentioned above.

‘Not Ripe’ class includes Orange fruit samples which will never ripe, which has a sour taste and sugar content is very low; these are not preferred in culinary uses. ‘Semi Ripe’ class contains Oranges which are half ripened; life span of these can be up to one month. ‘Ripe’ class stands for absolutely ripe Oranges mostly used for culinary purposes with life span of at most 5 days. ‘Over Ripe’ indicates Oranges with life span of hardly a day after which its’ tangy- sweet test gets distorted. First step of pattern recognition is data acquisition; this is achieved by taking color photograph of each fruit sample, taken in the imaging box designed in the said controlled environment.



(a) Imaging Chamber (b) Back ground Removal

Fig. 1. (a) Imaging Chamber; (b) Back ground Removal.

2.3. Preprocessing & Background Removal

As the images are taken in controlled environment & fix background, no blur/ haze is found; hence any pre-processing is not needed. The background chosen is white for proper reflection of the object color; still due to the color reflection from Orange, slight variation in the background color is observed. Value Threshold Algorithm is used to segment the object and remove the background as shown in fig. 1(b). Threshold cut off value is decided by taking mean of the four pixels values near four corners in the image, coding is done using MATLAB 10(b). This segmented image of fruit object extracted is scanned and only four key features are extracted. These features are Tot_pixel, Ravg, Gavg and Bavg,

- Tot_pixel = ∑ pixel in Obj. - - > This is total number of pixels in the segmented sample fruit.
- Ravg = ∑pixel [R]/Tot_pixel - - > Arithmetic mean of Red color in the fruit.
- Gavg = ∑pixel [G]/Tot_pixel - - > Arithmetic mean of Green color in the fruit.
- Bavg – ∑pixel [B]/Tot_pixel - - > Arithmetic mean of Blue color in the fruit.

2.4. Pattern Recognition Techniques Used For Classification

Following pattern recognition techniques are deployed on training data set using four key features for classification of Oranges; and outcomes of this are embedded in the software designed.

1) Nearest Neighbour Prototype-

This is similar to the single nearest neighbor technique, except that only one typical sample from each class is used for the reference set of possible neighbors[16]. This class prototype (reference point) can be chosen as typical or ideal cases or as the average feature vector of all the members of the class. In this research the reference point for each class is chosen as average feature vector of all the training samples in that class. For ex. reference seed point for Ripe class is (Rripe, Gripe, Bripe) where

- Rripe = ∑ Ravg / No_samples
- Gripe = ∑ Gavg / No_samples (Arithmetic Mean)
- Bripe = ∑ Bavg / No_samples

Hence we get four reference vectors for 'Not Ripe', 'Semi Ripe', 'Ripe', and 'Over Ripe' classes. The test sample is classified as belonging to the class of the closest prototype. Any distance metric and feature scaling techniques can be used for calculation of distances of an unknown sample from the classes designed[16, 15]. City Block distance metric found to be most suitable to calculate the nearest neighbor as compared to Euclidean metric, as it needs lesser computational time.

2) Edited Multi-Seed Nearest Neighbour Technique

The classes chosen are broad, hence the cluster of each class is big and scattered; also objects near the decision boundary are miss- classified at times. A new technique is invented, by blending three different techniques 1) K-nearest Neighbor, 2) Edited Nearest Neighbor, and 3) K-means. Multiple seed points / feature vectors for each class are evaluated based upon the breeds/ variety of Oranges. For each class it was observed that, the training samples inside that class are clustered towards two different points; these are selected as two seed points for this class. Applying this we get two seed point vectors for each class, hence in all 8 seed points. K-Means algorithm is used for evaluation of these two seed points for each class. Using City Block distance metric, distances of the unknown sample from all four different classes are evaluated. The distance of an unknown sample from a class is sum of its distance from both seed points of that class. The unknown sample is classified into the class which is nearest to the sample.

3) Linear Regression

Predicting the maturity /ripeness measure of the fruit is not achieved in earlier two techniques used as the classes are broad. Linear Regression technique used can not only classify an unknown orange fruit sample but can also predict the maturity level of it. Linear equations can be written using 3 features Ravg, Bavg, Gavg and maturity / ripeness measure as below.

$$\text{Ravg} * X + \text{Gavg} * Y + \text{Bavg} * Z = \text{Ripeness\_measure} \tag{1}$$

In this equation X, Y and Z are the coefficients for Red, Green and Blue color value respectively in a fruit. Ripeness_measure is the % of maturity of that fruit. The Ripeness measure for each fruit is decided by two human experts. Various equations are solved and values of X, Y, Z are evaluated. Any unknown Orange fruit sample's maturity can be predicted by solving equation (1). Fuzzy rule set is designed to classify the Oranges according to the maturity level, into the four different classes.

2.5. Prediction about the Size

'Tot_pixel' feature is used to predict the size of the Orange into 'Small', 'Medium' or 'Large' classes. 'Small' class Orange is less than 5cm in diameter; ‘Medium’ class’s diameter is in between 5 cm to 6.5 cm and ‘Large’ indicates oranges with diameter above 6.5cm which is preferred for exports. Training set orange fruits’ mean diameter is measured using Vernier caliber and then correlation between size and ‘Tot_pixel’ is established, as ‘Tot_pixel’ indicates area of the fruit. Fuzzy Logic is used for prediction of the size. Combining 3 classes of size and 4 classes of maturity level, total 12 classes can be defined; hence we can predict oranges into 12 classes.

3. Results & Discussions

Because of blending Background Removal and Feature Extraction algorithms together, only one scan of the whole image is needed, optimizing the computational time required. VGA images are with least resolution (640 x 480) hence processing time required is very less. More over unlike the other researcher works original color image is directly processed over three matrices for Red, Green and Blue color, resulted in reducing number of iterations with simplified code. Features extracted proved to be relevant to predict the ripeness measure and Grade/class of the fruit, along with the size of the fruit. Test sample set of 80 oranges, is used for testing the software signed; the success rate of each classification technique used can be seen in Table 1. Linear Regression can further classify the fruits as per the maturity/ ripeness measure into many classes.

Table.1 Successes of Techniques Used

Pattern recognition Technique	Nearest Prototype	Edited multi-Seed Nearest Neighbor	Linear Regression
Success %	92.93 %	89.90 %	97.98 %

The statistical analysis of the training data set is shown in Table. 2. It is observed that the standard deviation of Blue colour is very less, also its range is in between 29 to 48 for all samples; indicating Blue color is least significant for prediction and hence can be ignored. The standard deviation of Red and Green color, for some classes is above 10 indicates future scope of sub class creation.

Table. 2 Statistical Analyses of Oranges.

Class & no of samples		Red	Green	Blue
	Mean	64.21	76.57	41.64
Not Ripe (20 samples)	Standard Deviation	12.99	13.70	5.51
	Min , Max	39, 83	51, 97	33, 48
	Mean	85.66	88.33	39.00
Semi Ripe (20 samples)	Standard Deviation	10.51	7.88	3.81
	Min , Max	67,96	77, 99	32, 47
	Mean	120.26	94.80	34.66
Ripe (20 samples)	Standard Deviation	13.46	4.81	1.71
	Min, max	89, 141	83,102	32, 38
	Mean	128.93	88.60	33.53
Over Ripe (20 samples)	Standard Deviation	9.75	7.21	3.77
	Min, max	104, 145	77, 99	29, 44

As it can be observed from the plot below in Fig. 2, difference in Red and Green is very less for ‘Semi Ripe’ class. But over with increasing ripeness, the value of Red goes on increasing and slow growth in Green color is also seen. Except a few exception cases the difference between Red and Green is high in Ripe, Over Ripe classes. Blue color value decreases with ripeness.

Various curve fitting graphs are plotted as in fig. 3, blue color is neglected for plotting as its range is very small. The best curve fitting the points in the plot is power function; but value of R^2 (co-relation coefficient) is not above 0.9, hence this indicates that both the parameters Red and Green are essential for predication of Ripeness Measure.

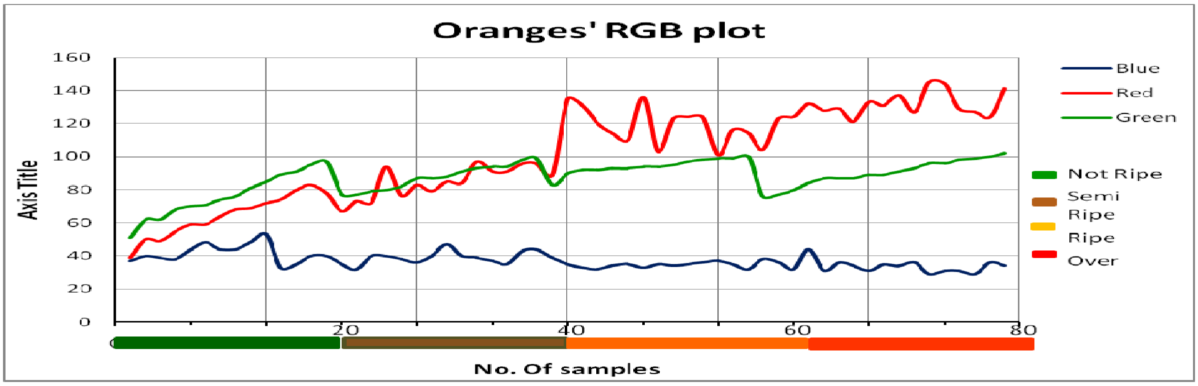


Fig. 2. Red, Green, & Blue plot of Oranges- class-wise

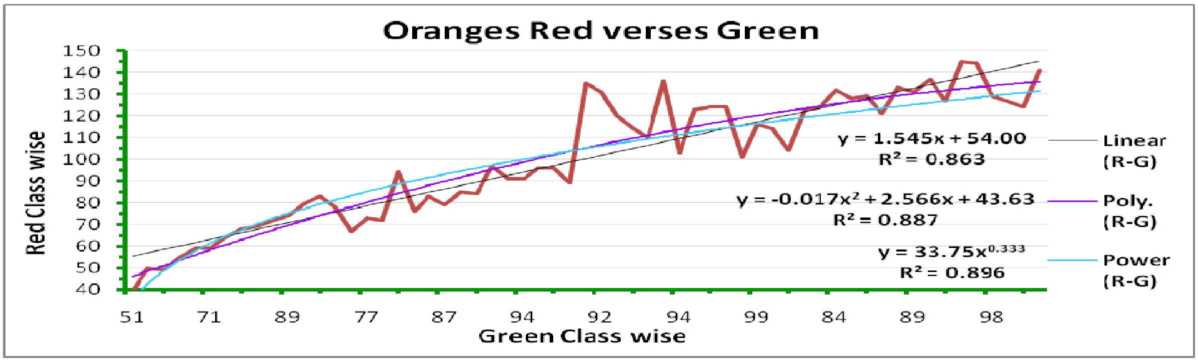


Fig.3. Red-Green Correlation Curve Fitting Plot of Oranges

4. Conclusion and Future Scope

This research showed that even four key features are sufficient to classify an Orange fruit. As it is observed that, Blue colour is least significant hence can be neglected for classification & maturity prediction; further reduces the key features required for prediction to three. Out of various techniques used Linear Regression proved to give best results and this technique can be further explored to predict the life span of the fruit. Edited Multi-Seed Nearest Neighbour technique designed is more or less equal effective like that of Nearest Neighbour Edited Multi-Seed technique devised can be more effective if used with more than two seed points.

Oranges doesn't ripe in uniform manner hence the image of other surface is need for proper prediction. All the misclassifications while testing are found to fall in this category. Hence future scope lies in using two images of opposite surfaces rather than using single colour image of a fruit (lemon sorting & cranberry sorting[2, 14])

Damaged Orange fruit detection is one big challenge which not handled in this research. It was observed that damages like Bruises, Rugs, Fungal Infections, Hail Stone marks, etc. produces marks on the skin; hence these can be also detected applying some advance algorithm. Further scope lies in detecting varieties like 'Battidar'(tight skin) and 'phopsa'(loose skin) Oranges.

References

1. Chandra Sekhar Nandi, Bipan Tudu, and Chiranjib Koley, "A Machine Vision-Based Maturity Prediction System for Sorting of Harvested Mangoes", in IEEE Transactions on Instrumentation and measurement, vol. 63, no. 7, july 2014, p.p.1722-1730J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–732.
2. M. Khojastehnazhand, M. Omid* and A. Tabatabaeefar, "Development of a lemon sorting system based on color and size", in African Journal of Plant Science Vol. 4(4), April 2010, pp. 122-127. Available online at http://www.academicjournals.org/ajps, ISSN 1996-0824
3. A.R. Jiménez, A.K. Jain, R. Ceres and J.L. Pons," Automatic fruit recognition: A survey and new results using Range/Attenuation images", Pattern Recognition, 1999, vol. 32 issue 10, pp. 1719-1736.
4. S.Arivazhagan, R.Newlin Shebiah, S. Selva Nidhyanandhan, L.Ganesan, " Fruit Recognition using Color and Texture Features ", Journal of Emerging Trends in Computing and Information Sciences vol. 1, no. 2, 2010, pp. 90 – 94. E-ISSN 2218-6301
5. Monika Sharma [1], Vibhuti P N Jaiswal [2], Amit Goyal, "Fruit Recognition with Multiple Features using Fuzzy Logic", IJCSC Vol. 4, No. 2 September 2013 pp.99-115. ISSN-0973-7391
6. Sachin Syal, Tanvi Mehta, Priya Darshni, " Design & Development of Intelligent System for Grading of Jatropha Fruit by Its Feature Value Extraction Using Fuzzy Logics", International Journal of Advanced Research in Computer Science and Software Engineering Vol. 3, No. 7, July 2013 ISSN: 2277 128X
7. Pragati Ninawe, Mrs. Shikha Pandey, "A Completion on Fruit Recognition System Using K-Nearest Neighbors Algorithm", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Vol. 3, No. 7, July 2014, pp. 2352 – 2356. ISSN: 2278 – 1323
8. Priyanka Sharma, Manavjeet Kaur, " Classification in Pattern Recognition: A Review", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Vol. 3, No. 4, April 2013 ISSN: 2277 128X
9. Youssouf Chherawala, Richard Lepage, Gilles Doyon, "Food Grading/Sorting Based on Color Appearance trough Machine Vision: the Case of Fresh Cranberries", IEEE , 0-7803-9521-2/06/2006.
10. Chandra Sekhar Nandi, Bipan Tudu, and Chiranjib Koley, " An Automated Machine Vision Based System for Fruit Sorting and Grading", IEEE Sixth International Conference on Sensing Technology (ICST) 978-1-4673-2248-5/12, 2012. Page(s): 195 – 200
11. Youssouf Chherawala, Richard Lepage' and Gilles Doyon, "Food Grading/Sorting Based on Color Appearance trough Machine Vision: the Case of Fresh Cranberries", IEEE Conference publications, ICTTA 2006, pp. 1540 – 1545. DOI: 10.1109/ICTTA.2006.1684612
12. Hamid AliMohamadi , Alireza Ahmadifard, "Detecting Skin defects of fruits using Optimal Gabor Wavelet Filter", IEEE International Conference on Digital Image Processing, 2009, pp. 402 - 406. DOI: 10.1109/ICDIP.2009.76
13. Trung Pham, Per Bro and José Luis Troncoso, In "Fuzzy Analysis of Color Images of Fruits". IEEE 2nd International Conference on Information Science and Engineering (ICISE), 2010, pp. 3714 – 3717. DOI: 10.1109/ICISE.2010.5689230
14. Zhang Ruoyu , Kan Za , Jiang Yinlan, " Design of Tomato Color Sorting Multi-channel Real-Time Data Acquisition and Processing System Based on FPGA ",International Symposium on Information Science and Engineering (ISISE), 2010, pp. 207 – 212. DOI: 10.1109/ISISE.2010.40
15. Abdulhamid Haidar, Haiwei Dong and Nikolaos Mavridis, "Image-Based Date Fruit Classification", IEEE Conference on Ultra modern Telecommunications and Systems 2012, pp. 369 – 375.
16. Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification", Second Edition, Wiley publications, 2007, ISBN 81-265-1116-8.
17. Earl Gose, Richard Johnson, Steve Jost, " Pattern Recognition & Image Analysis", Pearson Publication, 2007, ISBN 81-317-1537-X
18. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3831687/Date & Time of Accession: 11th April 2015, 7pm.
19. http://www.historyindia.com/show, show names – 'Vital Stats Of India', 'Food CIA' & 'Unwrapped'
20. www.hunterlab.com
21. http://www.slideshare.net/vasanthcgowda/lemon-production-in-india-and-karnataka, accessed on 21-10-2015, 10am.
22. *www.**nrcc**itrus.nic.in*