IJAAR PUBLISHING®
DOI: 10.46654

African Journal of Sustainable Agricultural Development | ISSN: **2714-4402**
Vol. 2, Number 4 (October-December, 2021) | www.ijaar.org

CLUSTERING METHOD OF TOMATO RIPENESS IDENTIFICATION (REVIEW)

Nura Muhammad Baba
Department of Office Technology and Management, Kano State Polytechnic, Kano, Nigeria.
Email: nabnurtrn@gmail.com

ABSTRACTS

This paper reviewed related works regarding clustering method of fruit (mainly tomato) ripeness identification. Most of the literatures reported in this paper used unsupervised learning techniques. Fruit ripeness identification become necessary in order to reduce high labour cost, labor fatigue, lack of trained labour, misclassification error caused by graders/experts that resulted in producing low quality farm product. Most researchers mainly used k – means, algorithms, nearest neighbor, otsu, threshold, Gaussian density function, fuzzy methods and achieve more than 80% success recognition rate. Almost all the researchers used two categories of fruit RT, and UT. In the present study, we will identify ripe, unripe, overripe, and under ripe varieties of tomato. Based on the reviewed literatures we concluded that our proposed methodology will use five algorithms employed by most researchers that achieved high recognition rate and compare the results to either integrate or come up with new algorithms/colour models to achieve higher recognition rates than previously done work.

Keywords: Tomato, otsu, fuzzy, k – means, Gaussian density function, nearest neighbor, clustering.

1.0 Introduction

Clustering is the process of organizing clusters into different groups. Clusters are group of objects that has similar features/attributes. Clustering algorithms have been used to conduct research on fruits such as tomato, banana, palm oil fruit, among others, [8],[9],[10]. Clustering can be considered as one of the most important techniques of unsupervised learning [1] as it aims at finding the structure in a collection of unlabeled data. Tomato as a fruit provides a number of health benefits that includes rich in antioxidants, cardiovascular support, support bone health, and anti-cancer treatments, among others [2]. However, tomato finds numerous uses in both fresh and processed form that includes ketch up, sauces, paste and juice, which results in earning income to a country that exports these products [3]. The colour of tomatoes remains one of the most important features that one can use to identify the ripeness of fruits. Thus, tomato colour helps one to know the ripeness of the fruit. If the tomato colour surface is mostly red then it is classified as mature tomato and if completely green then the tomato is classified as not matured. According to USDA (United States Department of Agriculture) tomato ripening stages, tomato is considered as red (matured) if 90% of the tomato surface is covered with red colour, green (not matured) if the whole surface of the tomato is completely green [4].

There are two types of methods in machine learning, supervised and unsupervised learning. In supervised learning there are defined rules and the outcomes are known, while in unsupervised learning the algorithm follows certain rules to learn by itself and come up with the result. Throughout the process of the research, four main steps are followed; acquire an image, clean image, segment and identify the image. Images are acquired in two ways: indoor and outdoor environment, in an indoor environment, certain factors have to be considered such as light, angle of taking the picture, image background, among others. While in outdoor environment the image is just captured in its natural environment. But mostly, when an image is acquired in outdoor environment, it undergoes more background cleaning which might affect the visibility of the captured image, thus reduces the chances of obtaining higher recognition rate of the fruit, especially if special camera were used [5]. Segmentation is the process of separating objects from its background. The important reason of conducting this research is that, it would help our local farmers and the industry to reduce high labor cost for identifying the ripeness of fruits (especially tomatoes).

2.0 Current Study on Fruit Ripeness Identification

A. Arefi and A. M. Motlagh [6] combined image processing and artificial neural network (ANN) to recognize ripe tomato from unripe tomato. Background images were removed by threshold segmentation and used RGB as colour space. Noises were removed based on morphological operation using opening by reconstruction operation. Touching tomatoes were separated by Euclidian distance. Ninety wavelet features are extracted based on two colour space RGB and HIS which reduced to sixteen effective features. Sensitive analysis was applied on the extracted features and used by the system designer to make decision on the extracted features and used as an input to the ANN. An MLP network was used by the ANN as input to classify RT from UT based on the wavelet features. They were able to combine image processing and ANN to recognized RT and UT with high accuracy.

The research conducted by X. Wang et al, [7] was about deciding the different maturity stages of tomato (breaker, turning, pink, light, and red stage)based on two colour model RGB and HIS, for a decision that decided the harvesting time of tomato by a robot. The images were acquired under natural conditions. And were segmented through OTSU method, background image was removed via logic processing. They were able to find out that, HUE and red green differences can be used to judge maturity stages of fresh tomatoes, but the Red green colour difference average method works better than HUE with a judgment accuracy of 96%.

H. Zhao and X. Zhou [8] suggested differentiating artificial ripening tomatoes from natural mature tomato using double parallel genetic network. Image were acquired in RGB colour space and converted to HSI colour space. Tomato colour used as an input and property as output to DPGNN. The real number coding system was used instead of binary code. The DPGNN used two algorithms; the genetic provided better recognition than BP algorithm which was reported as 93.8% accuracy.

X. Chen and S.X. Yang [9] discussed recognizing and localizing ripe and unripe tomato in a greenhouse by a robot. The Images were segmented using V colour classifier in the YUV colour space. Several experiments were conducted to determine the best colour model and achieve best result from V classifier of two class k - means algorithms. Simple edge detector algorithm was used to find the edges while contours tracing algorithm found the spatial boundaries of all objects. Finally, modified prewitt operator was used to locate and select the qualified edges and with this they were able to recognized and localized ripe tomato.

F. Quigchun, et al [10] designed a robotic vision system that can identify and locate mature tomatoes. Alaser projector was activated automatically on sighting the matured fruits. Images were converted to HIS from RGB colour model and segmented based on the result of the H and S value in threshold. They were able to identify matured tomatoes with only some errors due to the non-spherical shape of the fruit.

M. Chylinska et al. [11] aimed at depicting the spatial distribution of the main cell wall compounds in tomato tissue using Raman imaging. Raman imaging, principal component analysis and multivariate curve resolution methods were used. PCA was used for the imaging spatial distribution. PC1 couldn't distinguish between the different component while PC2 and PC3 provide the established pure component and PC3 also provides information about the cellulose. But the identification of the cellulose couldn't be obtained with the PCA. MCR technique was used to estimate the presence of pure components within the active map. Result had shown possibility to distinguish pectin and cellulose, while the distinguishing of hemi cellulose distribution was not possible due to overlapping characteristics band with cellulose and pectin's.

W. Lei et al. [12] introduced NMR and PCA analysis to characterize the metabolite profile of berry quality. They were able to used integrated NMR and multivariate analysis (PCA) to achieve the following: to compare the overall metabolite finger print of berry pulp and seed, to rapidly identify difference in bioactive compound among berry pulp and seeds and to detect possible biomarker for the separation of different tissues with different quality traits.

R. Kalaivani et al. [13] Identified good and bad tomato using automated sorting machine using various algorithms. Images were loaded into the mil software and saved as RGB format andwere converted from RGB to grey colour using linear intensity encoding of gamma expansion and finally to binary image. Noises were removed by median filter and images were segmented using canny edge detector algorithms. Results of the feature matching using correlated method were loaded into MATLAB for identification upon which we get a specific correlated value of good and bad tomato that were used for the identification. 80% accuracy was obtained for the identification of good and bad tomato.

The proposed paper by [14] discussed the use of simple algorithm to determine the ripeness of banana. The images of banana were captured until the banana become rotten. RGB colour format was used. Features were extracted by heuristic method and the feature vector by histogram method that was used to determine the ripeness of banana. An excel file was loaded into MATLAB that contains two folders of ripe and unripe banana that would be used as input to the neural network. They were able to achieve 96% recognition result.

The approach in [15] employed nearest neighbor for histogram-based palm fruit ripeness identification (ripe and unripe) in their paper. HSV colour space was used. They were able to find out that the value and colour lightness shows the most promising result of distinguishing between ripe and unripe palm oil fruit.

Y. Chiu et al considered an automated system to identify the harvest time of a crop through a cellular phone [16]. Features were extracted using Statistical moment. RGB colour space was used and the images were segmented through k-means algorithm. They were able to show that moment based k-means algorithm could efficiently be used to identify whether tomato was ripe or unripe, but the algorithm need to be implemented on the phone as a future work.

Segmentation on images of banana and tomato fruits at different stages of ripeness using partitional clustering algorithm namely k-means, FCM, GKB and GKPFCM were studied by [17]. RGB colour space was used. The image was segmented with each of the four algorithm with regard to the used of pixel intensity but spatial information was not taking into account. They discovered that as the various algorithms used different quantity and type of information and the more flexible the distance measure the better the result which is why the GKPFCM gave the best result that allowed for better matching among the regions identified and the object in the image.

The aim was to develop an autonomous picking robot that would operate in a greenhouse of tomato fruit by [18]. The robot has four parts: end effector, machine vision, robot carrier and control system. The machine vision had a lens that obtained actual 3D positions of each tomato features. The robot control system was operated by the Labview image processing software and the camera was also controlled by the image captured card of the Labview. HUE colour space was used. When the image was captured, the 3D coordinates of the tomato were calculated and then the robot arm will pick the tomato based on the instruction of the computer.

A Jaffar in their paper [19] used a photogrammetric grading system to inspect colour of palm oil fruit to determine the grading of its ripeness/maturity. The images were captured as an RGB and converted into L*a*b colour space. The images were preprocessed further using

binarization and morphological processing. K-means algorithm was used to segment the images which resulted in obtaining 3 clusters. Average digital no. of unsegment and segmented images were calculated and compared with the result of 3 clusters values to recognize the ripe and unripe palm oil fruit. K – means algorithm is integrated into the photogrammetric grading system. They were able to recognize ripe and unripe varieties of palm oil fruit.

The research produced in [20] designed an automated system that could detect and count gerbera flower images. The images were captured in an RGB format and were converted into HSV colour space. The images were preprocessed first, by removing the date portion that appears on the image of the flower using manual cropping and blurring. Features were extracted by histogram techniques and the images were segmented through threshold. The system achieved an accuracy of 100% for the red and yellow and 80.12% for pink. The overall accuracy of the system stands at 86.5%.

The research done by [21] aimed to design a computer mediated date fruit quality assessment and sorting system. The images were preprocessed by binarization threshold of image intensity histogram, and were converted to binary image through threshold. Edges were extracted through sobel edge operator. RGB colour space was used. Three manually criteria from grade 1 to 3 were defined and used as criteria for the two BPNN model for the classification. An 80% accuracy was achieved with the second BPNN model of the grade 2 classification.

The use of two steps k means algorithms to segment the images of banana finger and hand was discussed by [22]. RGB colour space was used. The first set of images was segmented through k means algorithm. Images with background were enhanced through colour inverse and noise removed by median filter while images without background were masked by enhancing the binary image to every channel of the RGB image. The 2^{nd} k means algorithm would be used to detect the flows in the image of banana without background. The first k means causes the flows to appear on the banana image but it disappear when the 2^{nd} k means was applied to it. The validation experiment had showed that it was possible to use double k means algorithms to segment the images of banana finger and hand.

The work by E. Guzman et al [23] was about improving a non-destructive method for fast online analysis of olive oil and suggested a direct method for determining the maturity of olive fruit online. The images were captured in duplicate mode, one mode in IR and the other in RGB format. To estimate the maturity index and external defects of the fruit, three steps were followed. Border four neighbor and edge based algorithm were used to extract the features. Gradient segmentation was used. A mask and application of the mask were created and added to the original image that simplifies the task of extracting features of the images. The original image is added to the mask and then colour and grey segmentation were performed. The results of the segmentation are implemented in matlab 7.4 and software image pro plus version 6.0, the proposed methodology can potentially be used for classification of the fruit online.

F S Mohamad et al in [24], proposed computer aided model to assess the quality of palm oil fruit based on ripe and unripe varieties. Images were saved as jpeg and converted to HSV colour format. Histogram was used to extract mean feature of the fruit. Distance feature were computed for each feature. Finally, images were identified by matching process based on the

calculation of distance measure (Euclidian, NN, furthest N and mean distance). Results had showed that HSV is the best colour model for palm ripeness identification. Value is the most dominant feature to indicate ripeness while the mean distance proved the best distance measure for both ripe and unripe categories.

O O Arjenaki et al [25] demonstrated an automated sorting machine based on image processing techniques for tomatoes. Tomatoes were group into desirable and undesirable based on maturity defects, shaped and size by an expert. HSV and RGB colour space was used. The images were transferred into grey scale and threshold with otsu algorithm. Shapes of the tomato were determined by calculating its eccentricity. To classify/determine the size, maturity and defects, their means colour was obtained and histogram was applied to the images. The overall system result stands at 98% while that of defects, size and shape stands at 84.4, 90.9 and 94% accordingly.

The paper by [26] developed an algorithm using machine vision by a grafting robot to grade and sort out tomato seedling into different categories. Features of the images were extracted through Polarized light filter. The images were processed by converting the grey scale image into binary image. The seed are graded based on the stem diameter and are sorted into the four groups with the last group being labelled as substandard or left for manual grafting. Results had showed an overall performance sorting rate of 97% with 3% accounting for manual grafting.

R S Sarkate et al used BPNN network to segment the images of gerbera flowers from offline polyhouse in [27]. 20 sets of images were captured under different lighting conditions in an outdoor environment. Cropping and filtering method were used to preprocess the images. HSV colour space was used and histogram analysis was applied to the images to get the corresponding colour values. A multi-layer feed forward network algorithm with back propagation was used to training the network.

The researchers in [28] focused on finding the most suitable colour model for palm fruit identification. RGB and HSV colour model were used. Images were captured by a certified grader and stored as jpeg format. The features were extracted with histogram feature vector and mean feature was obtained for both colour models. First experiment was conducted with the RGB model and second with HSV model. Results for RGB model had showed that it is less consistent since the values of RGB are not well separated and for the HSV model, the hue shows an overlapped in the images, while saturation show an improvement over HUE by showing less overlap but the value provided by the better result that shows the least possible overlap.

In [29], Shape feature extraction techniques was used to process the image, features were extracted using filter and wrapper techniques. The images were preprocessed and converted into HSV space, then to grey scale and finally to binary images. Noises were removed by morphological operation. Results had showed that an accuracy of 82.22% and 100% recognition was observed for tomato leaf and fruit habit using naïve based algorithm of wrapper based approach.

A Arefi et al developed a new algorithm based on colour quantification and fruit shape to recognized and localized RT and UT from the background [30]. HIS and YIQ colour space were used. The image background and noises were removed by Drg equation from RGB colour space.

The image was segmented with water shed algorithm in conjunction with define global disk. Three statistical criteria were used to recognized RT and UT. Their proposed algorithm was able to recognized tomato in 2.2 seconds and to localize the RT with 91.11% accuracy.

The proposed paper in [31] developed an algorithm and pseudo code to detect three diseases that affect palm oil trees through mobile device. A digital camera was used to capture the image of the leaf with a white background. RGB colour space was used. Images were filtered by BFS algorithm and six effective features were extracted. A mathematical model was developed so as to enable the technique to be implemented in mobile device using linear complexity to help reduced the processing time. A neural network with three hidden layers was designed. They were able to have 87.5% average accuracy.

U O Dorj et al in [32] designed an image analysis system to determine and estimate tangerine flower trees. The background noise of the image is first removed manually and gausian filter was applied to further reduce the noise. A colour detection algorithm with Regression analysis was used to count the number of white pixel and to use this value to determine the result of the work. They were able to find out that their percentage error is as low as 0% and as high as 55% and mean absolute error is 17% when compared with other previous research.

The research in [33], developed a machine vision system with selected algorithm to sort out tomatoes. As the tomatoes were conveyed by the belt, a light will be triggered so that the images were captured. Then the algorithm process the colour image by applying threshold and extracts the components of these colours R, G and HUE. The images were segmented through blob extraction level function. It checks the tomatoes for any defects and applied a decision sorting algorithm to the processed image to sort out the tomato. The decision sorting algorithm was designed based on ruler based method of two criteria decision logic and threshold value. Results had showed that the algorithm was able to detect the tomato with much higher accuracy as the threshold area is increased. Also, overall accuracy increases with increase in the shape factor. In conclusion, neural network detects cracks better while rule-based method detects calyxes better.

L Huang et al developed a machine vision system to recognized apple using image processing algorithm [34]. The images were converted from RGB format into L*a*b colour space. The images were segmented with fuzzy 2D entropy, and converted into binary image, and flood fill operations were conducted to remove numerous segmental fragments of the image. They were able to achieve the followings: l*a*b colour space provides the recognized system of apple with flexibility of real world scene of several apple farms, the fuzzy approach clearly separated the background and the object apart and finally a GA algorithm when compared with heuristic searching method enhanced the accuracy of the segmentation of the apple.

The aim was to device an image processing algorithm that can help measure the size and shape of a mango fruit, determine its defective area and grouped mangoes according to the extracted features in [35]. The RGB and Cb colour values of the image were calculated and segmented with iso data algorithm. Features were extracted through final segmented method. Nearest neighbor was used to determine the mango class and brute force and Euclidian distance were used to determine the mango quality. They were able to use nearest neighbor with

Euclidian distance to determine the quality of mango, and threshold value to identify the defective area of mango. Also, their algorithm can correctly segment the mango, but couldn't identify the stem due to its colour of green and brown. Green stem are classified as healthy while Brown as defective.

Jiang et al [40] designed a simple and effective identification method for tomato maturity by integrating color moments and physicochemical indices. The color moments were extracted by an adaptive K-means clustering image processing program, and firmness, soluble solid content and sensory evaluation were measured by professional techniques. To reduce the confusion between adjacent stages, a novel multinomial logistic regression with kernel clustering (MLRKC) method was designed to identify maturity, and the accuracy was 95.83% for tomato testing set.

Hu et al [41] proposed an automatic method for fast and accurate detection of ripe tomatoes on plant, which replaces manual labor with a robotic vision-based harvesting system, is a challenging task. The detection method combines deep learning with edge contour detection and separates target tomatoes from overlapping tomatoes to detect individual fruits. They used Gaussian density function of H and S in the HSV color space to help segment tomato regions from the background, followed by erosion and dilation on the tomato body to separate adjacent tomatoes and remove peripheral subpixels from all detected ripe tomatoes. The results of their study demonstrate that the proposed method improves tomato detection accuracy and that it can be further applied in the harvesting process of agricultural robots.

Toafik et al [42] developed a system to work on a mobile device with the ability to detect four levels of ripeness of tomato and chili fruits. In their proposal, they used K-Means Clustering Method for image segmentation while ripeness detection uses fuzzy logic. After an experiment, the results showed the level of success for fruit segmentation was 80% for tomato and 100% for chili and the level of success for detecting fruit ripeness is 80% for tomato and 90% for chili.

According to Viabhav et al [43] ripeness checking and grading is a well-known process that takes place in all the industries. Accordingly, they proposed a cost-effective system of detecting ripeness in order to suppress the sufferings of farmers. They took images using raspberry pi and the features are extracted which will help to find the ripeness and grading. Machine learning algorithm SVM is used for classification and to get the desired output. According to them, the system transmitted to different devices for large scale estimation.

Astrianda & Mohamad [44] argued that the maturity of tomatoes is usually determined by several parameters, including the parameters of size, weight, color features, texture, fragrance of the fruit and others. To this end, their research was based on identifying the maturity of tomatoes based on the main image-based colors. They deployed the neural network lavenberg-marquardt with different color models (HSV, YcBCr and CIElab) to find the best in the identification of tomato maturity. The results showed that tomato maturity was identified with an accurate rate of 96%.

Skolik et al [45] pointed that development and ripening of tomato fruit are important processes for the study of crop biology. To this end, they deployed a biospectroscopy approach using compact portable ATR-FTIR spectroscopy coupled with chemometrics and found that the

developmental and ripening processes showed unique spectral profiles, which were acquired from the cuticle-cell wall complex of tomato fruit epidermis in vivo.

Sari & Adinugroho [46] proposed a new method of segmenting tomato image for clustering its ripeness. The tomato images are taken from three types of smartphone camera in various lighting condition with white background. To evaluate the segmentation performance, clustering method is computed based on retrieved color of segmented image using K-Means, in which k=6 because of there are 6 stages of tomato ripeness. Color feature extraction by means of R, G, a*, and b* color channel are treated subsequently. The result showed that the system yield 1% Mean Square Error in clustering the ripeness of tomatoes.

Wu, et al. [47] presented a novel automatic algorithm for recognition of ripening tomatoes using an improved method that combines multiple features, feature analysis and selection, a weighted relevance vector machine (RVM) classifier, and a bi-layer classification strategy. The algorithm operates using a two-layer strategy. In the first-layer classification strategy, it is aimed at identifying tomato-containing regions in images using the colour difference information while the second classification strategy is based on a classifier that is trained on multi-medium features. Their proposal was presented to simplify calculations and improve recognition efficiency. The processed images are divided into 9 _ 9 pixel blocks, and these blocks, rather than single pixels, are considered as the basic units in the classification task. Six colour-related features, namely the Red (R), Green (G), Blue (B), Hue (H), Saturation (S) and Intensity (I) components, respectively, colour components, and five textural features (entropy, energy, correlation, inertial moment and local smoothing) were extracted from pixel blocks. The algorithm demonstrated the detection accuracy of 94.90% on 120 images which indicates that the proposed algorithm is effective and suitable for tomato detection.

3.0 DISCUSSION OF CURRENT STUDY

For the segmentation of the images different techniques were applied among the researchers, some used only one algorithm while others used many to compare and used the best algorithm that gave them better result. Among the researchers that use single segmentation algorithms are [9],[16],[17],[19]. [12] uses k means, though, according to author k means update its clustering constantly. But the method the researcher follows has a major setback of not distinguishing between mechanical injuries and the senescent spot [22]. Whereas [19][24] use L*a*b as it has uniformity in terms of distribution of colours and the way humans eyes views colours. [7],[25] uses otsu algorithm and obtained 96% and 98% accuracy respectively but according to [37] the method doesn't converge to a global optimum when working with images that are discontinuous in grey level, also [40] says that the performance will be degraded if the global distribution of the target image and its background varies widely. Genetic Algorithm methods was used by [3],[29] but it lacks the capacity to solve optimization problems[39]. Histogram method was used by [16],[24][27], but according to [38] says that the methods usually assign two different grey levels with different intensities to one grey level that resulted in reducing the efficiency of the method. Threshold was used by [20],[25], and the threshold method is one of the most prevalent technique for analyzing images in agriculture and food industry [22] but according to [22] threshold method incurred an over segmentation to the ripening stages of fruits. BP and BPNN algorithm were used by [8],[27] but the functionality of

the system can only progressed when the rules governing the working of the system is defined [39]. Canny edge detector was used by [13] but has a complex computation of time consumption [31]. [30] uses watershed algorithm but Splits fruits into several slices[30], and it also produces as over segmentation of the image [40] while fuzzy 2D algorithm was used by [34], but fuzzy is non deterministic in nature [38]. Iso data and nearest neighbor is used by [35] but there's tendency that nearest neighbor would produce a worsening predictive performance when presented with incorrect or insufficient database [39]. Other researchers like [8], [17], [22], [27], [32], [34], [35] employ more than one algorithm to segment the images in their research. Removing of noise in the image using spatial filter method was used by [22], but it's considered as a local procedure as the value of each pixel has to be change which makes the task to be repetitive and not fully computerize [22]. Sobel edge operator was employed by [21] to extract features, yet according to [40], it tends to produce spurious edges that are not present in the image. The method that [21] follow to grade the fruit based on human perception isn't effective because it didn't apply formal feature distribution as it can be used to determine the quality of fruit grading and improve classification accuracy [22]. [8,9] decreases their chances of obtaining accuracy and can cause failed harvesting which occur as the result of using special cameras to acquire the images. But [9] infer that only achieve good result during daytime could be attained since the proposed method requires the use of infrared lighting during nighttime [33]. Most of the researchers [14],[16],[17],[21],[22],[23], [31], [25] used RGB colour model, but [15] say RGB doesn't represent colour based on the way the human eyes views natural object. Also its inability to detect particular categories of fruit as it has correlation among the colour components [34]. Filter based method is used by [29] to carry out feature selection task but models can over fit one another as the weight put by the algorithm that does the ranking is not the same in term of the weight put by the classification algorithm [29]. [15][24][27][29] used HSV colour space as it is a device dependent model because It makes use of three concepts, [20][35]. [9] used YUV colour space but it tends to freak out when added and dark colour turns green when subtracted because its colour channels works out from the middle instead of going straight 0 – 255 like RGB in [35]. [47] presented a technique using relevance vector machine (RVM) classifier, and a bi-layer classification strategy which demonstrated the detection accuracy of 94.90%. The result obtained by [47] is similar with [40], [42] and [44] who applied different technique. The system advanced by [43] is said to be capable of transmitted to different devices for large scale estimation. It is also important to note [41] focused on colour detection using the Gaussian density function of H and S in the HSV color space to help segment tomato regions from the background while on the other hand [42] and [46] all work on mobile devices and explore its cameras to collect and detect the ripeness.

4.0 Proposed Methodology

First image would be acquired from a farm through digital camera, and then the image will be preprocessed by removing the background objects and noises. We will used various colour model such as RGB, HSV, HIS, among others on the image. We will then use five algorithms to segment the image namely: K – Means, Fuzzy, nearest neighbor, otsu and threshold method to identify the ripeness of tomato (ripe, unripe, overripe, and under ripe). Next, feature extraction would be performed, and then the result will be loaded into MATLAB for further processing and identification of the fruit images. Both the result of the corresponding colour model and

algorithm would be compared and use the result that would provide us with best fruit recognition rate when implemented in Matlab.

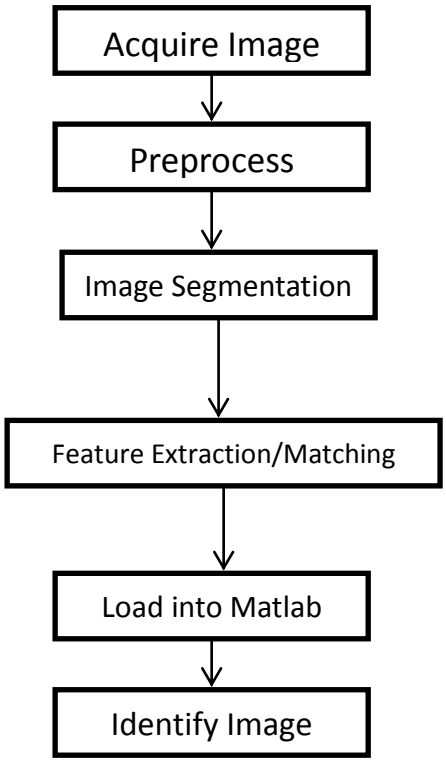


Fig. 1: Process Flow

5.0 Conclusions

This paper reviewed several studies carried out by other researchers on fruit ripeness identification, although most attention is focused on the tomato fruit. The review work presented here provides us with the idea on how we will conduct our fruit ripeness identification research. Most of the researchers used k means algorithm, mobile phone technology such as camera in their work while a robotic application is getting widespread use especially in the developing countries to replace human laborers working at the farm and greenhouses. Most of the researchers were able to achieve more than 80% to 90% performance rate. Based on the literatures reviewed, we concluded that in our proposed methodology we will use five algorithms used by most researchers that achieve more recognition rate and compare the result to either integrate or come up with new algorithms/colour models to achieve higher performance than previously done works.

Acknowledgement

We thank our creator for giving us the strength and ability to conduct this research. Special thanks go to Kano State Government for sponsoring this program.

References

[1] Alessandro Bevilacqua, (2012). "Pattern Recognition: Classification and Clustering", ACAT, Advanced School, Bologna. Pp 22.
[2] http://www.whfoods.com/genpage.php?tname=foodspice&dbid=44, Tomato, (visited on June 2, 2014)
[3] T. U. Ganiron (2014). "*Size Properties of Mango using Image Analysis",* International Journal of of Bio Sicence and Bio-Technology. Vol 6, No.2 pp. 31-34.
[4] C. Marita, (2013). *Ripening Tomatoes*. Fruit Ripening and Retail Handling Workshop Postharvest Technology Center, UC Davis, March 18-19, 2013. Slide 2.
[5] F. Qingchun, et al (2014). "Design of structured-light vision system for tomato harvesting robot", Int J Agric & Biol Eng. Vol 7 No.2, P19.
[6] A. Arefi and A. M. Motlagh (2013) *Development of an expert system based on wavelet transform and artificial neural networks*. Australian Journal of Crop Science AJCS 7(5):699-705.
[7] X. Wang, et. al, (2011) "*Vision-based judgment of tomato maturity undergrowth conditions".* African Journal of Biotechnology Vol. 10(18), pp. 3616-3623.
[8] H. Zhao and X. Zhou (2013) "*Recognition of Artificial Ripening Tomato and Nature Mature Tomato Based on the Double Parallel Genetic Neural Network".* Advance Journal of Food Science and Technology 5(4): 482-487.
[9] X. Chen, and S.X. Yang (2013) *"A practical solution for ripe tomato recognition and localization" J Real-Time Image Proc (2013) 8:35–51*
[10] F. Qingchun, et al (2014). "Design of structured-light vision system for tomato harvesting robot", Int J Agric & Biol Eng. Vol 7 No.2, 19.
[11] M. Chylinska et al. (2014) *"imaging of polysaccharides in the tomato cell wall with Raman microspectroscopy"* http://www.plantmethods.com/content/10/1/14
[12] W. Lei et al. (2013). "*NMR metabolomics of berry quality in sea buckthorn"* Mol Breeding, 31:57–67.
[13] R. Kalaivani et al. (2013) *"Identifying the Quality of Tomatoes in Image Processing Using Matlab"* IJAREEIE, Vol. 2, Issue 8
[14] M P Paulraj et al. (2009) "Color Recognition Algorithm using a Neural Network Model in Determining the Ripeness of a Banana", ICoMMS, 11 – 13 October 2009, Penang, Malaysia.
[15] Mohamad, F.S., et al, (2011). "Histogram-based Fruit Ripeness Identification Using Nearest-Neighbor Distance" ICCS, Vol 4, Pages 1296–1305.
[16] Bhatnagar, M., Kumar Singh, P., (2014). "An Efficient Method of Image Segmentation for Harvest Time Identification", IJCA, Vol 87, No.7.
[17] B. O. Magana et al. (2010) "Color Image Segmentation by Partitional Clustering Algorithms". IEEE.
[18] Y. C. Chiu et al. (2013). "*Study of an Autonomous Fruit Picking Robot System in Greenhouses",* EAEF 6(3) : 92-98, 2013.
[19] A. Jaffar et al. (2009) "*Photogrammetric Grading of Oil Palm Fruit Fresh Fruit Bunches",* International Journal of Mechanical and Mechatronics Engineering, Vol 09 No:10.
[20] R. S. Sarkate et al. (2013) "*Application of Computer Vision and Colour Image Segmentation for Yield Prediction Precision",* IEEE.

[21] Y. A. Ohali (2010). *"Computer Vision Based Date Fruit Grading System: Design and implementation*", Journal of King Saud University Computer and Information Sciences, (2011) 23, 29 – 36.
[22] M. H. Hu et al. (2013). *"The Potential of Double K-Means Clustering for Banana Image Segmentation",* Journal of Food Process Engineering ISSN 1745-4530.
[23] E. Guzman et al. (2013). *"Using a visible System for On-Line Determination of Quality Parameters of Olive Fruits",* Food and Nutrition Sciences, 2013, 4, 90-98.
[24] F. S. Mohamad et al. (2012). "*Exploiting Distance Measurement for Ripeness Identification*", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012.
[25] O. O. Arjenaki et al. (2013). "Online Tomato Sorting Based on Shape, Maturity, Size, and Surface Defects Using Machine Vision", Turkish Journal of Agriculture and Forestry, (2013) 37: 62-68".
[26] M. A. Ashraf et al. (2011). "*Use of Machine Seedling to Sort Tomato Seedlings for Grafting Robot",* Eaef 4(4): 119-125,2011.
[27] R. S. Sarkate et al. (2014). *"BPNN Approach in Pixel Classification based Precision Segmentation for Agriculture Images",* International Journal of Computer Applications (0975 – 8887), Recent Advances in Information Technology, 2014.
[28] F. S. Mohamad et al "*Exploiting Suitable Colour Model for Ripeness Identification*", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012.
[29] A. Hazra et al. (2013). "*Shape Oriented Feature Selection for Tomato Plant Identification*", International Journal of Computer Applications Technology and Research, Volume 2–Issue 4, 449 - 454, 2013.
[30] A. Arefi et al. (2011). "Recognition and localization of ripen tomato based on machine vision", AJCS 5(10):1144-1149 (2011).
[31] A. F. Aji et al. (2013). "Detection of Palm Oil Leaf Disease with Image Processing and Neural Network Classification on Mobile Device", International Journal of Computer Theory and Engineering, Vol. 5, No. 3, June 2013.
[32] U. O. Dorj et al. (2013). "*A computer Vision Algorithm for TangerineYield Estimation*", International Journal of Bio-Science and Bio-Technology, Vol.5, No.5 (2013), pp. 101-110.
[33] Md. Rokunuzzaman, H. P.W. Jayasuriya (2013). "Development of a low cost machine vision system for sorting of tomatoes", Agric Eng Int: CIGR Journal, Vol. 15, No.1 173.
[34] L. Huang et al. "Segmentation on Ripe Fuji Apple with Fuzzy 2D Entropy based on 2D histogram and GA Optimization" Segmentation on Ripe Fuji Apple with Fuzzy, Intelligent Automation & Soft Computing, 19:3, 239-251.
[35] k.leon, d.mery, f pedreschi, (2006). Colour management in l*a*b units from RGB digital images. J*ournal of food technology.*
[36] *Two Algorithms of Image Segmentation and Measurement Method of Particle's Parameters.* Applied Mathematics & Information Sciences International Journal, Natural Sciences Publishing Cor. Vol 6 No. 1S pp. 105S.
[37] X. Yuping, et. all, (2010). *Nearest neighbor and logistic regression analyses of clinical and heart rate characteristics in the early diagnosis of neonatal sepsis.* Med Decis Making. 2010 Mar–Apr; 30(2): 258–266.
[38] C. Marita, (2013). *Ripening Tomatoes*. Fruit Ripening and Retail Handling Workshop Postharvest Technology Center, UC Davis, March 18-19, 2013. Slide 2.

[39] B.Lakshmi, (2012) *An Enhanced Approach for Medical EDGE Image Enhancement using Genetic Algorithm.* IJCST Vol. 3, ISSue 1, Jan. - March 2012. pp273.
[40] Jiang Y, Bian B, Wang X, Chen S, Li Y, Sun Y. Identification of tomato maturity based on multinomial logistic regression with kernel clustering by integrating color moments and physicochemical indices. J Food Process Eng. 2020; e13504. https://doi.org/10.1111/jfpe.13504
[41] Hu, Chunhua, et al. "Automatic detection of single ripe tomato on plant combining faster R-CNN and intuitionistic fuzzy set." IEEE Access 7 (2019): 154683-154696.
[42] Taofik, A., et al. "Design of Smart System to Detect Ripeness of Tomato and Chili with New Approach in Data Acquisition." IOP Conference Series: Materials Science and Engineering. Vol. 288. No. 1. IOP Publishing, 2018.
[43] Viabhav, et al (2019) "identification of ripeness of tomatoes." International Journal of Engineering, Research & Technology (IJERT), Special Issue – 2019, Volume 7, Issue 08, 1-4.
[44] Astrianda, Nica, and Fatma Susilawati Mohamad. "Ripeness Identification of Tomato Using Different Color Models Based on Neural Networklevenberg-Marquardt." World Applied Sciences Journal 35 (2017): 57-61.
[45] Skolik, Paul, et al. "Determination of developmental and ripening stages of whole tomato fruit using portable infrared spectroscopy and Chemometrics." BMC plant biology 19.1 (2019): 1-15.
[46] Sari, Yuita Arum, and SigitAdinugroho. "Tomato ripeness clustering using 6-means algorithm based on v-channel otsu segmentation." 2017 5th International Symposium on Computational and Business Intelligence (ISCBI). IEEE, 2017.
[47] Wu, Jingui, et al. "Automatic recognition of ripening tomatoes by combining multi-feature fusion with a bi-layer classification strategy for harvesting robots." Sensors 19.3 (2019): 612.