Computer Vision Based Maturity Level and Disease Detection for Tomato

P Sai Krishna[1], M Karthik[2], S. Usha[3]
[1, 2, 3] Department of Electrical and Electronics Engineering
[1] PG Scholar, Kongu Engineering College, Erode, Tamilnadu, India
[2, 3] Assistant Professor (Senior Grade), Kongu Engineering College, Erode, Tamilnadu, India

Abstract *- In India, tomato consumers are nearly about 10-20 million on every day. Its farming is highly labor attentive and offers more employment all over India. After the cultivation, Tomato are mainly classified by three types of maturity level such as unripe, ripe and over ripped. The main aim of this proposed project is to analyses the maturity level and diseases using digital image processing, discriminative clustering and hybrid classifier techniques. Using the high resolution camera, different healthy and diseased images of tomato were collected. The collected Tomato digital image can be analyzed using MATLAB in version 9.1 developed on 2016 which gives a common pattern of image. Here the system stores both levels of maturity and diseased tomato at different time period. Using Discriminative Clustering based Segmentation the feature is extracted from both diseased and healthy tomato. By using hybrid classifier, it analyses whether the tomato is affected by disease or not and also to classify the maturity level.*

Keywords - Tomato; Discriminative Clustering based segmentation; PAC and GLCM; Minimum Distance Classifier;

I. INTRODUCTION

Tomato is used for daily life cooking and medicine. These tomato are classified by maturity level like ripe, unripe and over ripped and diseases like Black spot and Anthracnose. The captured digital images are used to find out the maturity level and diseases by using different methodology.

In image preprocessing the noise is removed by using Gaussian filter in acquired tomato images. CIELAB space model is used for color transformation. CIELAB color component are converted from RGB color component. ‘L’ indicates the color intensity of CIELAB color. The discriminative clustering based segmentation is helps to crop the fore ground (tomato) from its background image. Discriminative clustering crops the image by pixel clustering.

The PAC and GLCM feature is extracted for tomato image. The hybrid classifier (MDC) is helps to classify the disease affected tomato and to identify the maturity level of tomato.

II. METHODOLOGY

The captured image is pre-processed and the fore ground Tomato is segmented in segmentation process. The PAC and GLCM feature is extracted and moved to hybrid classifier.

The algorithm used here is to analyze the major three types of maturity and diseases in Tomato. The work flow of proposed algorithm is shown below in Fig. 1.

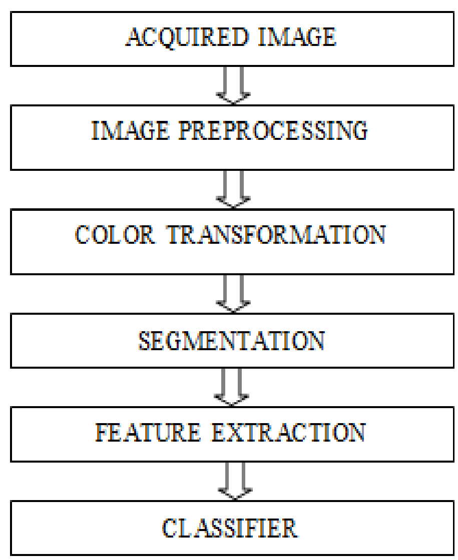

Fig.1 Flowchart of the work flow

III. ACQUIRED IMAGE

Image is captured from the tomato market. The resolution of captured image is 5312*2988. For processing the healthy and diseased tomato images are acquired. More over 150 images of different maturity level and diseased affected tomatoes are taken which is shown in Fig 2.

a.over ripe

b. ripe

c .un ripe

d. diseased

e.diseased

f. Anthracnose

Fig 2 Samples of tomato

IV. PREPROCESSING

To get efficient image the acquired image are pre-processed by enhancement and removal of noise by using different filters. The filter used is gaussian filter.

A. Gaussian filter

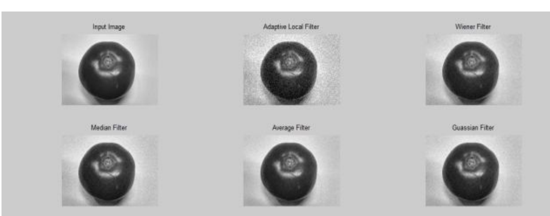

Fig.3.Different filters applied Tomato

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$$

Gaussian filter is achieved on the reduced in quality of image that contains both noise as well as original image. The average PSNR value while using gaussian filter is 30.61. The noise is effectively removed by Gaussian filter. Here the quality of image is increased and noises reduced. The Gaussian filter is more effective for the edge detection. The fig. 3 shows output of different filters.

V. IMAGE COLOR CONVERTION

The CIELAB color model is indicated by the Global Commission on Light. This is system freelance and describes the color that is seen by necked eye. The most segmentation was drained color area model. Every color allocate the values wherever physical property L* changes from zero to one hundred (zero as black & one hundred as white) and a*, b* represent that -120 to +120 are the range that gamut varies. The layer a* and b* are ranges from green to red and blue to yellow.

The basic (a*&b*) is predicated on the adverse color, that represent RGB cannot be at a similar time. The transformation from RGB area to CIELAB area is shown in Fig.4.

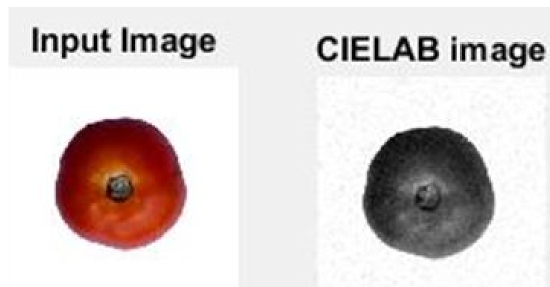

Fig. 4 CIELAB colour model of Tomato

VI. SEGMENTATION

The discriminative clustering methodology is based on positive definite kernels. Since our k-dimensional features are all histograms, we consider a joint n × n positive semi definite kernel matrix K (defined for all pairs of all pixels from all images) based on the χ2-distance, with entries:

$$K_{lm} = \exp\left(- \lambda_h \sum_{d=1}^{k} \frac{\left(x_d^l - x_d^m\right)^2}{x_d^l + x_d^m} \right)$$

Where h > 0. In the experiments, we use λ h = 0.1.Notethat we do not use the positions p j to share information through images in order to be robust to object location.

The restored image is superimposed on initial tomato image. The expanded image is retrieve by cropping highest marker level region. Finally the discriminative bunch is completed to get the segmental Tomato image. Fig.5 shows output of discriminative clustering.

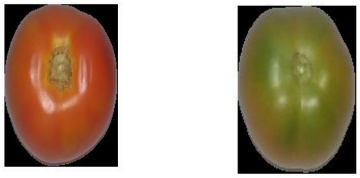

Fig. 5 Discriminative clustering

VII. FEATURE EXTRACTION

The extraction of PCA feature for Tomato are obtained by following steps

Principal component analysis

Step 1: Get input image

Let A (N, n) be the image matrix: N is that the range of pixels, n is that the range of dimensions. Calculate the mean.

Step 2: Subtract the mean

For PCA to work perfectly, you have to subtract the mean from each of the data dimensions. The mean subtracted is the average beyond each dimension. So, all the values have subtracted from them. This produces a data set whose mean is zero.

$\bar{x} = \frac{1}{z}\sum_{i=1}^{z} x_i$

Data = data – repmat (mn, 1, N)

Step 3: Calculate the covariance matrix

Covariance is such a measure. Covariance is always measured between 2-D. If you calculate the covariance between 1-D and itself, you get the variance. So, if you had a 3-dimensional data set (x, y, z), then you could measure the covariance between the x and y dimensions, then x and z dimensions, and then y and z dimensions.

Covariance = $\frac{1}{z-1}\sum_{i-1}^{z}(x_i-\bar{x})^T(x_i-\bar{x})$

Step 4: Calculate the eigenvalues and eigenvectors

Hence the co-variance matrix is sq., we will confirm the eigenvectors and eigenvalues for this matrix. These are relatively important, as they tell us useful information about our data.

Step 5: Choose the component and form the feature vector

Propose the n dimensional information on a p dimensional sub-space (p ≤ n), minimizing the error of the projections, here is wherever we tend to scale back the spatiality of the info. Order the eigenvalues from highest to lowest to urge the parts so as of understanding. Project on the p eigenvectors that corresponds to the best p eigenvalues. The eigenvector with the best eigenvalue is that the principal element of the information.

GLCM Feature

Gray co-matrix creates an GLCM by calculating how many times a pixel with grey level value (i) occurs horizontally neighbour to a pixel with value (j) Each value in (i,j) are replaced in gray co-matrix. Which is show in Fig.6 Gray co-props measures the value of contrast correlation, homogeneity and energy.

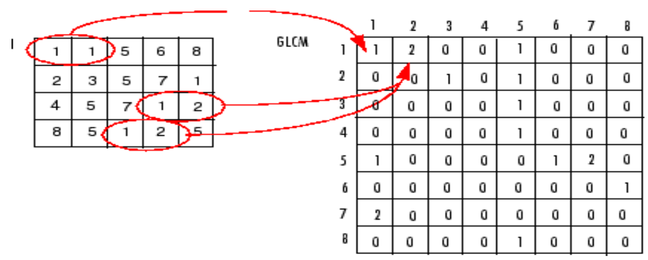

Fig. 6 Calculation of GLCM

Contrast:

The measure of capacity to distinguish between luminance of different levels in a static image is known as Contrast Sensitivity. The split of light and dark pixels are called Contrast.

Contrast = $\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P(i,j)(i-j)^2$

Where i, j is a grey level pairs, p is a GLCM element.

Correlation: Measure of linear dependency of gray levels

Correlation = $\frac{1}{\sigma_x\sigma_y}\sum_{i=1}^{G}\sum_{j=1}^{G}(i,j)P(i,j) - \mu_x\mu_y$

Where, μ_x, μ_y, σx and σ_y are the mean & SD (standard deviation) of the p_x and p_y.

Cluster Prominence: Measure of skewness of matrix

Cluster prominence = $\sum_{i=1}^{G}\sum_{j=1}^{G}(i+j-\mu_x-\mu_y)^4 P(i,j)$

Cluster Shade: GLCM is symmetric when cluster shade is low

Cluster shade = $\sum_{i=1}^{G}\sum_{j=1}^{G}(i+j-\mu_x-\mu_y)^3 P(i,j)$

Dissimilarity: Variation in intensity values

Dissimilarity = $\sum_{i=1}^{G}\sum_{j=1}^{G}|i-j|P(i,j)$

Energy: It provides the sum of squared elements in the GLCM.

Energy = $\sum_{i=1}^{G}\sum_{j=1}^{G}P(i,j)^2$

Entropy: Statistical measure of randomness

Entropy $= -\sum_{i=1}^{G}\sum_{j=1}^{G}P(i,j)\log(P(i,j))$

Maximum probability: Possible occurrence of grey level to the total grey level (N).

Maximum probability = $\frac{1}{N_g}\sum_{i=1}^{G}\sum_{j=1}^{G}P(i,j)$

Sum of squares: Statistical is a measure of heterogeneity

Sum of square=$\sum_{i=1}^{G}\sum_{j=1}^{G}(i-\mu)^2P(i,j)$

Gradient levels of tomatoes for various parameters are presented in Table.1.

Table 1 Gradient value for tomatoes

PARAMETER	MATURITY LEVEL AND DISEASE			
	Overripe	Ripped	Unripe	Disease
PCA Variance	0.0380	0.0157	0.0234	0.0012
Compactness	0.9035	0.8510	0.8990	0.7842
Color correction	0.7244	0.9189	0.9338	0.9561
Length width ratio	0.9668	0.5821	1.1648	1.6552
Circle area ratio	1.9644	1.1904	2.3304	3.2138
Square area ratio	10.004	6.0658	8.7537	5.9775
Triangle area ratio	1.2942	1.3026	1.2746	1.2367

VIII. CLASSIFIER

MDC (Minimum Distance Classifier) is used to identify the maturity level such as over ripe, ripe and unripe. The achieved classes from feature extraction are saved in the hybrid classifier. Diseased and non-diseased tomato is trained. For the classifier the gradient values of acquired image (untrained tomato) are given as the input. In the case of input values are close to the maturity class, then the classifier classifies that untrained tomato belongs to maturity. The results obtained are presented in Fig.7.

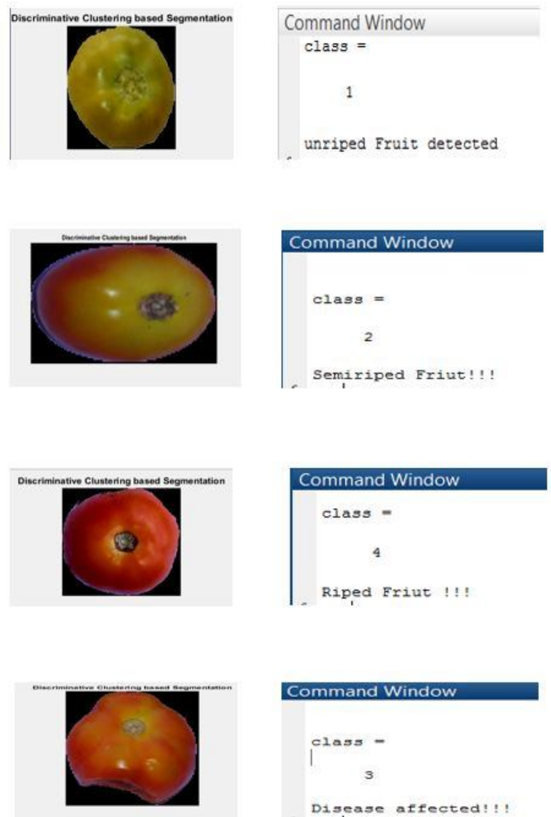

Fig.7 Classified tomatoes

IX.CONCLUSION

The proposed methodology is to classify the maturity of tomatoes and to identify whether they are affected with disease or not. In MATLAB the image processing is a best tool for detecting maturity as well as diseases affected tomatoes in short time with accurate and absolute result. In discriminative clustering based segmentation the foreground (tomato) is exactly segmented and neglected the background images. By using PCA and GLCM methods the different gradient value of tomato images are taken then, the Minimum distance classifier, exactly classifies the maturity level and identification of disease affected tomato with the help of extracted feature values of the given images in a short period of time.

REFERENCES

[1] Arazuri, S., Jaren, C., Arana, J.I., Perez De Ciriza, J.J., (2007). ‘Influence of mechanical harvest on the physical properties of processing tomato’. Journal of Food Engineering Vol.80 (1),pp 190–198.

[2] Arora, S., Kumar, S., (2005). ‘Studies on engineering properties of different varieties of tomatoes’. IE(I) Journal AG Vol.86 (12),pp.58–60.

[3] Anup Vibhute, Bodhe S K, (2012), ‘Applications of Image Processing in Agriculture: A Survey’, International Journal of Computer Applications (0975 – 8887), Vol.52(2).pp.34-40

[4] Antonio Cezar de Castro Lima, Camila Baccili, (2014), ‘Segmentation of somatic cells in goat milk using color Space CIELAB’, Journal of Agricultural Science and Technology.

[5] Blasco J , Aleixos N, Moltó E,(2007). ‘Computer vision detection of peel defects in Citrus by means of a region oriented segmentation algorithm’. Journal of Food Engineering, pp. 535-543.

[6] B.C.Russell, W.T. Freeman, A.A. Efros, J.Sivic, and A. Zisserman,(2006) ‘Using multiple segmentation to discover objects and their extent in image collections’. CVPR.

[7] Chen, Xuming and Yang, Simon X. (2013), ‘A Practical Solution for Ripe Tomato Recognition and Localisation’, Journal of Real-Time Image Processing, Vol. 8(1), pp. 35-51.

[8] Dah-Jye Lee, James K. Archibald, and Guangming Xiong,(2011)’ Rapid Color Grading for Fruit Quality Evaluation Using Direct Color Mapping’ IEEE Transactions on Automation Science and Engineering, Vol.8.pp.292-302.

[9] Devrim Unaya, Bernard Gosselinb,Olivier Kleynenc, Vincent Leemansc, Marie-France Destainc, Olivier Debeird,(2011) ‘Automatic grading of Bi-colored apples by multispectral machine vision’ ELSEVIER, Computers and Electronics in Agriculture, Vol.75.pp.204-212.

[10] Fakhri A, Nasir A, Nordin M, (2012), ‘A Study of Image Processing in Agriculture Application under High Performance Computing Environment’, International Journal of Computer Science and Telecommunications, Volume 3(8).pp.16-24.

[11] Liu C Q, Salman C, (2003), ‘Image Segmentation and Edge Detection Based on Watershed Techniques’, International Journal of Computers and Applications, Volume 25, no. 4, pp. 258-263.