Detection of Maturity Level and Disease in Tomato Using Discriminative Clustering Method

A. Stephen Sagayara(1), C. Kalaarasan(2), S. Gokulnath(3), R. Karthikeyan(4), K. Bharath(5)

Assistant Professor, Department of Electronics and Communication Engineering, Jay Shriram Group of Institutions, Tirupur, Tamilnadu, India

UG Scholar, Department of Electronics and Communication Engineering, Jay Shriram Group of Institutions, Tirupur, Tamilnadu, India

ABSTRACT: In India, tomato buyers are almost around 10-20 million on consistently. The farming is highly labor focused and offers more employment all over India. After the cultivation, tomato are mainly classified by three type’s maturity level such as unripe, ripe and over ripe. It places a major role in food product. The proposed idea is to analyse the maturity level and diseases of tomato using digital image processing, discriminative clustering. The different diseased and healthy tomato images are collected from various agriculture farms by using high resolution camera. The image is preprocessed by using Gaussian filter, median filter and linear filter to eliminate the noises. The image is segmented using discriminative clustering method and GLCM feature is used to extract the feature for detection of maturity level and diseases in tomato.

KEYWORDS: Tomato images; Discriminative Clustering based segmentation; PAC and GLCM;

I. INTRODUCTION

Tomato is a fruit which is highly utilized one in day by day life as for cooking and as medicine. These tomato are differentiated by their ripe, unripe, over ripe maturity conditions and the diseases like bacterial spot and anthracnose. A different methodology were utilized to find the maturity level and the diseases occurred in the tomato. Were the captured digital images of the tomato are used to find the maturity level and diseases of tomato. In image enhancement method the image is preprocessed by using filters like Gaussian filter, median filter and linear filter to remove the noises in the acquired tomato images. The LAB colour space describes mathematically all perceivable colours in the three dimensions L for lightness and a and b for the colour opponents green-red and blue-yellow. CIELAB colour space is used for colour transformation. RGB colour component are converted to CIELAB colour component. The colour intensity of CIELAB is indicated by the letter ‘L’. Discriminative clustering based segmentation is used to crop the foreground image from its background image. Here the cropping the image is done by pixel clustering. GLCM and PAC feature is extracted for tomato.

II. METHODOLOGY

The image that has been captured has to be preprocessed and the image colour is converted into RGB to CIELAB colour space. The foreground tomato image is cropped in discriminative clustering segmentation process. The GLCM and PAC feature is extracted for the collected image.

The algorithm followed here to classify the ripe, unripe, over ripe tomato’s and the disease affected tomato’s. The flowchart of proposed algorithm is shown below in
Fig. 1.

---

III. ACQUIRED IMAGE

Image is captured from the various agriculture farms by using high resolution camera. The resolution of captured image is 256*256. For processing the non-diseased and diseased tomato images are acquired. More over 100 images of different maturity level and diseased affected tomatoes are taken which is shown in Fig 2

(1)  
(2)
IV. PREPROCESSING

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

A. Gaussian filter

Gaussian filtering is used to blur images and remove noise and detail. In one dimension, the Gaussian function is:

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

![Image with Gaussian noise](image1)

![Image without noise](image2)

Fig. 3

Gaussian filter is resulted on reducing noise in the captured image that contains both noise as well as original
tomato image. The PSNR value of the image with noise is 48.7098. The removal of noise is done by Gaussian filter. The PSNR value of the image without noise is 48.7312. Here the quality of image is increased and noises reduced. The Gaussian filter is more effective for the edge detection.

V. IMAGE COLOR CONVERSION

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

VI. SEGMENTATION

The goal of image segmentation is to cluster pixels into salient image regions, i.e., regions corresponding to individual surfaces, objects, or natural parts of objects. The discriminative clustering method is used to segment the fore ground tomato image from the back ground.

Where \( h > 0 \). In the experiments, we use \( \lambda, h = 0.1 \). 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 overlaid on primary tomato image. The stretched image is recover by cropping highest marker level region. Finally the discriminative clustering is completed to get the segmental Tomato image. the segmented images shown in Fig.4.

1,2 segmented image Fig.4
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.

\[
\text{Mean} = \frac{1}{N} \sum_{i=1}^{N} A_i
\]

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.

\[
\text{Data} = \text{data} - \text{repmat} (\text{mean}, 1, N)
\]

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. 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**

A statistical method of examining texture that considers the spatial relationship of pixels is the GLCM. Feature

\[
\text{Covariance} = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{j=1}^{N} (i - \text{mean}) (j - \text{mean})
\]

Extraction is a method of capturing visual

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 content of images for indexing & retrieval. Primitive or low level image features can be either general features, such as extraction of color, texture and shape or domain specific features. This paper presents an application
of gray level co-occurrence matrix (GLCM) to extract second order statistical texture features for motion estimation of images.

The Four features namely, Angular Second Moment, Correlation, Inverse Difference Moment, and Entropy are computed using Xilinx FPGA. The results show that these texture features have high discrimination accuracy, requires less computation time and hence efficiently used for real time Pattern recognition applications.

Where, \( \mu_X \), \( \mu_Y \), \( \sigma_X \) and \( \sigma_Y \) are the mean & SD (standard deviation) of the \( p_X \) and \( p_Y \).

**Cluster Prominence**:  Measure of skewness of matrix

\[
\text{Cluster prominence} = \sum \gamma
\]

\[
\gamma = 1 \left( \overline{\gamma} + \frac{\gamma - \overline{\gamma}}{2\overline{\gamma}} - \overline{\gamma} \right) \overline{\gamma} \overline{\gamma}
\]

**Cluster Shade**: GLCM is symmetric when cluster shade is low

\[
\sum \gamma
\]

\[
\left( \overline{\gamma} + \frac{\gamma - \overline{\gamma}}{2\overline{\gamma}} - \overline{\gamma} \right) \overline{\gamma} \overline{\gamma}
\]

**Contrast**:
\[ \| \| - 1 \| \| (0, 0) \]

The split of light and dark pixels are called Contrast.

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

\[
\text{Contrast} = \sum_{i=1}^{2^m-1} \sum_{j=1}^{2^n-1} p(i,j)(i-j)^2
\]

\[
\text{Energy} = \sum_{i=1}^{2^m-1} \sum_{j=1}^{2^n-1} p(i,j)^2
\]

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

**Entropy:** Statistical measure of randomness

\[
\delta = \delta
\]

**Correlation:** Measure of linear dependency of gray levels

\[
\text{Entropy} = \sum_{i=1}^{2^m-1} \sum_{j=1}^{2^n-1} p(i,j) \log(p(i,j))
\]

\[
\text{Correlation} = \frac{1}{\sum_{i=1}^{2^m-1} \sum_{j=1}^{2^n-1} p(i,j)(i-j) - \mu \cdot \nu}
\]

**Maximum probability:** Possible
\[ \sum_{o}^{c} = \]  

occurrence of grey level to the total grey level (N).

\[ \sum_{(i,j)}^{c} \]

\[ \sum_{0}^{c} \]

\[ \sum_{1}^{c} \]

\[ \sum_{-1}^{c} \]

**Sum of squares:** Statistical is a measure of heterogeneity

Table 1 Gradient value for tomatoes to identify maturity level and diseased tomatoes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Maturity level and diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ripped</td>
</tr>
<tr>
<td>PCA variance</td>
<td>1.7475</td>
</tr>
<tr>
<td>Compactness</td>
<td>0.7082</td>
</tr>
<tr>
<td>Colour correction</td>
<td>0.8616</td>
</tr>
<tr>
<td>Length width ratio</td>
<td>1.4917</td>
</tr>
<tr>
<td>Circle area ratio</td>
<td>3.9234</td>
</tr>
<tr>
<td>Square area ratio</td>
<td>6.1223</td>
</tr>
<tr>
<td>Triangle area ratio</td>
<td>1.3740</td>
</tr>
</tbody>
</table>

**VIII. CONCLUSION**

The proposed methodology is used to analyze the maturity and diseases of tomato. the DCM (discriminative clustering method) is used to segment the foreground image from background and feature is
extracted by using GLCM and PCA method. In future KNN classifier and ANFIS classifier is will use for classification accuracy.

REFERENCES