



Disease Detection in Plants using a Pseudo Color Co-Occurrence Matrix

Jibrael Jos, K A Venkatesh

Abstract: This study reports a color based texture classification for a machine vision system for the identification of disease in plants from color leaf images. We applied the texture features in literature and studied which subset will be effective for Mango and Tomato plants. Effectiveness of each statistical functions were studied in classifying the pattern using a Support Vector Machine. For textures which are different like smooth new leaves, dry leaves and growth Gray Level Co-occurrence based statistics was effective but values failed to discriminate in tomato diseases. We propose a novel method which uses second order statistics on a pseudo color based co-occurrence matrix which resulted in a better classification for three tomato diseases. This method can be applied for early Disease Detection for any plant and help farmers take corrective measures to avoid loss of yield.

Keywords : CCM, Disease Detection, Pattern Classification, Texture Analysis, Tomato Plants.

I. INTRODUCTION

A mango tree has to face different kinds of environmental challenges. Rain is good but if it happens in the flowering or during fruiting season it can be detrimental. Extreme heat and cold can also affect the quality of the fruit. A wide variety of insects can cause havoc at different stages. Bacterial and Fungal diseases can also result in loss of production. Farmers in India work in cooperatives and self-help groups to ensure the experiences in managing the infections of the trees in their orchards are shared across other farmers in the village. How can a cost effective solution be provided which can help them proactively handle infection and pest outbreak in the orchards. The system created has a Color, Texture and Shape Feature Extraction modules. This research paper focuses on the texture of the leaves in helping ascertain disease symptoms. Texture features extracted using this approach along with color and shape features have been used to identify different mango features and tomato diseases.

II. LITERATURE SURVEY

The mango leaves are either lanceolate or elliptical with long stalk and a texture which is thick, smooth at the top and leathery in terms of its feel. But there are finer elements in recognizing leaves say from three different mango trees. An orchard manager is able to recognize the mango variety by

squeezing the leaf and smelling it. For instance the Totapuri and Neelam mango from Karnataka can be recognized by the smell of its crushed leaf. With experience they are able to visually identify mango diseases like anthracnose, powdery mildew, sooty mold, insect infections like Mango Hopper.

The collective wisdom of hundreds of orchard managers would be effective for any manager. System aims to learn over time different symptoms seen across India and classify them better. The disease symptoms can be manifested in the bark, leaf or fruit from an external diagnosis. Detecting at the fruit stage is good from quality assurance point of view but a bit late from the farmer's point of view. If the problem is detected early then a solution can greatly help the farmer to yield a better harvest. Textures of a leaf can be analyzed by a Structured and Statistical approach. The former is useful in case of artificial textures. Since this research is for mango leaves the statistical approach was selected. Two statistical methods have been used to study, the Histogram of Intensity and the Gray Level Co Occurrence Matrix. The gray co-occurrence matrix method was proposed by **Haralick**[1]. Second order statistics are derived from it. Statistical measurement in co-occurrence method could use different values of the angle and distance varied from 1, 3 and 5 pixel distance and later averaged for one distance **Camargo et al**[2]. They got an 83% accuracy while using just texture based features for cotton crops. Fractal Dimension uses a Box counting algorithm, **Li et al** [3] used an enhanced counting algorithm to classify textures. Lacunarity using a gliding box algorithm by **Plotnick et al** [4] for classifying landscape texture. There are also method where the feature extraction is not specific but uses a general method like approaches based on Convulsion Neural Network like a deep learning model for plants was suggested by **Ferentinos** [5].

A detailed survey on plant extraction and segmentation can be found in **Barbedo**[6] and **Hamuda et al** [7]. Detection of unhealthy leaves using texture features extracted from co-occurrence matrix has been done by **Azivazhagan et al**[8]. **Huang et al**[9] used a mixture of color and texture features. **Guerrero et al**[10] used SVM to classify weeds for maize crop.

The histogram computes the occurrences of each intensity in the image, whereas GLCM is taking into consideration of relative occurrence of two intensity.

Statistical functions which can be applied to the **histogram** are:

Mean is a measure of average intensity

Standard Deviation is a measure for average contrast

Smoothness measures the smoothness in the intensity values in a region

Revised Manuscript Received on April 18, 2020.

* Correspondence Author

Jibrael Jos*, Department of Computer Science, Christ University, Bengaluru, India. Email: jibrael.jos@res.christuniversity.in

K A Venkatesh, Math and Computer Science, Myanmar Institute of Information Technology, Mandalay Myanmar. Email: ka_venkatesh@miit.edu.mm

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Table- I: Statistical formulas applied over of intensity histogram

Property	Formula
Mean	$m = \sum_{i=0}^{L-1} z_i p(z_i)$
Standard Deviation	$\sigma = \sqrt{\mu_2}$ $\mu_2 = \sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)$
Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$
Entropy	$e = - \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$

Uniformity value is high if intensity are all same and decreases with variations

Entropy is a measure of randomness, in case the probability of some intensities are higher the log factor penalizes the overall value.

Table- II: Statistical formulas applied over of Grey Level Co-occurrence Matrix(GLCM). (d=1,2,3)

Property	Formula
Contrast	$\sum_{i=1}^K \sum_{j=1}^K (i - j)^2 p_{ij}$
Correlation	$\sum_{i=1}^K \sum_{j=1}^K \frac{(i - m_r)(j - m_c)p_{i,j}}{\sigma_r \sigma_c}$
Energy (Angular Second Moment)	$\sum_{i=1}^K \sum_{j=1}^K p_{ij}^2$
Homogeneity	$\sum_{i=1}^K \sum_{j=1}^K \frac{p_{ij}}{1 + i - j }$

Contrast is a measure of the intensity contrast between a pixel and its neighbor.

Correlation is a measure of how correlated a pixel is to its neighbor. Means for the row(m_r)and column(m_c) in GLCM has to be calculated along with standard deviation along the row(σ_r) and column(σ_c).

Energy is the sum of squared elements in the GLCM. Advantage of squaring is that preference is given for large values.

Homogeneity also known as uniformity and angular second moment. Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

III. MATERIALS AND METHODS

For *Mango* trees four different textures have been studied: dry, growth/spots, new leaves, healthy mature leaves. For *Tomato* three diseases Early Blight, Yellow Leaf Curl Virus (YLCV), Leaf Mold along with healthy leaves. Three columns in the figure 1 belong to each of the disease respectively.



Fig. 1. Leaf samples of Early Blight, YLCV and Leaf Mold

A. Existing Texture Features

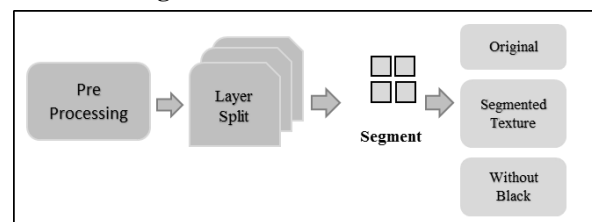


Fig. 2. Process Flow for Texture Analysis

The texture feature system has the following steps

1. Preprocessing and Background Identification
2. Clustering to Separate Region of Interest
3. Segmenting the leaf into smaller block
4. Texture Analysis based on
 - a) GLCM on original image block
 - b) Extracted Texture
 - c) Eliminate leaf and background co-occurrence
5. Support Vector Machine(SVM) based classification along with shape, color and texture features

B. Proposed Method for Pseudo CCM

A variation of GLCM uses color based co occurrence method (CCM) as used by **Bashish** et al [11]. It creates three CCM matrices for Hue, Saturation and Intensity. Some extracted texture features include Angular Moment, Mean Intensity, Variation of Image Intensity, Correlation, Produce Moment, Contrast, Entropy related ones. **Choi** et al [12] proposed color features based on Gabor Wavelets and Local Binary Pattern. Another variation of GLCM for example using color information as done by **Hossain** et al [13]. **Benco** et al [14] proposed a color based texture features. A variation of CCM to identify a froth texture **Gui** et al [15]. The proposed method is a variation of CCM and uses a specified set of Pseudo Colors

Step 1: Convert color Image to Pseudo Color Matrix.
Values as shown in Table III.

- Step 1a: Each Pixel is converted to pseudo color based on its H, S, V value based on the HSV color model.
- Step 1b: Black and White are based on Saturation and Value rather than Hue
- Step 2: Create a color co-occurrence matrix, with offset of 1 pixel in direction of 90, 45, 0, -45 degrees
- Step 3: Statistical functions are applied on the matrix to identify Correlation, Contrast, Energy and Homogeneity
- Step 4: Distance between two color is based on circular model with Black and White considered 6 points away and all color with Black/White are considered 3 points away

Table- III: Pseudo Colors for CCM

Pseudo Color	Criteria	Index
Black	Value < 0.1	0
White	Value > 0.1 Sat < 0.1	13
Red	Hue 0	1
Orange	Hue 30	2
Yellow	Hue 60	3
Yellow Green	Hue 90	4
Green	Hue 120	5
Green Cyan	Hue 150	6
Cyan	Hue 180	7
Cyan Blue	Hue 210	8
Blue	Hue 240	9
Violet	Hue 270	10
Magenta	Hue 300	11
Red Magenta	Hue 330	12

Difference is calculated based on a cyclic value (360 is 0) such that difference between 7 and 12 is 5 as normally done but 1 and 12 is 1 (from other side) 3 and 11 is 4 (like a clock)

Difference is used in the formula for Contrast and Homogeneity
This method is a follow up of the Hue based color features extracted for disease detection. So apart from dominant color and ratios of pseudo colors we can apply texture based on colors itself.

C. Image Set

Mango leaves were collected and scanned from orchards in Karnataka, India. This set has 500 images. The leaves are numbered 1 to 500 and leaves are either having a gray background or a black background. Leaves have been collected from different trees, varieties, orchards and seasons. Tomato leaves have been taken from **Plant Village** dataset [16]. Four hundred leaves have been taken for classifying three diseases.



Fig. 3. Leaf split into regions of interest

In the pre-processing phase the leaves are put through a segmentation which splits leaf into multiple layers using a k-means clustering based on the color for Mango. For tomato a HSV based color mask helped in eliminating the background and healthy portion of leaf leaving only the region of interest. Splitting the leaf into layers can be seen in the Fig 3. First image is the original leaf, second and third is the same leaf split into layers. In the second image we can see that the background is black. **Guijarro** et al [17] used a texture segmentation approach but here a color based Segmentation was used using a K-Means clustering method **Bashish**[11]. **Zhang** et al[18] used a variation of the to achieve a better level of segmentation. If the texture features are extracted on the original image. Contrast and Homogeneity values get dependent on the texture of the leaf and veins too. To make the texture features reflect the texture itself. The segmented layer can be seen in Fig 4.

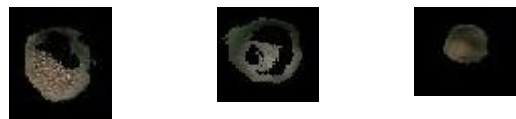


Fig. 4. Leaf Segment with growth for texture analysis

Once the GLCM is generated on this region of interest, the black background has to be ignored else the statistical values will get effected by black with black and black with texture co-occurrences.

Like in the case of energy black background effects the intensity, hence energy without lowest row and column in GLCM is more accurate.

Overall Features:

- Shape:5 (5 relevant out of 9)
- Texture: 21 (60 measurements)
- Color: 8 features (32 measurements)

IV. RESULTS AND DISCUSSION

Textures when applied to Tomato disease was not accurate and did not help in classification. It was noticed that if the image was broken into four quadrants then statistical functions in the four quadrant were different in most cases. Presence of a vein in an image could throw the statistical functions off. In figure 5 the min, max and average values can be seen for three leaf diseases (sample size of 3 diseases of 100 images each)

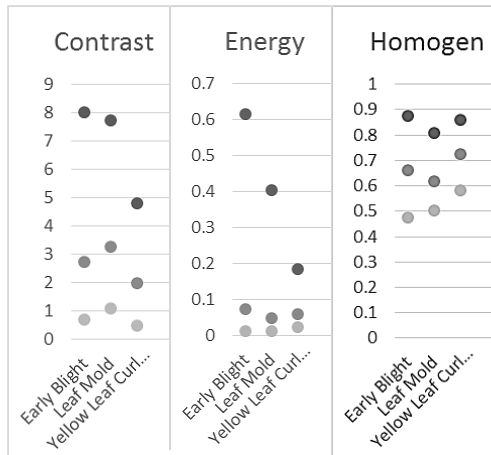


Fig. 5. Graph showing GLCM based texture results, plots maximum average and minimum values for each disease So experimentation was done to see in which cases can texture be actually used directly (Part A) an then impact of proposed Pseudo Color Matrix was analysed(Part B)

A. Texture Feature Analysis

The texture features along with basic shape and color features extracted using the process was fed into a Support Vector Machine with two batches of 100 mango leaf textures. The SVM was cross validated and then model was tested for a test data.

In the system created color features themselves give a high level of accuracy in classifying different disease symptoms. So the attempt here is to classify textures based on how they look and feel, to see if the model created can differentiate between the different classes.

Growth: Normally circular in nature, should have intensity variation due to it being a bump on the plane.

Dry Leaves: Rough or coarse, the intensity profile should be more jagged and should differ from a symmetric profile in case of a circular growth

Young Leaves: Can be easily recognized due to different color but from a texture point of view. It is quiet smooth and a certain amount of reflectance

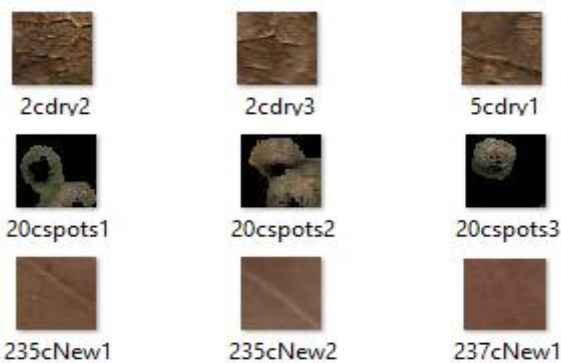


Fig. 6. Process Flow for Texture Analysis

In the table IV there are three texture categories: Dry leaves, leaves with growth/spots and new leaves Hue will be different for all three types but we will discuss texture here:

Variance and Standard Deviation

New<Dry<Spot

Homogeneity

Spot<~Dry<New

Energy

Spot~Dry<New

Contrast

New<Dry<~Spot

Where the symbols mean

< **less than**

~ **values close**

The results of two batches of 100 have been tabulated in the table IV, as it can be seen the accuracy of detecting the healthy leaf segment was 87% on average.

Table- IV: GLCM Classification Results

Class A	Class B	SVM Classification	
Growth Area	Healthy/Other	95%	92%
Dry Areas	Healthy/Other	94%	NA
Young Leaves	Healthy/Other	88%	85%
Healthy	Other	86%	88%

In batch 1 the growth, dry, young were 20 each and healthy were 40.

In batch 2 whole set of new images were taken and instead of dry leaves others with a mix of non-healthy symptoms were included. The purpose was to check how effective was classification of healthy leaves and growth areas based on texture.

Bashish et al[11] model gave an accuracy of various diseases 92% on the average

Texture Features accuracy can be seen in Table V. This is based on a *block wise (n pixel by n pixel)* value and not for the whole leaf

Table- V: Texture Feature Group Accuracy

Feature	Settings	Group Accuracy (Average over multiple batches)
Contrast Correlation Energy Homogeneity	Using GLCM with d=1,2,3 Angle: 0,45,90,-45 and all neighbor Based on H and V channel in the HSV model Includes only <i>texture of interest</i> and not background	Above 70% (Without Color, Shape and Other texture features)
Mean Standard_Deviation Entropy RMS Variance	On the intensity histogram relative position of intensity is ignored here	Above 50% (Without Color, Shape and Other texture features)

Mean, Standard Deviation and Variance were able to differentiate textures but entropy values and RMS were not effective and for the similar textures gave very different values for different sample.

Camargo et al [2] texture features overall gave 83% accuracy and around 53% individually.

Changing the distance value from 1,2,3 the advantage was not much just d1 with other features gave an equally good result. The table shows impact when only 4 features for each value were independently applied.

Contrast and Energy is not effective in detecting a new leaf versus a healthy mature leaf. But when comparing a dry region or growth the values are different from healthy leaves.

Shape features seem to classify well for smaller dataset but as dataset size increase it tends to classify textures wrong. Eccentricity showed values close to each other for textures with growth in it.

Homogeneity as expected in a growth will be different from a new healthy leaf. Though color features are effective in differentiating certain features. The study was able to understand the key features which are effective in disease detection in case of Mango trees

B. Pseudo Color Co Occurrence Analysis

Pseudo CCM based Contrast and Energy gave varied value for different diseases, though homogeneity and correlation did not prove very useful for detection of these three diseases. Homogeneity was useful for detecting whether a given leaf is healthy or not.

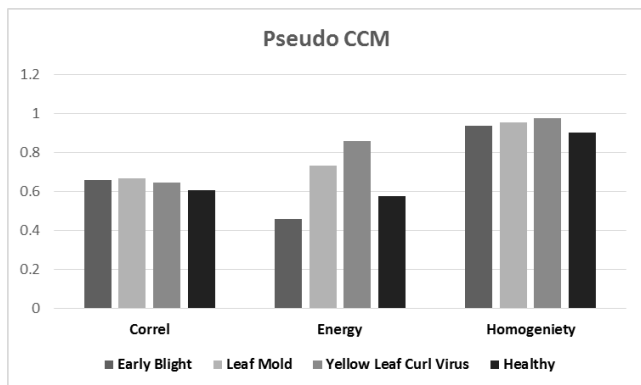


Fig. 7. Results of Second Order Stats for three tomato diseases

In the tomato dataset there are leaves whose disease symptoms is negligible hence there are cases where wrong classification was done. Along with texture features using color features was found to be more effective as can be seen in Table VI.

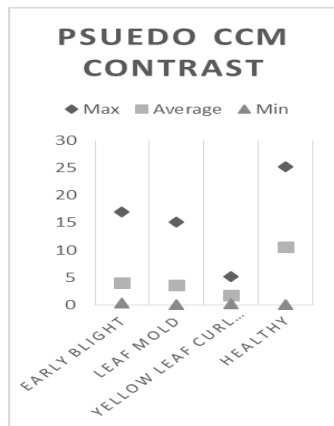


Fig. 8. Pseudo CCM based contrast

As can be seen in the graph in figure 8 contrast gives a more varied value for the three diseases but some outlier with very less symptom does create confusion in classification. The contrast values are close to zero for some images. Just the texture gave accuracy of around 73% along with CCM the accuracy went up to 79% and along with color features it went up to 90%. Results in the table VI below are for 400 images with each class having around 100 images. The main problem is just the texture cannot differentiate between early blight and leaf mold. Color based texture gave a higher level of accuracy.

Table- VI: PCCM Classification Results

Class	Texture	PCCM, Dom Hue	Text,P CCM	Text,P CCM, Color
Early Blight	49	87.3	86.3	82.4%
Leaf Mold	65	65	80.4	86.3%
Yellow Leaf Curl Virus	85.3	85.3	90.2	90.2%
Healthy	92.2	80.4	82.4	89.2%
Overall	73%	79%	85%	87.0%

ROC curve for SVM for tomato diseases showed area of 95%, 94%, 96% respectively under the curve.

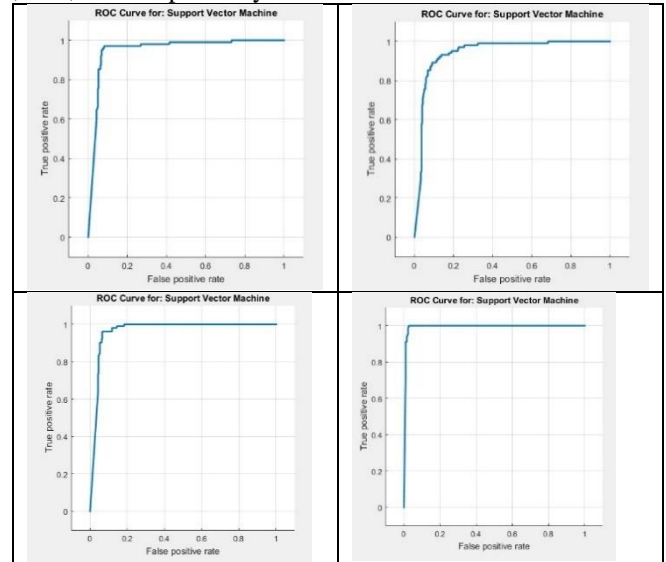


Fig. 9. ROC Curve for Early Blight, Leaf Mold, YLCV, Healthy Leaves

V. CONCLUSION

The texture features extracted can be used to create a feature bucket to identify disease in Mango Trees and Tomato Plants. The texture analysis of the feature along with shape and color were able to classify three diseases effectively. But just the GLCM based features were not enough. They worked well only when regions of interest were distinctly different. Eliminating the healthy and background pixels from the texture made the statistical values more effective in classifying different patterns. Proposed Pseudo CCM worked better and in the case of Tomato Diseases were effective in classifying 3 diseases. Pseudo CCM could also be customised based on disease being studied by adding more pseudo colors like dark green and shades of brown, shades of red. Pseudo CCM method also helped in adding features like dominant hue and second most dominant hue, Adjacency of colors like Yellow with Green, Brown with Green.

REFERENCES

1. Haralick R M, Shanmugam K, Dinstein Its'Hak (1973) Textural features for image classification, Systems, Man and Cybernetics, *IEEE Transactions*, vol. SMC-3, no. 6, pp. 610-621.
2. Camargo, A., and J. S. Smith. "Image pattern classification for the identification of disease causing agents in plants." *Computers and Electronics in Agriculture* 66.2 (2009): 121-125.
3. Li, Jian, Qian Du, and Caixin Sun. "An improved box-counting method for image fractal dimension estimation." *Pattern Recognition* 42.11 (2009): 2460-2469.

4. Plotnick, Roy E., Robert H. Gardner, and Robert V. O'Neill. "Lacunarity indices as measures of landscape texture." *Landscape ecology* 8.3 (1993): 201-211.
5. Ferentinos, Konstantinos P. "Deep learning models for plant disease detection and diagnosis." *Computers and Electronics in Agriculture* 145 (2018): 311-318.
6. Barbedo, Jayme Garcia Arnal. "Digital image processing techniques for detecting, quantifying and classifying plant diseases." *SpringerPlus* 2.1 (2013): 660.
7. Hamuda, Esmael, Martin Glavin, and Edward Jones. "A survey of image processing techniques for plant extraction and segmentation in the field." *Computers and Electronics in Agriculture* 125 (2016): 184-199.
8. Arivazhagan, Sai, et al. "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features." *Agricultural Engineering International: CIGR Journal* 15.1 (2013): 211-217.
9. Huang, Kuo-Yi. "Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features." *Computers and Electronics in agriculture* 57.1 (2007): 3-11.
10. Guerrero, José Miguel, et al. "Support vector machines for crop/weeds identification in maize fields." *Expert Systems with Applications* 39.12 (2012): 11149-11155.
11. Al Bashish, Dheeb, Malik Braik, and Sulieman Bani-Ahmad. "Detection and classification of leaf diseases using K-means-based segmentation and." *Information Technology Journal* 10.2 (2011): 267-275.
12. Choi J-Y, Ro Y-M, Plataniotis K N (2012) Color local texture features for color face recognition, *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1366-1380.
13. Hossain K, Parekh R (2010) Extending GLCM to include color information for texture recognition, International Conference on Modeling, Optimization and Computing, Book Series: *AIP Conference Proceedings*, vol. 1298, pp. 583-588, ISSN: 0094-243X, ISBN: 978-0-7354-0854-8.
14. Benco M, Hudec R (2007) Novel Method for Color Texture Features Extraction Based on GLCM", *Radio Engineering*, Vol.16, No.4, pp. 64-67, ISSN 12102512.
15. Gui, Weihua, et al. "Color co-occurrence matrix based froth image texture extraction for mineral flotation." *Minerals Engineering* 46 (2013): 60-67.
16. Hughes, David, and Marcel Salathé. "An open access repository of images on plant health to enable the development of mobile disease diagnostics." *arXiv preprint arXiv:1511.08060* (2015).
17. Guijarro, Maria, et al. "Automatic segmentation of relevant textures in agricultural images." *Computers and Electronics in Agriculture* 75.1 (2011): 75-83.
18. Zhang, Shanwen, et al. "Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG." *Optik* 157 (2018): 866-872.

AUTHORS PROFILE



Jibrael Jos was an Associate Professor at Christ University for 8 years. He has worked in the IT Industry for 12 years. He has worked with MindTree and CMC Limited. He has a Masters in Computer Science from Birkbeck College, London University. Jibrael has completed his MPhil in Computer Science and is pursuing Ph.D at Christ University in the field of Artificial Intelligence and Computer Vision. He has papers published in Data Warehousing, Image Processing and Artificial Intelligence. Currently he is teaching as an Adjunct Associate Professor at Christ University, Hosur Road, Bengaluru, India.



Dr. K A Venkatesh has 3 decades of teaching and 4 years of Industry, served as CTO. He served as Registrar, Dean-Research, Head of Department, Principal. He has published 65 papers and coauthored a book on Discrete Mathematics. He has completed his Ph.D from Alagappa University. Also authored 5 book chapters. He is a member of Systems Society of India, RMS, Constrained Programming, USA, Academy of Discrete Mathematics. Serving as Adjunct Professor at IIIT-B and Professor of Math and Computer Science, Myanmar Institute of Information Technology, Mandalay, Myanmar.